

NOVEL MATHEMATICAL APPROACHES TO CT IMAGE ANALYSIS IN ASTHMA

**Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester**

by

**Alexander James Bell MMath
Department of Respiratory Sciences
University of Leicester**

September 2019

Novel mathematical approaches to CT image analysis in asthma

Alexander J. Bell

Abstract

Asthma is a complex respiratory disease characterised by spatial ventilation heterogeneity (VH) in the lung. Multiple breath washout (MBW) and forced oscillation technique (FOT) testing has been shown to quantify VH at the mouth. Methods to provide fresh insight into the mechanisms driving VH are urgently needed to support treatment development.

Computed tomography (CT) with parametric response map (PRM) analysis provides spatial and functional lung data. Integrated data sets of imaging features and VH markers present the problem of extracting meaningful information from high dimensional objects; topological data analysis (TDA) has demonstrated recent success in solving this problem.

In this thesis topological methods are applied to extract information about lung anatomy and function from imaging data to study association with VH in innately high dimensional integrated data sets. It was sought to describe the spatial distributions of voxel based metrics, visualise the lungs using TDA methods, and combine PRM with simulated airway tree models to improve computational modelling of FOT.

Data was obtained from a clinical trial involving 52 subjects, 41 asthmatic (11 healthy). An algorithm was developed to quantify directional Hounsfield unit change (Δ HU) gradients. TDA was applied to integrated data sets and voxel sets with measures of airway tree depth. FOT modelling was augmented with Δ HU about simulated acinar airways.

Inferior-to-superior Δ HU gradient best differentiated VH markers. Visualisations of the lungs illustrated gradient reversal and increased heterogeneity in Δ HU in asthmatics. FOT modelling using Δ HU parameterisation showed increased correlation between clinically measured and simulated FOT small airway resistance (R5-R20).

In this thesis additional insight is presented into mechanisms affecting VH markers derived from MBW and FOT. Topological methods to analyse imaging data were developed and implemented, including workflows to visualise the lungs and lung function. Linking simulated airway tree geometry with Δ HU showed improvement in FOT modelling.

Acknowledgements

I wish to acknowledge Prof. Salman Siddiqui and Prof. Alexander Gorban, my supervisors from the medical and mathematical disciplines respectively, for their close guidance and support throughout this interdisciplinary research project, and whose collaboration secured funding from the Sir Jules Thorn charitable trust for my PhD studentship; I want to gratefully thank the Sir Jules Thorn trust for funding this work.

Salman provided close instruction and guidance on the medical problems and literature motivating the work, and was chiefly involved in the conceptual design of all studies presented in this thesis. He has been extremely energetic and engaged in my project throughout providing advice and direction on all aspects of the work, and extensive mentoring in academic life skills, particularly writing for journals.

Alexander advised me on the mathematical challenges perceived within the work and provided expert advice on novel methods developed during the project. He provided useful critique of results generated throughout my research, and insightful explanation of the outcomes observed. I am very grateful to have had the opportunity to enjoy so many entertaining and educational discussions with a truly great mathematician.

I wish to thank Dr. Pranabashis Haldar, Dr. Evgeny Mirkes and Prof. Jeremy Levesley, for participating in my progress review panel. Throughout my project they have provided excellent feedback on the work I presented and objective advice on how to deal with the many challenges that arose throughout my research experience.

Prof. Craig Galbán provided excellent support and co-supervision from the very beginning of my project, with expertise in image registration methods that underpinned my entire project. I am grateful for having had the opportunity to visit the University of Michigan (USA), and further thank Aleksa, Ben, Kevin and Amanda for their friendly support during my visit, and Tania and Carrie for their administrative assistance.

Dr. Matthew Richardson has been incredibly supportive throughout my PhD course, sharing his expertise and valuable experience in programming and applied statistics; particularly bio-statistical analysis. Matt provided excellent co-supervision throughout my project advising on my statistical methods and results, and participating in the learning and applying of TDA mapper using Ayasdi software.

I want to thank Marcia Soares for her friendship and support as a colleague where I was based in Glenfield Hospital in Leicester (UK), alongside Sherif, Latifa, Lorna and Michael who were present and very supportive during my initial years in that environment. I also wish to thank Rob and Bo who were present during the later years and provided advice and support regarding bioinformatics and operations using the Linux OS. Additionally I am very grateful to Susan Kapadia, our departmental post-graduate research administrator, for her excellent support throughout my course and energy organising the post-graduate student staff committee in which I participated.

I wish to acknowledge Brody Foy, who has been an extremely energetic and capable collaborator from about the middle of my project, providing support and collaboration on work involving extended airway centreline models and in particular his expertise in FOT modelling; his assistance in data sharing, understanding and independent investigation significantly supported all of the studies presented in this thesis.

Finally I wish to thank my family; my parents Graham and Fiona, my siblings Jonathan and Rebecca, and my uncle Adrian, for their extensive support throughout my undergraduate and post-graduate studies at the University of Leicester, enabling me to achieve my academic goals, and without whom I would not be where I am today.

Statement of work personally performed

I was primarily engaged in the designing and implementation of algorithms using Matlab to extract features from CT data, followed by statistical and visual analyses of the resulting data sets. I received raw imaging data in the form of .mat files and .dcm series and performed QC/QA using dicom viewer and my own scripted software for all subjects. Additionally I curated non-CT data sets from MS Excel spreadsheets, from multiple sources in the clinical workplace, for the purpose of focused comparative analysis with CT imaging data.

I developed the stratified axial analysis and other described methods using image registered voxel sets. I was initially provided with registered imaging data at the start of my project by Prof. Craig Galb  n from the University of Michigan (USA); I later visited the centre for molecular imaging (CMI) lab directed by Craig and learned to perform image registration on a sabbatical visit in June and July of 2018, returning with expertise and code to undertake expiration-to-inspiration registration for the final study presented in this thesis. Additionally I developed BASH scripts and configured a Linux workstation to run automated image registration in batches using Imbio software.

I designed and wrote algorithms aligning and evaluating point sets of registered lung voxels and simulated airway extended centrelines supported by Brody Foy, a post-graduate at the University of Oxford at the time of work. I designed, implemented and analysed all work done using TDA mapper in both phenotyping population data and generating visualisations of the lungs; this includes Python scripts automating network generation and persistent homology study. I performed all statistical analyses presented in the studies, and generated visualisations of joint density histograms, min-max projections and tree structures with functional summaries annotated to branch points.

In addition to work presented in the thesis, I processed over 350 subjects from the ‘Atlantis’ study, from DVDs stored on site at Glenfield (UK), through E2I image registration and performed Hounsfield unit (HU) sampling in all cases for normalisation study using rods of known density, using my own scripts automating the sampling and data storage processes; this data set was a significant output of the work that is yet to be analysed due to time constraints and limited availability of non-CT features.

Abstracts and publications arising from this thesis

Published original article

Bell AJ, Foy BH, Richardson M, Singapuri A, Mirkes E, van den Berge M, Kay D, Brightling C, Gorban AN, Galbán CJ, Siddiqui S. Functional CT imaging for identification of the spatial determinants of small-airways disease in adults with asthma. *Journal of Allergy and Clinical Immunology*. 2019 Jan 22.

Supported publications

Foy BH*, Soares M*, Bordas R, Richardson M, **Bell A**, Singapuri A, Hargadon B, Brightling C, Burrowes K, Kay D, Owers-Bradley J. Lung computational models and the role of the small airways in asthma. *American journal of respiratory and critical care medicine*. 2019 May 18(ja). * shared first-authorship.

Siddiqui S, Shikotra A, Richardson M, Doran E, Choy D, **Bell A**, Austin CD, Eastham-Anderson J, Hargadon B, Arron JR, Wardlaw A. Airway pathological heterogeneity in asthma: Visualization of disease microclusters using topological data analysis. *Journal of Allergy and Clinical Immunology*. 2018 Nov 1;142(5):1457-68.

Abstracts

Foy B, **Bell A**, Siddiqui SH, Kay D. Low frequency lung resistance is a global bronchoconstriction detection measure, but is still sensitive to small airways disease. In B65. ASTHMA: PATHOPHYSIOLOGY AND CLINICAL TRIALS 2018 May (pp. A3928-A3928). American Thoracic Society (ATS) 2018.

Bell A, Richardson M, Singapuri A, Mirkes E, Gorban A, Galbán C, van den Berge M, Brightling C, Siddiqui S. Parametric response map registered CT feature and small airway physiology analysis in asthma. *European Respiratory Society (ERS)* 2017.

Contents

1	Introduction	1
1.1	Overview	1
1.2	Asthma	4
1.2.1	Definition	4
1.2.2	Classification	4
1.2.3	Nasal Transcriptomics	5
1.2.4	Lung Structural Elements	6
1.2.5	Ventilation Heterogeneity	8
1.3	Lung Function Testing	10
1.3.1	Body Plethysmography	10
1.3.2	Spirometry	11
1.3.3	Multiple Breath Washout	13
1.3.4	Impulse Oscillometry	15
1.4	Imaging	17
1.4.1	Modality Overview	17
1.4.2	Parametric Response Map	21
1.4.3	Computational Modelling of FOT	24
1.5	Data Analysis	26
1.5.1	Big data	26
1.5.2	Statistics	27
1.5.3	Topological Data Analysis	30
1.6	Aims and hypotheses	33
1.6.1	Perform feature extraction and statistical analyses on PRM registered CT data to study association with clinical markers of asthma and VH.	34
1.6.2	Apply TDA to CT based data sets for visualisation of the lungs and lung function in asthma	35
1.6.3	Link functional CT imaging data with airway tree structures to support FOT modelling with regional compliance parameterisation.	36
1.7	Structure of thesis	37

2	Methods	39
2.1	Overview	39
2.2	Clinical Procedures	41
2.2.1	Subjects	41
2.2.2	Questionnaires (ACQ/AQLQ)	42
2.2.3	Atopy Testing	42
2.2.4	Induced Sputum Tests	42
2.2.5	Nasal Transcriptomics	43
2.3	Lung Function Tests	43
2.3.1	Body Plethysmography	43
2.3.2	Spirometry	43
2.3.3	Multiple Breath Washout Testing	44
2.3.4	Impulse Oscillometry	44
2.4	CT Imaging	45
2.4.1	Acquisition	45
2.4.2	Segmentation and Registration	45
2.4.3	FOT Modelling	50
2.4.4	Set Alignment	52
2.5	Feature Extraction	54
2.5.1	Features from Joint Density Histogram	54
2.5.2	Global Lung Features	56
2.5.3	Stratified Axial Analysis	58
2.5.4	Inter-Lung Features	60
2.5.5	Voxel SO Score	62
2.5.6	Branch Δ HU assignment	62
2.6	Analysis	64
2.6.1	Statistical Methods	64
2.6.2	Min-max Projection	71
2.6.3	Topological Data Analysis	71
2.6.4	Tree Visualisation Algorithm	78
2.6.5	Constriction / Correlation Analysis	80

3	Studies.....	82
3.1	Functional CT imaging for identification of the spatial determinants of small airways disease in adult asthma	82
3.1.1	Introduction.....	83
3.1.2	Methods	84
3.1.3	Results.....	90
3.1.4	Discussion.....	95
3.1.5	Figures and tables	99
3.2	Topological data analysis to visualise the lungs and small airways in adult asthma.....	127
3.2.1	Introduction.....	128
3.2.2	Methods	130
3.2.3	Results.....	132
3.2.4	Discussion.....	134
3.2.5	Figures and tables	136
3.3	Functional CT image analysis supports computational modelling of small airway resistance in asthma.....	147
3.3.1	Introduction.....	148
3.3.2	Methods	150
3.3.3	Results.....	153
3.3.4	Discussion.....	155
3.3.5	Figures and tables	158
4	Conclusions	171
4.1	Discussion of results.....	171
4.2	Future work	177
4.3	Closing statement	180
5	Appendices	181
6	References	193

List of Tables

Table 3.1-1: Clinical characteristics of asthmatic and healthy subjects.	113
Table 3.1-2: Computed tomography imaging biomarkers and ventilation heterogeneity based stratification.	115
Table 3.1-3: Reference list describing extracted CT attributes investigated in this study.	116
Table 3.1-4: CT data across all groups, with treatment rationalized GINA stratification applied to asthma cohort.	117
Table 3.1-5: Non-CT features with ventilation heterogeneity based stratification.....	118
Table 3.1-6: Computed tomography imaging biomarkers and ventilation heterogeneity based stratification (full).	119
Table 3.1-7: Summary of linear statistical analyses: average correlation and linear regression (raw value & PCA).	120
Table 3.1-8: Linear regression co-efficient table for linearity reduced CT subset regressions.	121
Table 3.1-9: Principal component loading scores for components used in linear regression.	122
Table 3.1-10: Linear discriminant analysis based classification percentage accuracy with selected feature sets.	123
Table 3.1-11: Linear discriminant analysis coefficients on combined feature sets.	124
Table 3.1-12: Genes highlighted from GE hierarchical cluster analysis (Figure 3.1-13).	125
Table 3.1-13: TDA network group comparison (refer to Figure 3.1-14).	126
Table 3.2-1: Clinical characteristics for selected asthmatic (n=2) and non-asthmatic (n=2) subjects.....	145
Table 3.2-2: Complete list of 12 features selected for upload to Ayasdi server.....	146
Table 3.3-1: Clinical characteristics across all subjects grouped by asthma diagnosis.	169
Table 3.3-2: Branch radii, ΔHU and resistance across all subjects, grouped by asthma diagnosis.	170

List of Figures

Figure 1-1: Sources of data analysed in this thesis from multiple length scales.	3
Figure 1-2: H-tree plotting after four and ten iterations from the root branch.	6
Figure 1-3: Strahler order assignment to an exemplar tree structure.	7
Figure 1-4: Increasing ventilation heterogeneity (VH) with decreasing asthma control. .	9
Figure 1-5: Body plethysmography box and volume-displacement diagram.	10
Figure 1-6: Flow-volume loop from spirometry.	11
Figure 1-7: Volume-time graph from spirometry.	12
Figure 1-8: Washout curve with synced flow-time graph from MBW.	14
Figure 1-9: Phase III slope analysis of washout curve segment from MBW.	14
Figure 1-10: Equipment diagram for impulse oscillometry.	15
Figure 1-11: Resistance/reactance curves as a function of frequency.	16
Figure 1-12: CT scanner equipment showing significant operating components.	18
Figure 1-13: MRI scanner equipment diagram.	20
Figure 1-14: Image registration and parametric response map voxel classification.	22
Figure 1-15: Demonstration of topologically driven PRM improving disease phenotyping.	23
Figure 1-16: Airway centreline from CT and associated circuit model of impedance. ...	25
Figure 1-17: Global data storage estimations over the past two decades.	27
Figure 1-18: Dimension of a set of points, association with principal components.	29
Figure 1-19: The seven bridges of Königsberg and associated representative graph.	30
Figure 1-20: Simplicial complex with associated Betti numbers β_0 and β_1	31
Figure 1-21: Data compression using TDA mapper algorithm.	32
Figure 1-22: Diagrammatic overview of study aims and work flow.	33

Figure 2-1: Imbio lung density analysis implementing PRM (I2E).	46
Figure 2-2: Lung volume segmentation and airway centreline computation using Mimics.	47
Figure 2-3: Yacta volume identification and segmentation with CT data.	48
Figure 2-4: Image registration using Matlab with CMI libraries (Michigan, USA).....	49
Figure 2-5: Extended centreline modelling from CT derived airways.	50
Figure 2-6: Manual alignment with visualisation of point sets.	53
Figure 2-7: Joint density histogram (JDH) of voxel HU with associated feature extraction.	55
Figure 2-8: ΔHU as a projection from eHU iHU plane, and relation to PRM.....	57
Figure 2-9: Mean ΔHU in slices, anterior-posterior (x) and inferior-superior (z) directions.....	58
Figure 2-10: Visual evaluation of 2-means lung identification.	61
Figure 2-11: Cell array assignment for efficient checking of point proximity sets.....	63
Figure 2-12: JDH ellipse approximation in disparate example cases.	70
Figure 2-13: Coronal and transverse plane min-max projection of PRM^{fSAD} in one subject.	73
Figure 2-14: Graphical representation of Reeb graph calculation from the torus.	73
Figure 2-15: TDA mapper from first principles to k-NN based output.....	74
Figure 2-16: SO1 ΔHU assignment visualisation of averages assigned to upper airways.	79
Figure 2-17: Histograms of mapping raw ΔHU values to airway adjustment factors. ...	81
Figure 3.1-1: JDH and association with obstruction (spirometry).....	99
Figure 3.1-2: Subject postures with respect to gravity and lung density distribution. .	100

Figure 3.1-3: Decile-wise tertile analysis, min-max projections and phenotype observed.	101
Figure 3.1-4: Coronal heat maps of ΔHU and PRM^{fSAD} in SAAz low and high tertiles.	102
Figure 3.1-5: Visual overview of linear discriminant analysis on disparate data domains.	103
Figure 3.1-6: Effect of computation model based regionalisation on small airway physiology.....	104
Figure 3.1-7: ΔHU derivation and inferior-superior SAA technical illustration.	105
Figure 3.1-8: ΔHU gradient measure methodology and combinatorial voxel features.	106
Figure 3.1-9: PRM ellipse area and PRM^{fSAD} association to VH markers S_{acin} and R5-R20.....	107
Figure 3.1-10: Decile-wise tertile analysis of ΔHU changes, anterior-to-posterior direction.	108
Figure 3.1-11: Comparing ΔHU across PRM^x voxel groups in $\Delta\text{HU}^{\text{IS}*}$ low/high tertiles.	109
Figure 3.1-12: Complete original data matrix correlation visualization and collinearity reduction from feature selection.	110
Figure 3.1-13: Hierarchical cluster analysis of nasal gene expression discriminating $\Delta\text{HU}^{\text{IS}*}$	111
Figure 3.1-14: Centrality based TDA network linking asthma with $\Delta\text{HU}^{\text{IS}*}$	112
Figure 3.2-1: Overview of TDA image analysis workflow developed in this study.	136
Figure 3.2-2: TDA network for older non-asthmatic subject (low VH).....	137
Figure 3.2-3: TDA network for younger non-asthmatic subject (minimal VH).....	138
Figure 3.2-4: TDA network for severe asthmatic subject (GINA 4, high VH).	139

Figure 3.2-5: TDA network for mild asthmatic subject (GINA 3, mild VH).....	140
Figure 3.2-6: Interpreting network heat maps for spatial features.....	141
Figure 3.2-7: Anatomical segmentation using TDA.....	142
Figure 3.2-8: High level methodological overview of TDA based organ visualization workflow.....	143
Figure 3.2-9: Strahler order (SO) score assignment method with k-NN.....	144
Figure 3.3-1: Graphical abstract of study: raw CT data to augmented FOT modelling.	158
Figure 3.3-2: Calculation of ΔHU scores assigned to terminal branch nodes.	159
Figure 3.3-3: Airway radii distribution and constriction as a function of ΔHU	160
Figure 3.3-4: Histogram of ΔHU distribution for all and SO1 proximal voxels across all subjects.....	161
Figure 3.3-5: Branch assigned ΔHU visualisations.	162
Figure 3.3-6: FOT resistance correlations across varying constriction intensities and SO ranges.	163
Figure 3.3-7: Scatter and Bland-Altman plots for selected measured and simulated R5- R20 distributions.....	164
Figure 3.3-8: Heat maps of terminal branch radii across all subjects.....	165
Figure 3.3-9: Heat maps of terminal branch assigned ΔHU across all subjects.	166
Figure 3.3-10: Histograms of ΔHU over all segmented lung volume voxels across all subjects.....	167
Figure 3.3-11: Histograms of ΔHU over all SO1 proximal voxels across all subjects.	168

Abbreviations

$\overline{\Delta HU}^{IS*}$	Average change in mean ΔHU in the inferior-superior direction
1KGP	1000 Genomes Project
ACOS	Asthma COPD overlap syndrome
ACQ	Asthma control questionnaire
AHR	Airway hyper responsiveness
AirPROM	Airway disease predicting outcome in patient specific computer models
ANOVA	Analysis of variance
AQLQ	Asthma quality of life questionnaire
ATS	American thoracic society
AX	Area of reactance
Bash	Bourne again shell
BD	Bronchodilator
BMI	Body mass index
BRC	Biomedical research centre
BT	Bronchial thermoplasty
CBF	Cerebral blood flow
CBV	Cerebral blood volume
CDI	Convection-dependent inhomogeneity
CF	Cystic fibrosis
CFD	Computational fluid dynamics
Chaste	Cancer heart and soft-tissue environment
CMI	Centre for molecular imaging (University of Michigan, USA)
COPD	Chronic obstructive pulmonary disease
CT	Computed tomography
CXR	Chest x-ray
DCDI	Diffusion- and convection- dependent inhomogeneity

DICOM	Digital imaging and communications in medicine
DNA	Deoxyribonucleic acid
DSC	Dynamic susceptibility-weighted contrast
EBUS	Endobronchial ultrasound bronchoscopy
eHU	HU at expiration (FRC)
ERS	European respiratory society
FDG	Fluorodeoxyglucose
FEV	Forced expiratory volume
FEV₁	Forced expiratory volume in 1 second
FOT	Forced oscillation technique
FRC	Functional residual capacity
F_{res}	Resonant frequency
fSAD	Functional small airways disease
FVC	Forced vital capacity
GINA	Global initiative for asthma
GOLD	Global initiative for COPD
HP	Hyperpolarised
HRCT	High resolution computed tomography
HU	Hounsfield unit
I2E/E2I	Inspiration-to-expiration / expiration-to-inspiration
ICS	Inhaled corticosteroids
IgE	Immunoglobulin E
iHU	HU at inspiration (TLC)
IOS	Impulse oscillometry
IVC	Inspiratory vital capacity
JDH	Joint density histogram
K_{co}	Carbon monoxide transfer coefficient
k-NN	k-nearest neighbours

LABA	Long-acting beta-agonist
LCA	Latent class analysis
LCI	Lung clearance index
LDA	Linear discriminant analysis (Fisher)
LLL	Left lower lobe (lung)
LNN	Lower limit of normal
LUL	Left upper lobe (lung)
Matlab	Matrix laboratory (programming language / software)
MBW	Multiple breath washout
MRI	Magnetic resonance imaging
NaN	Not a number
NELSON	Dutch-Belgian lung cancer screening trial
OCT	Optical coherence tomography
PC₂₀	Provocative concentration of methacholine causing 20% fall in FEV ₁
PCA	Principal component analysis
PEF	Peak expiratory flow
PET	Positron emission tomography
PExA	Particles in exhaled air
PFT	Pulmonary function test
PIF	Peak inspiratory flow
PRM	Parametric response map
QC/QA	Quality control / quality analysis
R20	Resistance at 20Hz
R5	Resistance at 5Hz
R5-R20	Small airway resistance (IOS)
RF	Radiofrequency
RLL	Right lower lobe (lung)
RML	Right middle lobe (lung)

RNA	Ribonucleic acid
RNG	Random number generator
RUL	Right upper lobe (lung)
RV	Residual volume
SAA	Stratified axial analysis
S_{acin}	Acinar regional VH index
SAD	Small airways disease
S_{cond}	Conductive regional VH index
SDK	Software development kit
SEM	Structural equation modelling
SF₆	Sulphur hexafluoride
SO	Strahler order
SO1	Strahler order 1 (terminal airway branch)
SPECT	Single photon emission computed tomography
STL	Stereolithography
TDA	Topological data analysis
TLC	Total lung capacity
V_A	Alveolar volume
VC	Vital capacity
VH	Ventilation heterogeneity
VOI	Volume of interest
Yacta	Yet another CT analyser (medical image analysis software)
ΔHU	HU change between expiration and inspiration (eHU-iHU)
χ	Euler characteristic

1 Introduction

1.1 Overview

Asthma is a complex chronic heterogeneous (1,2) respiratory disease that effects more than 1 in 10 people in the UK (3) and over 300 million worldwide (4), has increased in prevalence across the western world in recent decades (5–7), and was estimated to cost the UK over £1 billion and was the cause of over a thousand deaths in 2016 (8). Severe asthma in particular is responsible for a disproportionate amount of this burden (9).

Asthma may not be a leading cause of death globally (10), but has a significant impact on quality of life for both patients and carers, which appears largely due to poor control (11–13). There are strong economical and ethical motivations to improve understanding of asthma (14,15), to contribute towards better disease control and patient outcome.

Ventilation heterogeneity (VH) has been established as a fundamental characteristic feature in subjects with mild to severe asthma (16), predicts asthma control (17), and has been linked to airway hyper-responsiveness (18). Inert gas washout testing, using the multiple breath washout (MBW) technique, was developed on theories of gas mixing and ventilation distribution in the lung (19), and implemented in clinic to measure VH at the mouth (20). Impulse oscillometry (IOS), based on the forced oscillation technique (FOT), has been developed to provide a relatively non-invasive method to measure lung mechanics (21,22), is considered to be particularly useful for application in subjects with limited cooperation such as children (23), and has been validated as a small airways dysfunction detection tool in asthma (24).

A highly active and rapidly developing area of research for the study of asthma, and in particular the spatial distribution of ventilation in the lungs, is the application of numerical methods to data arising from 3D imaging of the thoracic region (25,26). Hyperpolarised 3-helium (^3He) magnetic response imaging (MRI) has been used to study the difference in VH between asthmatic and healthy subjects (27), and computed tomography (CT), typically providing higher resolution data, has been applied in the assessment of airway wall thickness (28), and with segmentation can be used to extract lung structures (29) and study density differences between inspiration and expiration

lung states (30). The parametric response map (PRM) was developed to assess the regional changes in a volume of interest (VOI) as it geometrically deforms between two or more states (31), and has been applied to study density changes between lungs at total lung capacity (TLC) and functional residual capacity (FRC), states (32). A comparison between CT and MRI based ventilation metrics has been performed (33).

The human conducting airway tree consists of approximately 16-17 generations of branches from the trachea, and accurate segmentation of this structure from CT data has been a long studied challenge (34). Accurate segmentation of airways from CT images is typically limited to the first 7-8 generations at most, that is between 2^7 and 2^8 branches (35,36). Recently an algorithm to simulate the complete 16-17 generation airway tree, given an initial segmentation and lobar surface stereolithography (STL), has been developed (37) and used for modelling of FOT VH metrics in asthma (38).

The analysis of data arising from imaging features is primarily conducted using the traditional statistics of group comparisons, discrimination and predictability. Clustering has been applied to effect phenotyping (39,40), inspired by the systems medicine approach (41,42). The strikingly rapid development of computing devices and the capacity for data collection and storage over recent decades has led to the modern problem of data mining on ‘big data’ (43) and expectation of a central role in many areas of society including medicine (44). In many areas of research data sets are becoming more innately high dimensional with growing numbers of features and records, leading to a relatively alien space for traditional statistical methods (45).

Principal component analysis (PCA) was developed by Karl Pearson around 1900 and has been widely used to extract information from high dimensional data sets, including phenotyping asthma (40); however PCA can be very sensitive to outliers (46) and information may be lost in the projection from a higher dimension (47). Topological data analysis (TDA) has been proposed as one method for robustly visualising data with high dimensional properties (48,49), to draw qualitative insights from looking at the shape of the underlying data object. TDA has been successfully applied to phenotype type-2 diabetes (50), breast cancer (51) and asthma (52).

In this thesis, statistical methods, computational algorithms and TDA were developed and applied to PRM registered CT imaging data sets, with lobar and airway tree segmentation; this was to visualise the lungs and lung function, evaluate novel CT

biomarkers for association with physiological measurements from multiple length scales (see **Figure 1-1**), including MBW and IOS derived VH markers, and support FOT modelling of small airway VH in adult asthma. This introduction will briefly cover the background of research areas from both the biomedical and mathematical literature that play a part in the story leading up to this project, culminating in a statement of the aims and objectives, and description of the thesis structure.

Figure 1-1: Sources of data analysed in this thesis from multiple length scales.

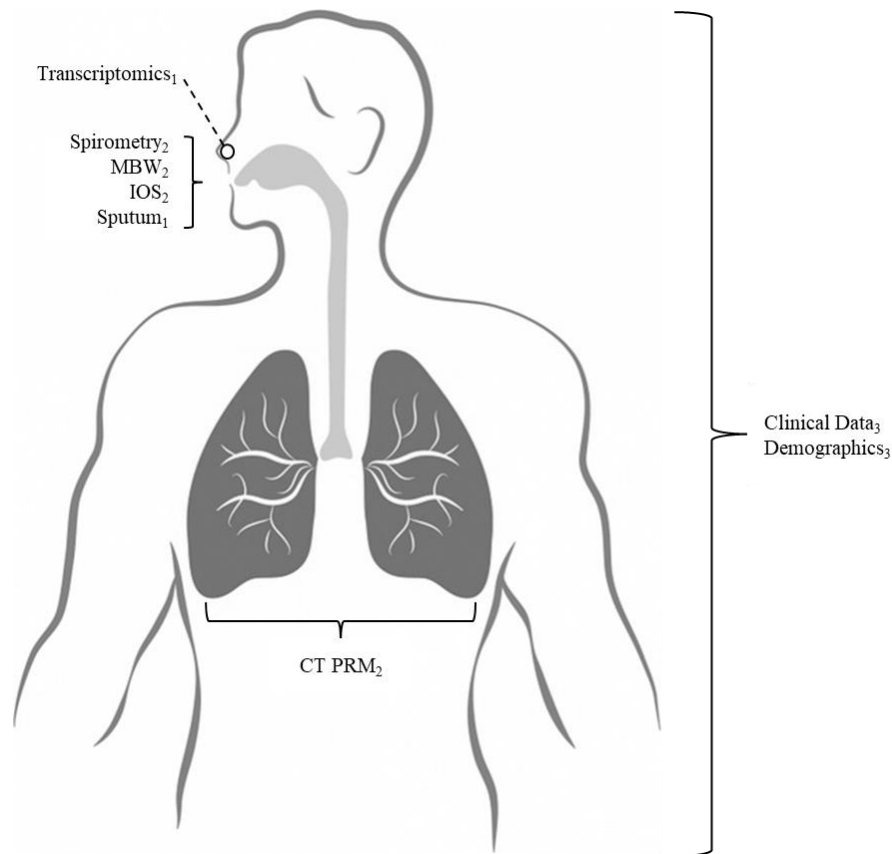


Illustration of physical sources of data measured for analysis in this thesis. Lung function testing with spirometry, impulse oscillometry (IOS) and multiple breath washout (MBW) take measurements using air flow, sound waves and gas concentration respectively at the mouth. Sputum is collected through induced expulsion via the mouth. CT imaging uses radiation to obtain three dimensional density based images of the lungs; image registration and the parametric response map (PRM) provides functional information from paired inspiratory-expiratory CT image sets. Transcriptomic data is collected through scraping of the nasal passage. Data on the whole organism is collected through questionnaires and other clinical procedures. Subscripts refer to apparent length scale: 1, cell / gene (microscopic); 2, organ; 3, organism.

1.2 Asthma

This section discusses the definition of ‘asthma’ with complications arising therein, some of the common classification schemes used, and how ventilation heterogeneity has been understood to play a significant role in the disease pathogenesis.

1.2.1 Definition

The Global Initiative for Asthma (GINA), formed in 1993, provided a definition for ‘asthma’ in a 2006 report (53) as follows:

“Asthma is a chronic inflammatory disorder of the airways in which many cells and cellular elements play a role. The chronic inflammation is associated with airway hyper-responsiveness that leads to recurrent episodes of wheezing, breathlessness, chest tightness, and coughing, particularly at night or in the early morning. These episodes are usually associated with widespread, but variable, airflow obstruction within the lung that is often reversible either spontaneously or with treatment.”

This definition suffices to convey the most common observations that one associates with the disease, such as airway hyper responsiveness, unbalanced inflammatory response; however due to the apparent complexity and the ever evolving understanding of disease heterogeneity (54,55), it is a definition subject to debate and revision (56). Clinician diagnosis of asthma is based on patient history of symptoms including recurrent wheezing, breathing difficulty and chest tightness, and demonstration of reversible airflow obstruction measured using spirometry; a common procedure being observation of the increase in forced expiratory volume in 1 second (FEV₁), using spirometry, in response to use of short-acting β_2 -agonist (57).

1.2.2 Classification

Patients with asthma can be classified by severity of the disease, which is typically measured using GINA step, a treatment based scheme that places a subject into one of five categories depending on the intensity of treatment required to control symptoms (58). A similar scheme is used for chronic obstructive pulmonary disease (COPD), GOLD (global initiative for COPD) (59). The situation is complicated by the potential for two or more distinct classifications to exist for a given subject, as observed with

asthma COPD overlap syndrome (ACOS), where subjects present with symptoms of both diseases (60).

Allergy is recognised to have a close relationship with asthma (61) leading to classifications of allergic and non-allergic asthma (62). Immunoglobulin E (IgE) are antibodies produced by the immune system and play a central role in defining allergic asthma (63). Atopy is closely related allergy, though specifically refers to a genetically inherited disposition to higher levels of IgE (64) and is argued not to play a dominant role in the disease (65). Thus there also exist atopic and non-atopic classifications. Inflammation of the airways is a key feature of asthma (66), and is associated with white blood cells called eosinophils (stained by eosin) and the more abundant neutrophils, which have been studied in relation to asthma and COPD (67–69). High eosinophil levels are now considered indicative of severe asthma and the term ‘eosinophilic asthma’ has been coined for this state (70). High neutrophil counts are reported to be more predominant in COPD relative to asthma (71).

To classify level of asthma control and impact on quality of life, questionnaires have been developed and implemented in the clinical setting. Prominent standard questionnaires widely used in the medical sciences were published by Juniper et al, in 1992 presenting the asthma quality of life questionnaire (AQLQ) (72), and later in 1999 presenting the asthma control questionnaire (ACQ) (73). A standardised version of the AQLQ, used in studies presented in this thesis, was also published and validated in a study by Juniper et al in 1999 (74). A shortened version of the ACQ taken with omission of FEV₁, ACQ-6 (6 items), was used in the studies presented (75).

1.2.3 Nasal Transcriptomics

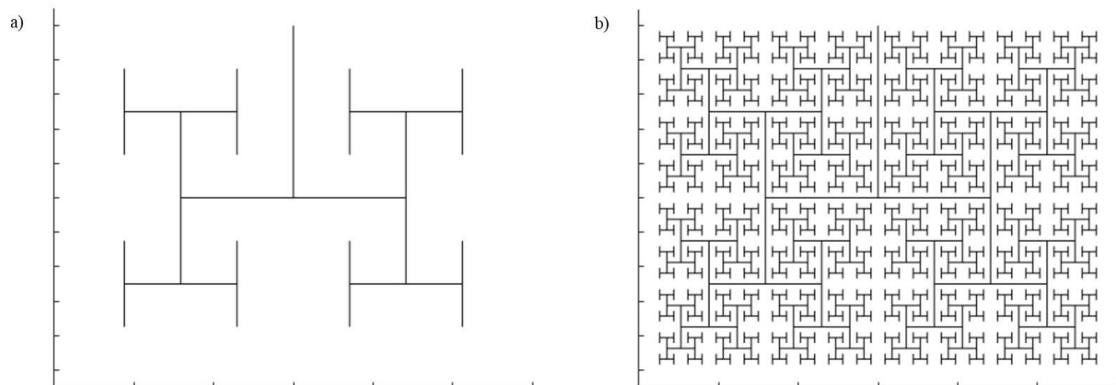
Airway inflammation and excessive mucus production are features in asthma governed by an inflammatory response involving epithelial cells lining the airways, often in reaction to some airborne stimuli e.g. dust mite droppings, pet dander and pollen. The transcripts of these cells contain information determining the nature of this response, thus it is of interest to examine transcript gene expression levels to identify association with complex disease traits (76). Obtaining cells from the epithelium of the small airways is impractical for large scale in vivo studies; fortunately it has been demonstrated that nasal epithelial cells may be suitable surrogates for analysis (77), and are much easier to access for sampling in the clinical setting. Gene expression data,

analysed in the studies presented in this thesis, was obtained using Affymetrix GeneChip array technology (78), widely used in the medical sciences for measurement of gene expression levels using photolithographic techniques (79).

1.2.4 Lung Structural Elements

The anatomical structure and physiological properties of the lung are key factors in determining pulmonary ventilation / perfusion spatial distributions and lung function, and understanding malfunction that arise and develop over time due to disease states. Ewald Weibel, from the University of Bern in Switzerland, published an excellent review article on the '*morphometric basis of lung function*' entitled 'What makes a good lung?', in 2009 (80); he illustrates in great detail the evolutionary design mechanisms of the human lung which provides maximal surface area within a fixed volume and dichotomous branching structure that consequentially mimics mathematical properties of the 'H tree' from fractal geometry, see **Figure 1-2**.

Figure 1-2: H-tree plotting after four and ten iterations from the root branch.



Plots produced using custom Matlab code of the H-tree using a starting branch directed from the origin to (0,-1) and using the standard $1/\sqrt{2}$ scale factor to determine new branch length on iterating orthogonal splitting of leading branches. A, fractal pattern after 4 iterations. B, fractal pattern after 10 iterations, note the effective covering of space contained within a bounded area.

Strahler order (SO) is a simple and widely used concept for evaluating tree complexity and characterisation of branches, developed by geoscientists Strahler and Horton in works published about the mid 20th century studying stream erosion and topology (81,82). First terminal branches are assigned an order of 1. Then branches are assigned

an order of $n+1$ where n is the highest order of any connected branch; this repeats until all branches have an order assigned to them. The result characterises the complexity of a tree by the SO assigned to the root, and for any branch the designated SO effectively conveys maximal distance from the leaves of the tree; see **Figure 1-3**.

Figure 1-3: Strahler order assignment to an exemplar tree structure.

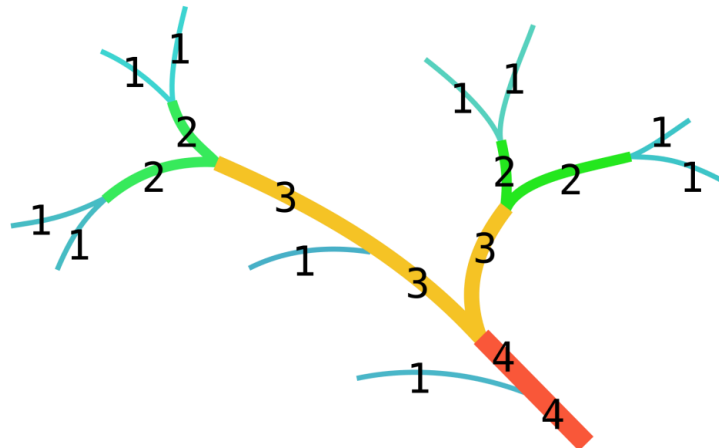


Illustration of Strahler order (SO) labelling of a simple branching tree structure.

Terminal branches (blue), often called leaves, are defined as those with no branch connected to their distal end. SO assignment begins from the leaves and characterises a branch by the maximal distance of that branch from at least one leaf. Tree complexity is characterised by the largest assigned SO (observed at the root). Image reused from: https://en.wikipedia.org/wiki/Arthur_Newell_Strahler, under creative commons license CC BY-SA 3.0: <https://creativecommons.org/licenses/by-sa/3.0/>.

Gas exchange within the healthy human lung is optimised through the spatiotemporal alignment of ventilation and perfusion facilitated by the lung structure, locally optimising the ratio V_a/Q of alveolar ventilation (V_a) to blood flow (Q) (83) in acinar units that tend to be ventilated in series and perfused in parallel; in disease this ratio can be disrupted through airway remodelling and obstruction (84,85), leading to local hypoxaemia (oxygen deficiency) and subsequent tissue damage.

Natural spatial heterogeneity in V_a and Q across the lung exists due to gravitational field effects (86), lung elasticity (87), pleural pressure gradients (88) and posture (89,90). Heterogeneity of ventilation and perfusion in the lung have been studied with respect to posture and gravitational effects using magnetic resonance imaging (MRI) techniques

(91), showing V_a/Q to be more spatially homogenous in the prone position relative to supine; in the latter case, typical of a subject undergoing a CT scan, ventilation and perfusion was markedly higher in the posterior and inferior regions of the lungs.

1.2.5 Ventilation Heterogeneity

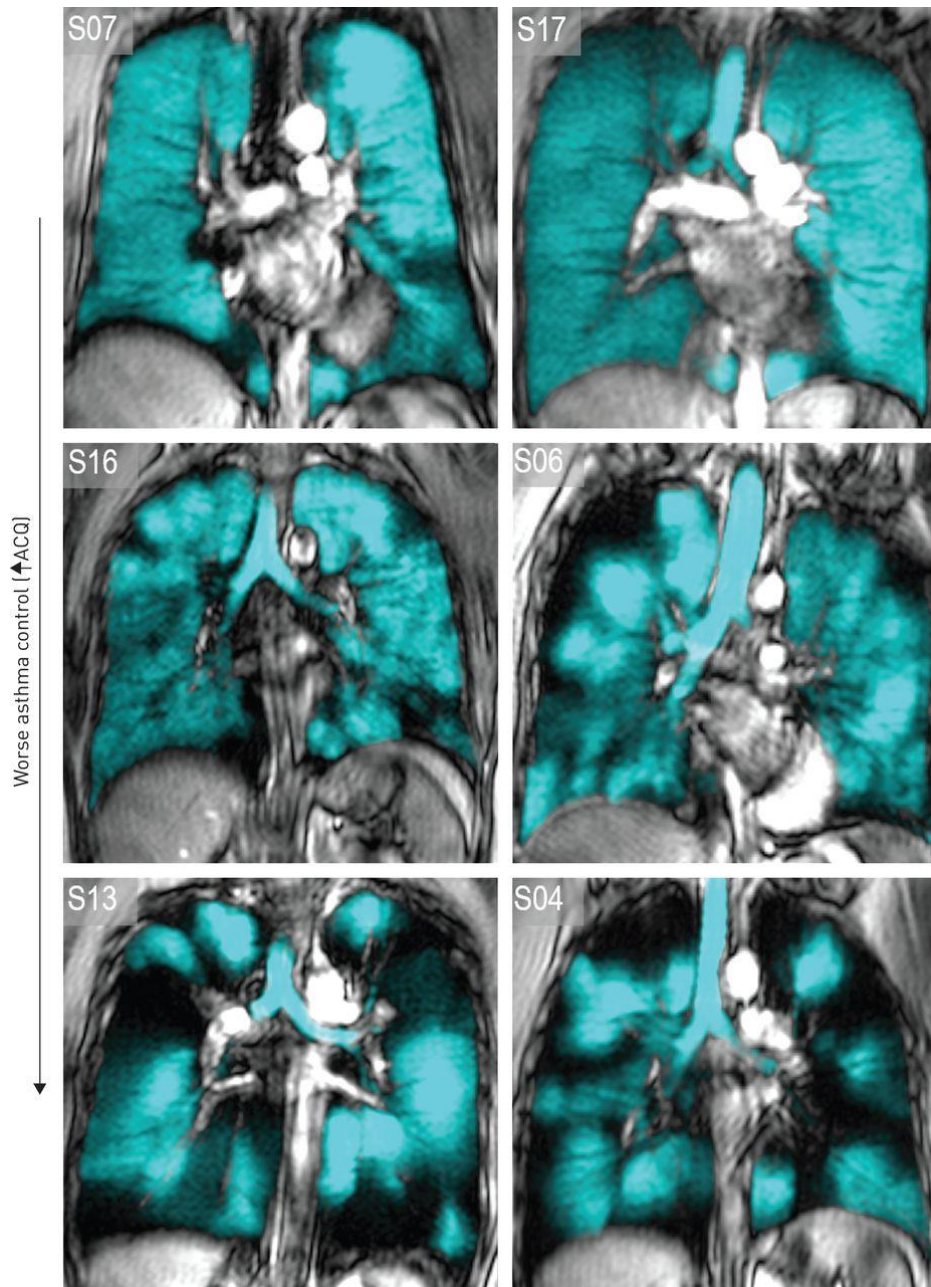
Gas exchange in the lungs requires the transport of air through the airway tree network, primarily via conduction for the first 14-15 generations from the trachea, then transitioning gradually into diffusion mechanics in the smaller airways thereafter (80). A study of convection-dependent inhomogeneity (CDI) and diffusion- and convection-dependent inhomogeneity (DCDI) within the human lung was published by Crawford et al in 1985 (92), followed by a landmark review by the co-authors, Paiva and Engel, on the theory of gas mixing and ventilation distribution in the lung in 1987 (19); this work led to the widespread use of the multiple breath washout (MBW) technique for global VH quantification at the mouth, described in section 1.3.3., where indices of conductive and acinar VH, S_{cond} and S_{acin} , are derived from CDI and DCDI respectively (16).

VH has been linked to airway hyper responsiveness (AHR), a fundamental trait of asthma typically measured by response in lung function performance to some challenge (e.g. methacholine inhalation, exercise, cold air), where S_{cond} was shown to predict AHR features (93). The effects of age on MBW characterised VH were reported in a study by Verbanck et al (94), showing a steady increase in VH through increasing MBW markers S_{acin} , S_{cond} and lung clearance index (LCI). VH has been studied in COPD, where S_{acin} was shown to correlate with increasing disease severity and related to impulse oscillometry (IOS) measurements (95); IOS is another popular technique for VH quantification (see section 1.3.4). Evidence was presented in a 2013 study suggesting VH is more closely associated with AHR in asthma relative to COPD (96).

Significant abnormal VH in the lungs has been observed in mild and severe asthma (16), first witnessed with imaging using inhaled ^{133}Xe (97), xenon-133, and more recently with ^3He MRI (27). In the healthy human lung, in an orthostatic posture, gravity induces a pleural pressure gradient that causes the lungs to be most ventilated in the inferior regions (98,99). This gradient of ventilation presents some natural heterogeneity of ventilation in the human lung, which can alter based on posture (91), due to the direction of gravity and lung tissue compression (87). In patients with at least mild asthma, the structure of the airways is perturbed through airway narrowing from

chronic inflammation (100) and remodelling (85,101,102) resulting in a significantly increased non-uniform heterogeneity in the spatial distribution of ventilation. A 2016 study linked increased VH observed in 3-He MRI to worsening asthma control (ACQ) (103), see **Figure 1-4**, where significant VH is apparent in poorly controlled asthmatics.

Figure 1-4: Increasing ventilation heterogeneity (VH) with decreasing asthma control.



Reproduced with permission of the © ERS 2019; European Respiratory Journal 48 (2) 370-379; DOI: 10.1183/13993003.00393-2016 Published 31 July 2016. Hyperpolarised 3He MRI illustrating increasing heterogeneity in ventilation spatial distribution in selected subjects with decreasing asthma control quantified by ACQ.

1.3 Lung Function Testing

This section introduces methods for lung volume and resistance measurement using body plethysmography, lung function measurement using spirometry, and two particular methods proposed and implemented in clinical research for the measuring of V_H in the lungs. N.B. lung function testing may also be referred to as pulmonary function testing with the common acronym PFT.

1.3.1 Body Plethysmography

Body plethysmography is a well-established PFT technique based on ideas described from the late 19th century and technically developed in the mid-20th century (104,105). The method relies on Boyles law (106), stating the relationship between pressure P and volume V of an ideal gas under isothermal conditions (constant temperature) as $PV=k$ for some constant k ; for the same substance at different times: $P_1V_1=P_2V_2$. This can be used to determine lung volume given known volume and pressure of a sealed box in which a subject is contained and respires, initial conditions of the box, and box pressure measured post full expiration. Thoracic gas compression and airway resistance can be measured using a volume-displacement plethysmograph (107,108); see **Figure 1-5**.

Figure 1-5: Body plethysmography box and volume-displacement diagram.

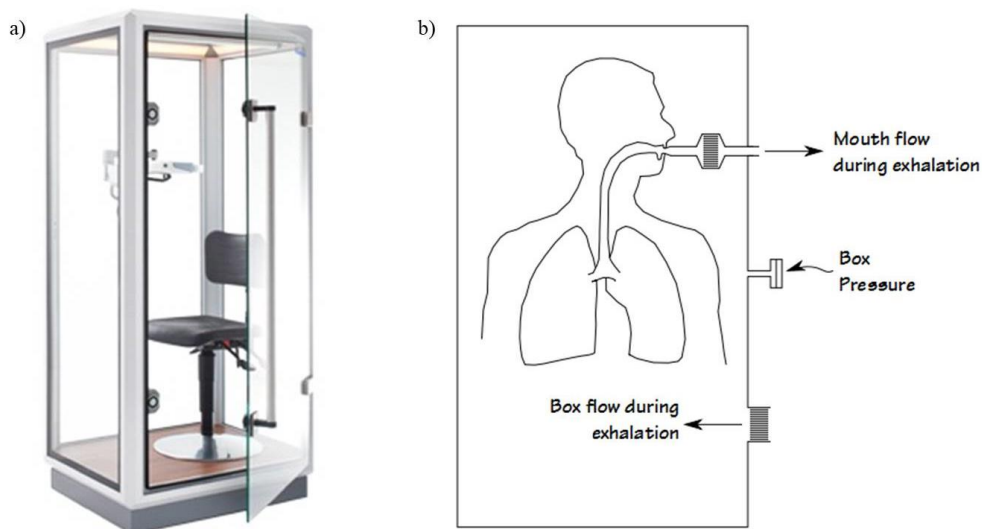


Illustration of body plethysmography equipment; body box in which a subject sits (a), and volume-displacement setup (b). Image (a) reused from: <https://en.wikipedia.org/wiki/Plethysmograph>, under creative commons license CC BY-SA 3.0: <https://creativecommons.org/licenses/by-sa/3.0/>. Image (b) reused from <https://www.pftforum.com/blog/measuring-thoracic-gas-compression/>, under creative commons license CC BY-NC 4.0: <https://creativecommons.org/licenses/by-nc/4.0/>.

1.3.2 Spirometry

The most common form of PFT is spirometry (109,110), which monitors the volume / flow rate of air, through a mouthpiece a subject breathes through, in time as they respire. The subject will typically breath normally, then by instruction breathe in as far as they can, then breathe out as fast as they can (forced expiration). It is the primary tool used in the diagnosis and management of asthma, and is typically assessed using the resultant flow-volume loop, see **Figure 1-6**, and volume-time curve, see **Figure 1-7**.

Figure 1-6: Flow-volume loop from spirometry.

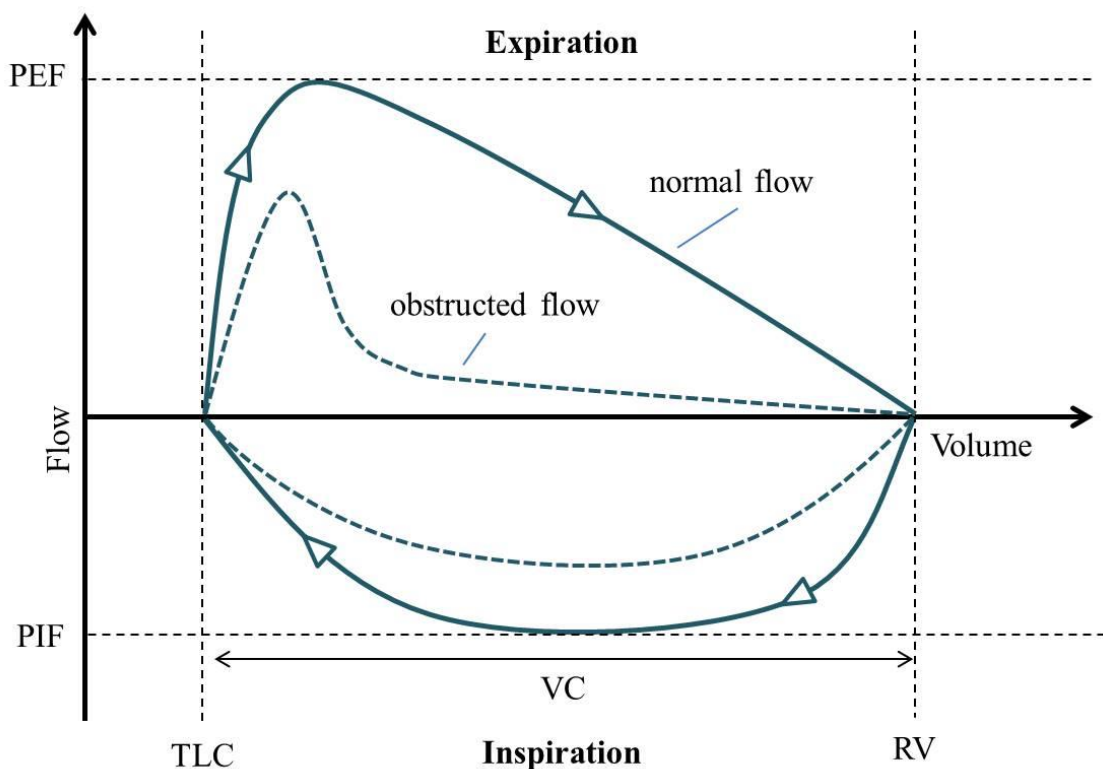
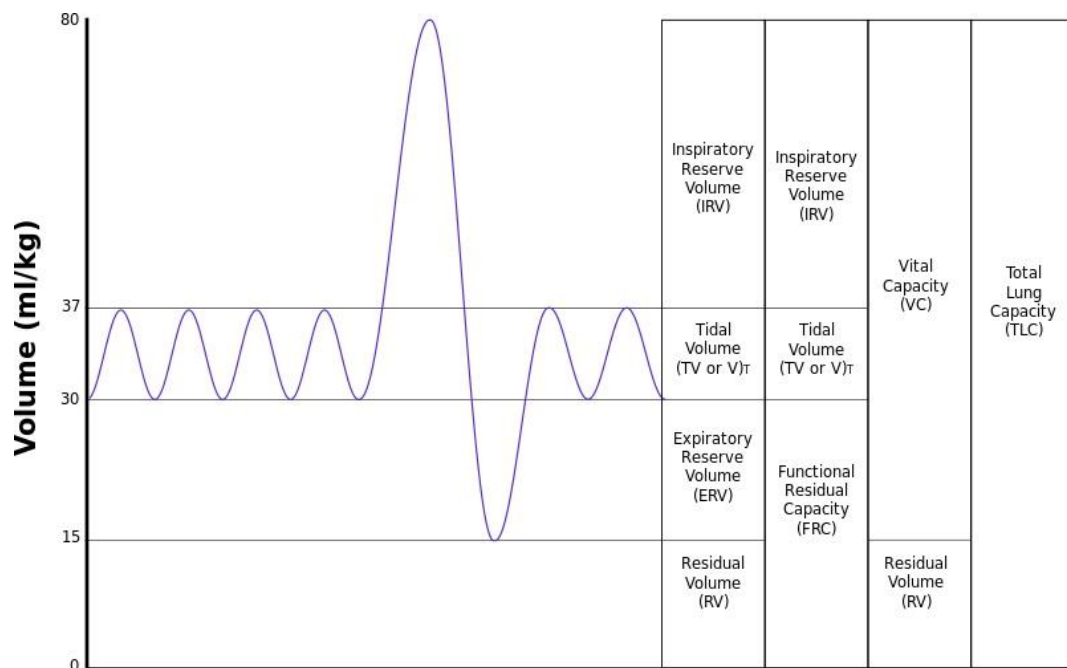


Illustration of typical flow-volume loop arising from one forced expiration manoeuvre during spirometry testing. PEF, peak expiratory flow; PIF, peak inspiratory flow; TLC, total lung capacity; VC vital capacity; RV, residual volume. Subject expires above the horizontal (positive flow rate) and inspires below the horizontal (negative flow rate). Profile of typical normal (solid) and obstructed (dashed) flow-volume loops demonstrated, illustrating shape change due to airway obstruction.

There are a number of significant features arising from this procedure to be aware of as they occur commonly in respiratory data. Forced vital capacity, FVC, reports the total volume of air expired forcefully in one breath; Forced expiratory volume in one second,

FEV₁, is the volume forcefully expired within the first second of the described ‘FVC manoeuvre’; the peak expiratory flow, PEF, is determined as the maximum flow rate achieved through expiration; the peak inspiratory flow, PIF, is the maximum flow rate achieved through inspiration; inspiratory vital capacity, IVC, is the total amount of air that can be inhaled after complete expiration (111).

Figure 1-7: Volume-time graph from spirometry.



*Illustration of typical volume-time curve arising from spirometry, during tidal breathing and performance of one forced expiration manoeuvre (peak followed by trough). Useful reference for respiratory volumes, compare with **Figure 1-6**. Image reused from: <https://en.wikipedia.org/wiki/Spirometry>, under creative commons license CC BY-SA 3.0: <https://creativecommons.org/licenses/by-sa/3.0/>.*

Expiratory flow limitation and in particular FEV₁ reduction is associated with chronic airway diseases such as asthma (112). Equations modelling expiratory air flow have been developed to describe the limiting mechanisms (113,114). FEV₁/FVC is a widely used ratio, expressing the relative proportion of FVC expired in 1 second, to define airway obstruction from spirometry (115).

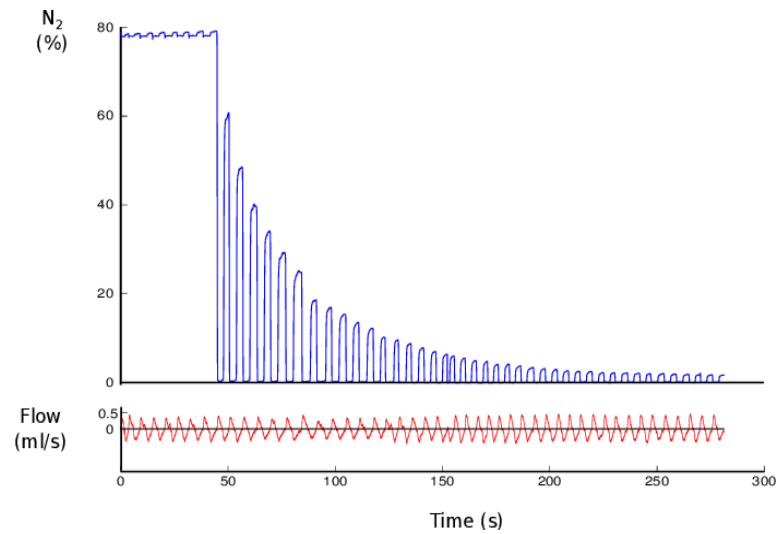
1.3.3 Multiple Breath Washout

MBW using the inert gas Nitrogen (N_2) was described in 1950 by Robertson et al (116) based upon principles of gas mixing and ventilation perfusion matching in the lung. A machine monitors tracer gas concentration as it washes out of a patient's lungs through tidal breathing of pure oxygen (117); besides nitrogen (N_2), helium (He) and sulphur hexafluoride (SF_6) have been studied, the latter described in 1985 (118), requiring an initial 'wash-in' phase of the tracer gas, not required for N_2 since this occurs naturally in air. Differences between N_2 and SF_6 have been reported (119). MBW has been developed alongside technological advances to provide a method for evaluating VH in clinical research, and has demonstrated improved sensitivity over spirometry for early detection of some pulmonary diseases such as cystic fibrosis (CF) (120). Recommendations on standards for performance of MBW testing are published (20).

An MBW test performed during tidal breathing produces a 'washout curve', see **Figure 1-8**. Each time the subject exhales, three phases are evident, see **Figure 1-9**, in which the gas concentration increases, reaches an angled plateau which persists along some gradient until full expiration and the subject must inspire again. Computational isolation and analysis of the phase three slopes is performed, and from this biomarkers of VH are calculated. Automated algorithms have been designed and implemented to calculate VH markers from the washout curve data, for example a study published in 2011 described an approach using 'segmented linear regression' or 'breakpoint' analysis, in Matlab, to identify the start of the alveolar plateau (121); comparison with manual measurement showed no statistically significant differences for studied markers.

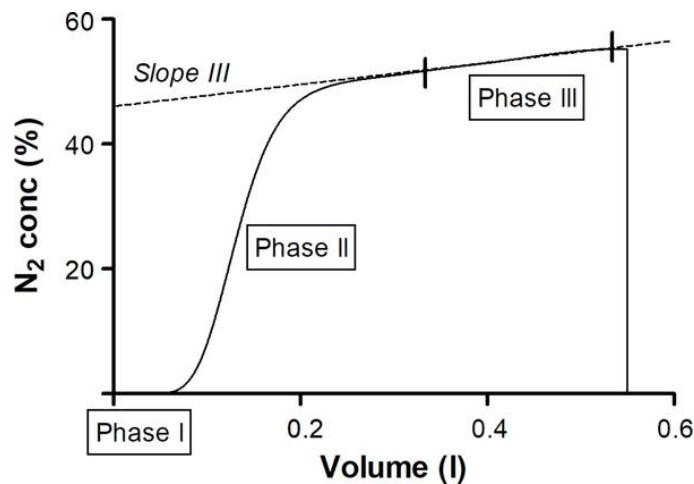
Lung clearance index (LCI) is essentially the number of tidal breaths required before the gas concentration falls below some threshold; specifically it has been calculated as the net expired volume up to the point the gas concentration first falls below $1/40^{\text{th}}$ of the starting concentration, divided by the patient's FRC (122). Conductive (S_{cond}) and acinar (S_{acin}) regional VH markers are calculated, as previously described by Verbanck et al (123,124), from the phase three slopes; linear regression is applied to determine gradients of each phase three slope, followed by normalisation by the mean expired gas concentration, then these normalised gradients are plotted against turnover. S_{cond} is defined as the gradient change between the 1.5^{th} and 6^{th} turnover, and S_{acin} is the gradient of the first breath minus the contribution due to S_{cond} .

Figure 1-8: Washout curve with synced flow-time graph from MBW.



Typical washout curve showing fall in inert gas concentration, nitrogen (N_2) in this case, with flow-time curve beneath synchronised in time showing subject tidal breathing. Up to about the 50 second mark the subject is breathing ambient air (78% N_2), then the gas supply is switched to pure O_2 and the steady fall in N_2 % is observed. Figure reused from: [arXiv:1902.09026v1](https://arxiv.org/abs/1902.09026v1), with permission from the author; (125).

Figure 1-9: Phase III slope analysis of washout curve segment from MBW.



*Illustration of phase III slope analysis including three defined phases associated with plateaus occurring in the typical washout curve (compare to **Figure 1-8**). Phase III is defined between points at which 50% and 95% of volume is expired (annotated vertical lines). Image reused from: <https://doi.org/10.1371/journal.pcbi.1007079> (126), under creative commons license CC BY-SA 4.0: <https://creativecommons.org/licenses/by/4.0>.*

1.3.4 Impulse Oscillometry

IOS is a method based on FOT described by DuBois et al in 1956, modified by Michaelson et al in 1975 (127,128), in which sound waves are superimposed onto tidal breathing through the use of mouthpiece connected to a loudspeaker, or piston, emitting forced oscillations over a range of frequencies (129); see **Figure 1-10** for illustration of typical components and configuration used in IOS testing. The relatively non-invasive nature of IOS, where a subject need only breathe normally whilst their cheeks are held to limit shunting effects, makes it particularly suitable for study in difficult adults and children where cooperation may be limited (23,130–132). Airway impedance (Z) is assessed at the mouth, as a function of pressure (P) and flow (V'), by applying Fourier transform to the P/V waves in time to map to the frequency space, i.e. P/V' as a function of wave frequency; resistance (R), describing mechanical properties of the respiratory system, is calculated as the real part of Z , and reactance (X), relating to energy storage and capacity, as the imaginary part (133).

Figure 1-10: Equipment diagram for impulse oscillometry.

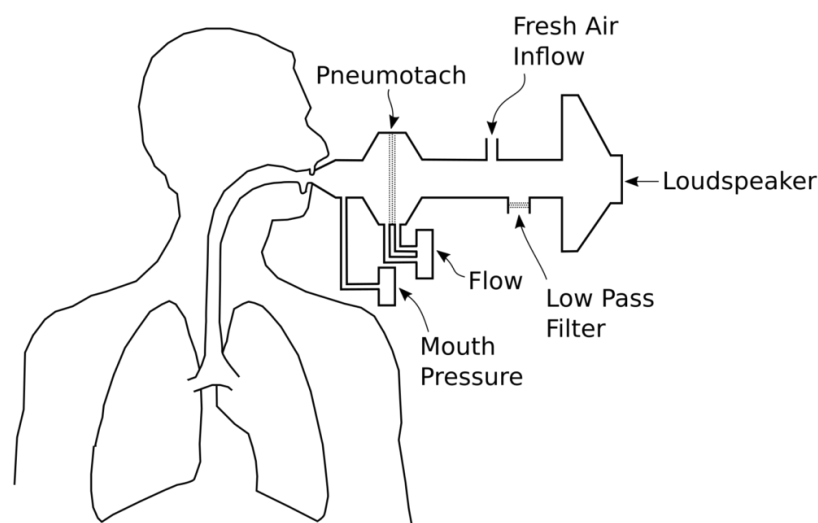
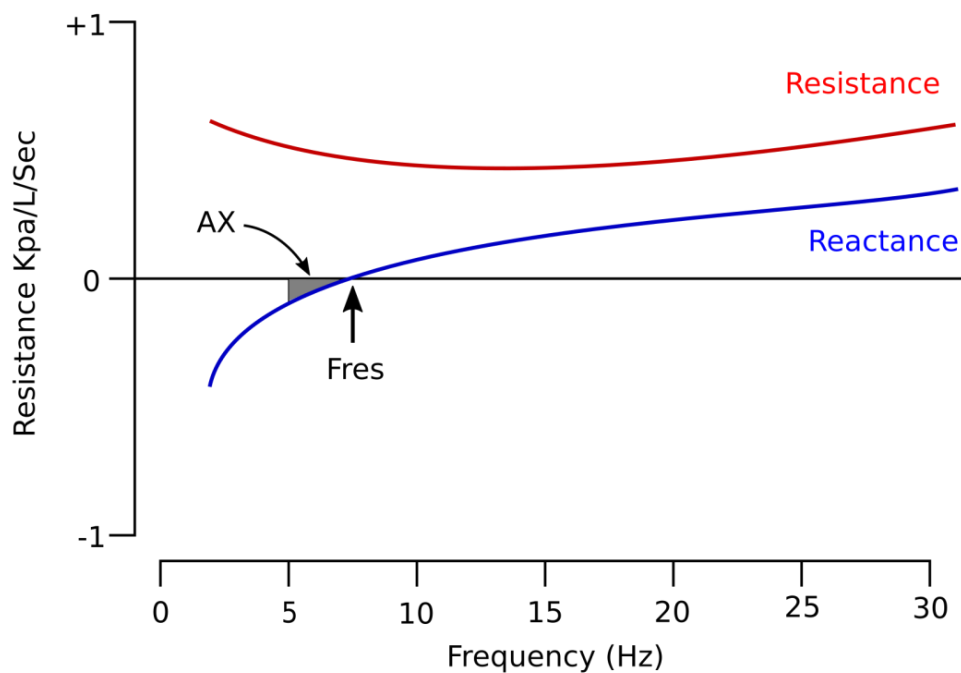


Illustration of equipment and setup for IOS measurement. Key measurements of flow and pressure are taken at the mouth. Air flows through a pneumotach, a light screen producing a known drop in pressure in direct proportion to air velocity. Oscillatory pressure is generated by loudspeaker. A low pass filter impedes high frequency signals to prevent waveform diversion / blunting. Image reused with permission (author Richard Johnston) from: <https://www.pftforum.com/blog/oscillometry/>, under creative commons license CC BY-NC 4.0: <https://creativecommons.org/licenses/by-nc/4.0/>.

An illustration of plotting resistance and reactance against frequency is given in **Figure 1-11**. Resistance at x Hz (hertz) is denoted R_x , lower frequencies tend to travel further in space, and R_5 and R_{20} are typically used as measures of whole and conductive lung resistance respectively (134); the difference $R_5 - R_{20}$ is thus presented as a measure of small airway resistance (135) and has been validated as a tool for the detection of small airways disease (SAD) (24). IOS derived parameters have been associated with acinar VH through comparison to MBW measures (95). Reactance at y Hz is denoted X_y . F_{res} , resonant frequency, is defined as the frequency at which the reactance curve crosses the frequency axis, i.e. the frequency t where $X_t = 0$. AX , the area of reactance, is defined as the area between the reactance curve and the frequency axis, between X_5 and X_{Fres} .

Figure 1-11: Resistance/reactance curves as a function of frequency.



Plot of a typical airway resistance (red) and reactance (blue) graph produced by IOS testing. The resonant frequency F_{res} is indicated where the reactance curve cuts the frequency axis. The area of reactance AX is indicated as the shaded area above the reactance curve between 5Hz and F_{res} Hz. Image reused with permission (author Richard Johnston) from: <https://www.pftforum.com/blog/oscillometry/>, under creative commons license CC BY-NC 4.0: <https://creativecommons.org/licenses/by-nc/4.0/>.

1.4 Imaging

This section will provide background on the primary source of data analysed in this thesis, computed tomography (CT), alongside closely related imaging methods. The mapping of imaging derived volumes in space through geometric deformation, and application in PRM method, is discussed, and the use of airway segmentation in computational modelling of FOT is described.

1.4.1 Modality Overview

Advancements in mathematics, physics and computing over the past two centuries have seen in their intersection the exponential rise of technologies for the imaging of three-dimensional structures, in particular the visualisation of internal components of such objects, and the clear application to medical science (136).

Wilhelm Roentgen won the Nobel prize in physics in 1901 for discovery of ‘x-rays’ a few years prior (137), perhaps the most widely recognised form of imaging used in medicine. The basic idea is that an object is exposed to a radiation (x-ray) source from one side, whilst on the other a sheet of material that visibly reacts to radiation absorption is positioned. The level of absorption at a point is inversely proportional to the density of matter the ray had to pass through to reach the sheet at that point, thus an image based on the density distribution within the object being imaged is produced. Skeletal structure is most immediately recognisable and culturally associated with the term x-ray since bone has significantly higher density than most other tissues in the human body. The Hounsfield unit (HU) is the most widely used standard for quantifying radiation absorption due to x-ray imaging (138).

Tomography is the imaging of an object by slices, using any kind of penetrating wave. In modern CT scanners the subject lies supine as the radiation source / detector rotates 180 degrees capturing multiple images of density absorption, which then traverses perpendicular to the slice plane (139), see **Figure 1-12**. The problem then is to infer the three dimensional structure of an object, with respect to varying density, given a set of two dimensional projections from varying angles. This ‘reconstruction problem’ has a number of possible solutions, varying in computational cost and accuracy; popular algorithms include Fourier-domain reconstruction, back projection and iterative reconstruction (140,141). The speed at which the scanner moves, both in rotation and

translation, and exposure frequency are significant in determining the resolution properties of resultant images.

Figure 1-12: CT scanner equipment showing significant operating components.

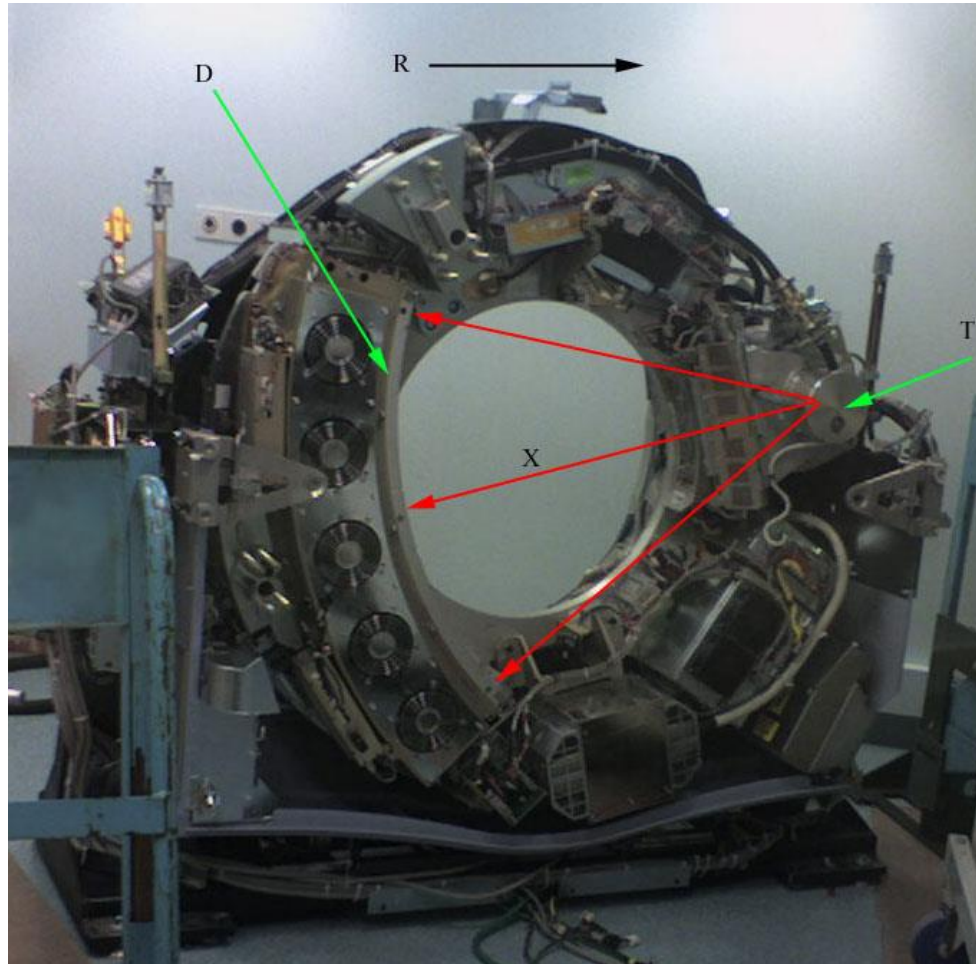


Photo of the internal components of a multi-detector CT scanner. D, x-ray detector array; T, x-ray emission source; X, x-ray beam travel during operation, from the source to the detectors; R, indication of direction in which the setup rotates during scanning. Image reused from: https://en.wikipedia.org/wiki/CT_scan, under creative commons license CC BY-SA 3.0: <https://creativecommons.org/licenses/by-sa/3.0/>.

In the study and treatment of respiratory disease, the two-dimensional x-ray image is focused upon the thoracic section to capture density variation in the lungs and the shape of the airways; this is often referred to simply as a ‘chest x-ray’, or CXR. In asthma a CXR is typically not standard unless complicating comorbidities are suspected (142). CT imaging is more common and recommended as a tool for evaluation, particularly in severe asthma, though it is not currently part of the diagnostic process (143).

Qualitative studies of CT in asthma have typically focused on airway wall thickening and mucous plugging (144–147). Bronchiectasis, a chronic condition where airways become unnaturally enlarged coupled with mucous pooling, has also been studied in CT imaging (148,149), and associates with disease duration (150). Gas trapping is a significant feature of respiratory disease, and appears commonly as regions of relatively low density on images taken at expiration, and has been studied in association with asthma (151). Treatment response due to inhaled corticosteroids (ICS) has been evaluated in a number of studies (152,153), and comparisons with extra-fine ICS have been made (154). Significant reductions in airway wall area due to inflammation targeting treatment, evaluated using CT, were reported by Haldar et al (70), in refractory eosinophilic asthma.

Magnetic resonance imaging (MRI), formerly nuclear MRI (155), is a technique that was described independently by Lauterbur and Mansfield in 1974 (156), based on earlier works by celebrated physicists Bloch and Purcell (157,158). MRI scanners induce a magnetic field using powerful magnets in an outer ring that surrounds the patient, with an intermediate ring of gradient coils and inner ring of radiofrequency (RF) coils (see **Figure 1-13**), which obtains images through observation of the effect the magnetic field has on proton spin in hydrogen atoms, common in water and fat tissues (159). Though historically MRI had initially lagged behind CT in application to pulmonary imaging due to technical and economic factors (160), such as relatively low proton density in lung parenchyma (161), the development of hyperpolarised (HP) gas, in particular helium-3 and xenon-129, methods for direct visualisation of ventilated airspaces in the lungs, has since significantly increased the application of this modality in the study of lung diseases (162–165).

The clinical application of HP gas MRI to severe asthma has been assessed looking at response to bronchial thermoplasty (BT) (166), a therapy that reduces the mass of smooth muscle in airway walls through heat energy applied locally via radiofrequency emission (167,168). Patterns of VH have been studied using HP ³He MRI in relation to eosinophil count from induced sputum in severe asthmatics (169). Most studies in humans have used ³He as the tracer gas for MRI, however advances in the cheaper ¹²⁹Xe gas techniques suggest this may change in the future (170,171).

Additional imaging modalities methodologically distinct from or supplementary to CT or MRI are positron emission tomography (PET), single photon emission computed tomography (SPECT), endobronchial ultrasound bronchoscopy (EBUS) and optical coherence tomography (OCT); the latter two involving more localised imaging captured physically from within the major airways. PET has been applied to study VH in the asthmatic lung through the analysis of self-organised patchiness (172), assess regional pulmonary perfusion and specific ventilation of atopic asthma (173), and study arterial inflammation in asthma using 18F-fluorodeoxyglucose (FDG) tracer (174). SPECT has been utilised in the study of increased eosinophil uptake in obese asthmatic subjects (175), assessment of VH through deposition pattern of inhaled ciclesonide particles in the mild asthmatic lung (176), and validation of computational fluid dynamics (CFD) models developed from CT data of patients with mild asthma (177). EBUS has been compared with high resolution CT (HRCT) in studies demonstrating a comparable ability to evaluate airway wall remodelling and in ovine and human (asthmatic and healthy control) lungs (178,179), and OCT was applied in a cohort of severe asthma patients to safely study the post-treatment acute airway effects of BT (180).

Figure 1-13: MRI scanner equipment diagram.

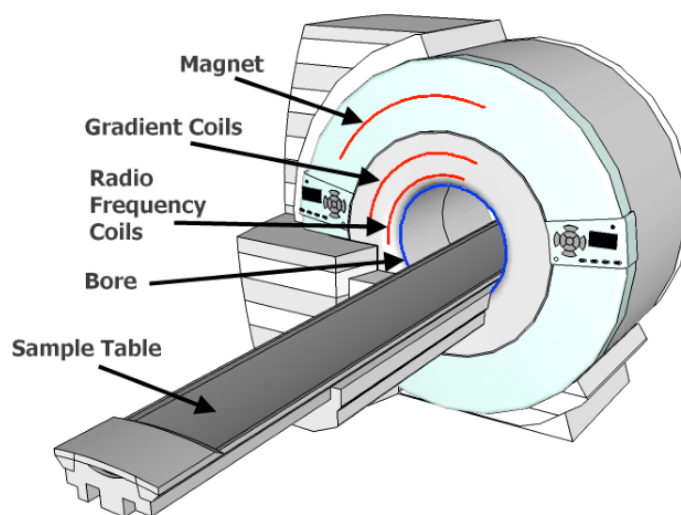


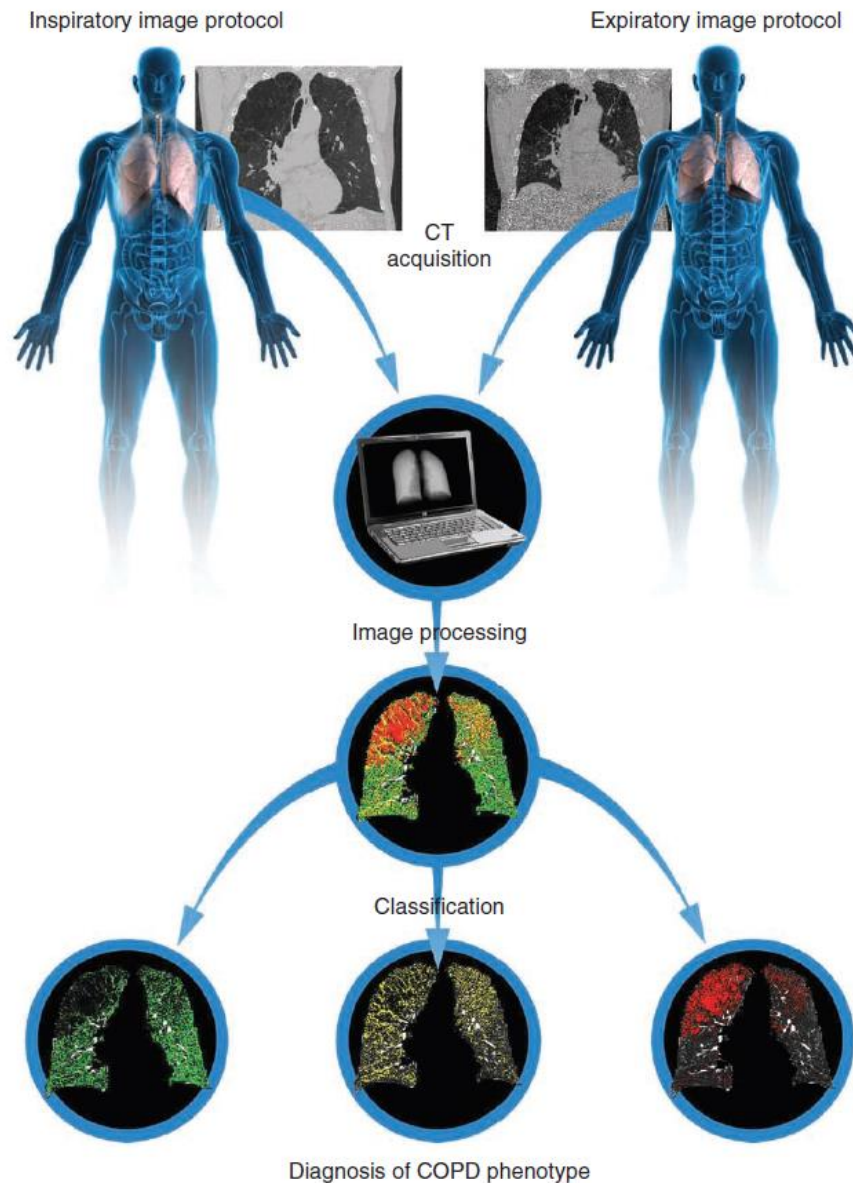
Diagram illustrating key components included in the construction and operation of a typical magnetic resonance imaging (MRI) scanner. Image reused from 'Geomorphological Techniques' ISSN 2047-0371 Page 45, under creative commons license CC BY-SA 4.0: <https://creativecommons.org/licenses/by/4.0>.

1.4.2 Parametric Response Map

In 2009 the parametric response map (PRM) algorithm was described as a voxel-wise approach in image analysis that was demonstrated to predict mortality in a cohort of subjects with high grade glioma (31); imaging data was obtained using dynamic susceptibility-weighted contrast (DSC) MRI, measuring cerebral blood flow (CBF) / volume (CBV) with contrast agents (181,182) . PRM demonstrated improved prognostic power over previous studies using simpler whole-tumour average methods (183). The algorithm relies on mapping points in space of a volume as its shape changes in time, following an entropy based approach to compute the geometric mapping, described by Meyer et al (184). Voxels in the reference image at one time step are then paired with voxels in a transformed homologous image at another time step. PRM features characterising each voxel into a discrete classification are then calculated based on the HU values at both points; the increase from one mean value to a number of classifications based on relative HU differences, and thus multiple mean values, explains the increased sensitivity and potential of this algorithm as a biomarker for disease prognosis. PRM was later additionally validated as an effective survival biomarker in cancer of the head and neck cancer, breast and prostate (185–187).

PRM was first applied to respiratory imaging data (CT) in a 2012 study published by Galbán et al in Nature Medicine (32), where the reference and homologous volumes were lungs segmented at expiration and inspiration states respectively, see **Figure 1-14**; VOI were segmented using a previously described automatic segmentation algorithm (188). Taking only voxels with HU in the open interval (-1000, -500) for both expiration and inspiration states, classification was derived by taking a HU value below -950 at inspiration to denote emphysema, and a HU value below -856 at expiration to denote gas trapping; a voxel equal to or above both threshold values was classified as normal. Specifically, using iHU and eHU to denote HU at inspiration and expiration respectively, classes were defined as: PRM^{Norm} if $iHU \geq -950$ and $eHU \geq -856$, PRM^{fSAD} if $iHU \geq -950$ and $eHU < -856$ and PRM^{Emph} if $iHU < -950$ and $eHU < -856$; the case $iHU < -950$ and $eHU \geq -856$ was unclassified. Subjects were part of the genetic epidemiology of COPD (COPDGene) multicentre observational study (189). The study shed light on the pathogenesis of COPD as voxels were observed to change state from PRM^{Norm} through PRM^{fSAD} to PRM^{Emph} , that is it illustrated functional small airways disease preceding emphysema in the disease pathogenesis.

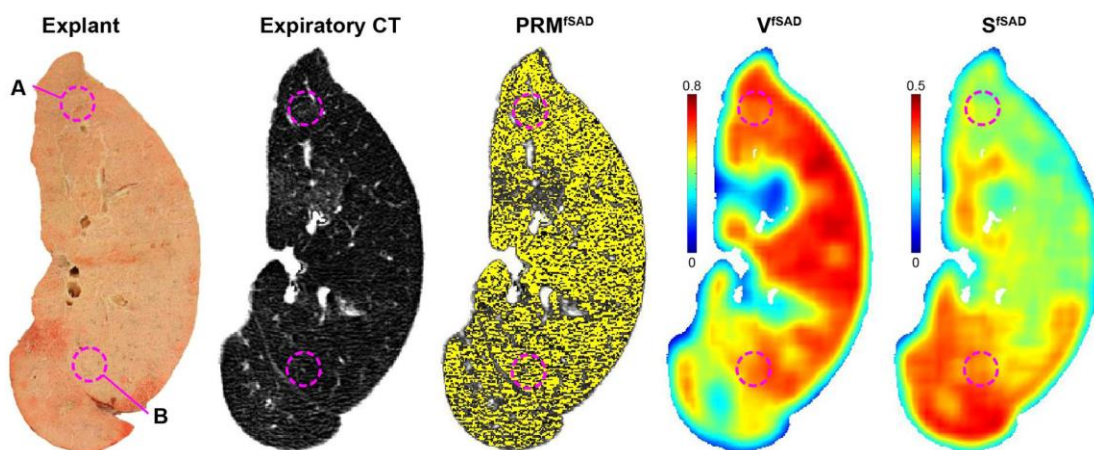
Figure 1-14: Image registration and parametric response map voxel classification.



Graphical overview of PRM for lung voxel classification in COPD from influential paper published in Nature Medicine (32). CT series of lungs at inspiration (left) and expiration (right) are image registered to obtain set of voxel positions with associated HU at inspiration and expiration for the same lung tissue. Classification as normal (green), fSAD (yellow) and emphysema (red) is then applied as a function of HU from both states. Reprinted by permission from Springer Nature: Nature Medicine. Reference: Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. Galbán, C.J. et al. © 2012.

The PRM algorithm has since seen application to respiratory imaging in numerous studies, focused primarily on COPD. The longitudinal relationship between regional gas trapping and emphysema was explored in a 2016 study by Boes et al (190), supporting the argument that fSAD is an intermediate stage as lung tissue transforms from healthy to emphysematous; the same group followed up later that year with a study on the impact of bias and noise on PRM in CT (191), concluding PRM measurement was relatively insensitive to CT scanner manufacturer or reconstruction kernel. Participants in the Dutch-Belgian lung cancer screening trial (NELSON), one of Europe’s largest running lung cancer trials (192), underwent PRM analysis to evaluate the relative value of PRM when compared against other CT biomarkers in multivariate logistic regression (193); PRM showing the highest diagnostic value of those tested. In 2017 PRM was utilised to study the relationship between age and SAD in COPD patients (194), evaluate risk factors in a five-year follow up in adult smokers from COPDGene (195), in which asthma was identified as a significant risk factor for exacerbations, and a topological augmentation to the algorithm published by Hoff et al (196) was developed and applied, using Minkowski functionals to study surface topology of a moving window (197), expanding the potential for disease phenotyping, see **Figure 1-15**. PRM was more recently applied in an asthma focused 2019 study by Bell et al studying the relation between a panel of CT based markers, including PRM, against asthma severity and disease characteristics (198) ;see the first study presented in this thesis.

Figure 1-15: Demonstration of topologically driven PRM improving disease phenotyping.



Transverse cross section of a lung illustrating topological PRM features providing additional insight into disease heterogeneity. Compare circled regions: S^{SAD} , local surface area of PRM^{SAD} , is distinct where other metrics are not. Explanted lung (left) supported observations with micro-CT analysis. Figure reused from (196), under license CC BY-SA 4.0: <https://creativecommons.org/licenses/by/4.0>.

1.4.3 Computational Modelling of FOT

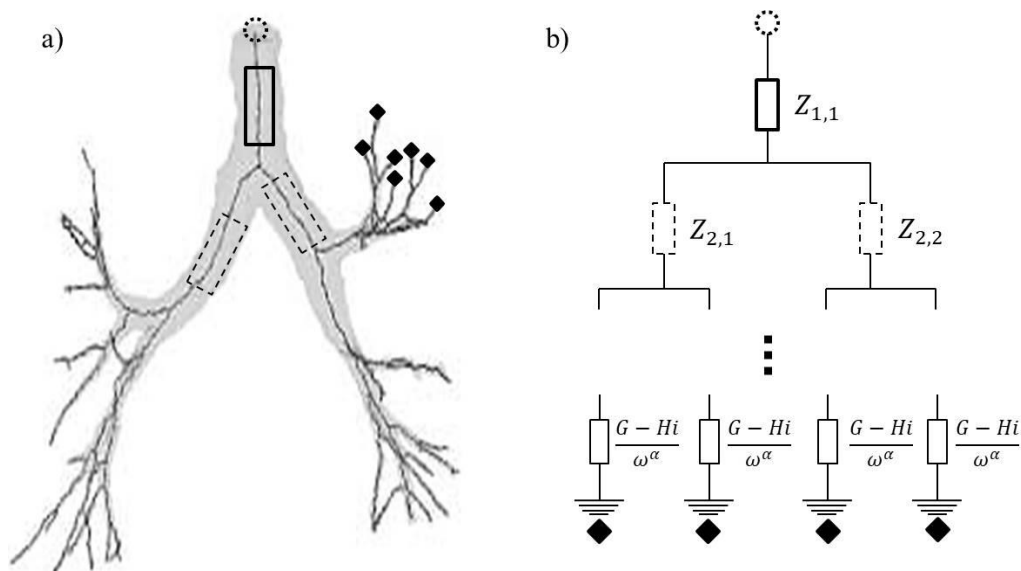
In recent decades the computational modelling of human anatomy and physiology has seen a significant rise towards increasing application in the biomedical sciences (199). In 2013 the AirPROM (airway disease predicting outcomes through patient specific computation modelling) consortium published on a framework designed to facilitate multi-scale computational modelling of airways in respiratory disease involving both patients and healthcare professionals (200). This body funded a study by Bordas et al in 2015 that developed a workflow to simulate patient-based complete conducting airways and therein model and analyse airflow (37), showing significant correlation with both FEV_1 and FEV_1/FVC against simulated resistance. The algorithm for growing a complete conducting airway tree, given an initial segmentation from CT data, is based on previously described methods (201), incorporating anatomical data published by Weibel (80). This complete airway tree generation and modelling functionality has now been augmented into the cancer heart and soft-tissue environment (Chaste) software, a project initiated in 2005 to provide a generic mathematical modelling framework for simulation in the biomedical setting (202).

A model for lung impedance simulation in the complete airway tree has been developed using the simplified analogy of an electric circuit, see **Figure 1-16**, where calculations of branch impedance are based on the propagation of soundwaves and periodic flow of incompressible fluid in a tube (203,204). Approximated solution to the 1D wave equation in Fourier space defines branch impedance as a function of resistance and reactance, where impedance of the terminal branches is determined by a constant-phase viscoelastic model previously described (205,206). Parameters, commonly noted G and H , for tissue dampening and elastance respectively, are estimated using previously reported experimental data (207).

The described airflow and impedance models, in simulated complete conducting airway trees, have been applied in numerous studies over recent years. Airflow modelling has been applied by Foy et al to study response of MBW and MRI to airway constriction (208) and the effect of gravity on MBW markers (209). In a 2017 study Soares et al evaluated the frequency dependence of resistance using both patient-based computational modelling and 3D physical modelling, showing concordance in the conclusion that isolated airway obstruction was not associated with frequency

dependence (210). More recently, in 2019, impedance modelling was applied to validate the pronounced effect observed in small airway resistance sensitivity, measured by R5-R20, to the inferior-superior distribution of pulmonary density changes in CT (198), and was later applied by Foy et al to study the role of the small airways in asthma, showing R5-R20 is a direct measure of small airway narrowing (38). Computational airflow modelling of MBW and impedance modelling of FOT has been cross examined by Foy et al under varying patterns of airway constriction, demonstrating MBW is primarily sensitive to the net magnitude of constriction whilst FOT appears more sensitive to airway depth (211).

Figure 1-16: Airway centreline from CT and associated circuit model of impedance.



Application of electric circuit modelling of impedance to airway tree centrelines. A, a major airway volume segmented from CT at inspiration (TLC) with centreline calculated (black central lines) from mid trachea (dotted circle) to feasible limit from CT resolution (6-8 generations maximum). Trachea and major bronchi outlined for reference, and black filled diamonds illustrated about some terminal branch points. B, associated electrical circuit diagram. Impedance Z is calculated for every branch, given terminal branch impedance Z_0 determined by a constant phase model:

$Z_0 = (G - Hi)/N\omega^\alpha$, where G and H parameterise dampening and elastance, $i=\sqrt{-1}$, $\omega = 2\pi f$, f is flow frequency and $\alpha = 2/\pi \tan^{-1}(H/G)$. N.B. major airway tree used for illustrative purpose, in practise Z_0 is assigned at terminal points of extended centreline simulation models, typically with order 10^4 terminal branches.

1.5 Data Analysis

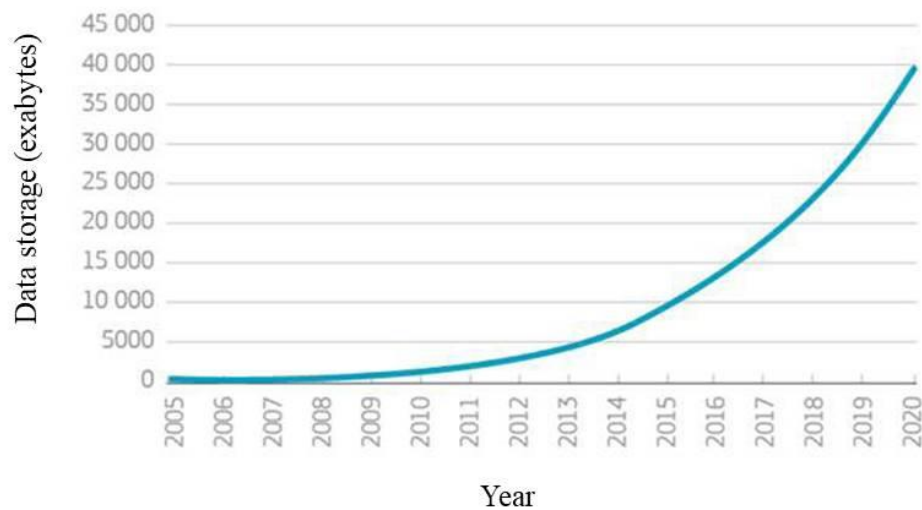
In this final introductory section, the problem of ‘big data’, a fast growing modern phenomenon, is described, statistical and clustering methods as they have been applied to biomedical data are discussed, and one novel mathematical approach to data visualisation using topology is introduced.

1.5.1 *Big data*

The term ‘big data’ has seen a fast growing and evolving definition in recent times, emerging in popular usage from around 2011 to describe the unprecedented increase in the volume of data being generated and stored globally (212), see **Figure 1-17**. One definition given by Laney in 2001 coined the ‘three V’s’ approach to understanding the significant factors (213); these are volume, variety and velocity. Roughly a decade later it appears to have gained traction in developing a standard framework (214,215). Data sets with high levels of volume and complexity are becoming more commonplace in many sectors of research, including the biomedical sciences and healthcare (43,216). Traditional data analysis is the application of proper statistical methods to extract useful information from data, and may still be applied for big data analyses (217); popular methods include cluster analysis, factor analysis, correlation analysis and regression analysis. Big data analytics are defined by the problem of rapid extraction of useful information, rapid being the key word, as standard methods tend to result in computationally difficult tasks under the weight of the three V’s.

The role big data may play with respect to asthma has been explored and elucidated in a number of studies over recent years. Disease prevention potential was explored in a 2013 study (218), presenting case studies where big data analytics is a promising means to improve asthma outcomes. Analysis of big data sources using artificial neural networks in 2015 demonstrated prediction of asthma-related emergency department visits (219), this used data from the ubiquitous companies Google and Twitter with environmental data. A 2017 review into the impact of big data on investigating the heterogeneity in asthma emphasized, in a cautionary tone, the importance of rigorous hypothesis testing to counteract the danger of readily accessible data analysis tools producing a decoupling between data mining and the scientific method (220); the closing statement emphasizing a pressing need for inter-disciplinary research involving scientists, clinicians and data analysts, to understand better the heterogeneity in asthma.

Figure 1-17: Global data storage estimations over the past two decades.



Growth of worldwide data storage over recent decades. Figure adapted from: https://ec.europa.eu/epsc/publications/strategic-notes/enter-data-economy_en, under creative commons license CC BY-SA 4.0: <https://creativecommons.org/licenses/by/4.0>.

1.5.2 Statistics

The most well-known and widely applied statistical methods to study heterogeneity in biomedical research are subgroup analyses for difference of some sample statistic such as the arithmetic mean, now a standard part of reporting in clinical trials and studies (221,222), efforts have been made to standardise this practise with protocols to limit misuse of techniques or misinterpretation of results (223), and the increasing importance statistical literacy has been emphasized (224). Selection of the proper statistical test depends on a number of properties, most prominently the number of subgroups; for one, two or three / more there is one-sample, two-sample and multiple group comparison testing respectively, the latter often requiring post-hoc testing for control of inflated p-values from the multiple comparisons (225).

Many traditional tests are designed with the assumption of a normal distribution in the feature being tested, and so it is necessary to test for normality by, for example, using the one-sample Kolmogorov-Smirnov test, based on works by the accredited early 20th century mathematicians (226,227). Attributes from a normal or non-normal distribution are commonly referred to as parametric or non-parametric respectively. If there exists a one-to-one relationship between pairs of samples between groups, e.g. the same subject

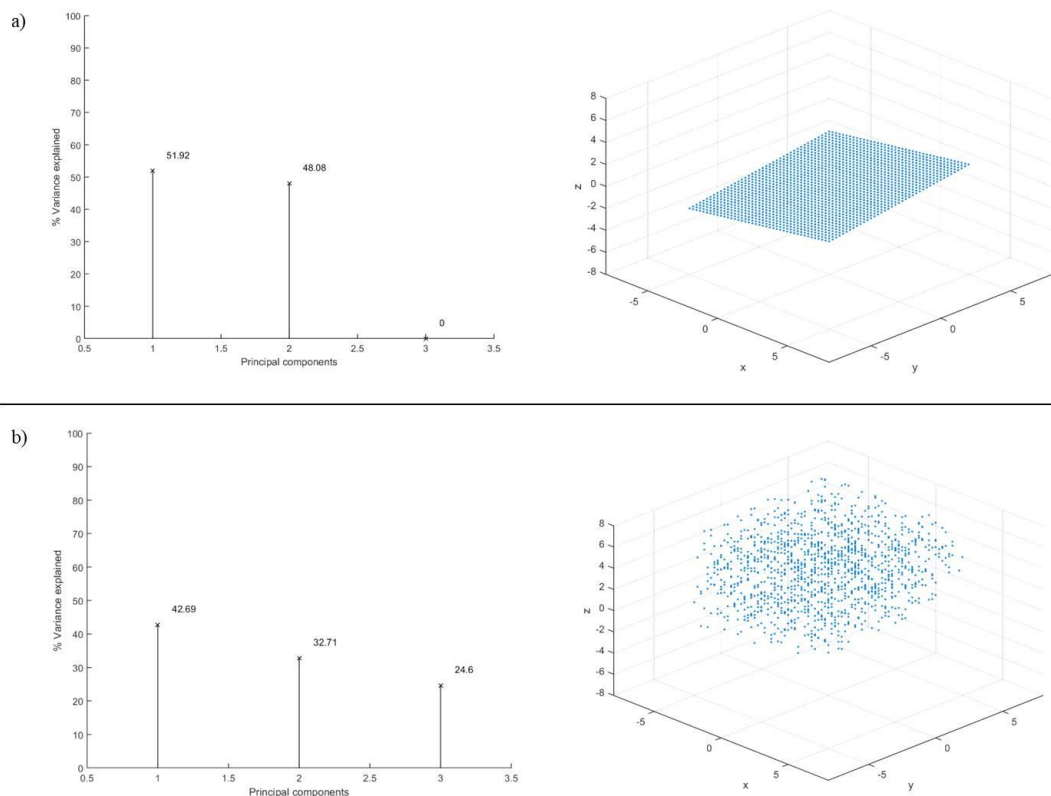
before and after some treatment, the samples are paired (or dependent), otherwise they are unpaired (independent). In the case of unpaired data, ubiquitous in the subgroup comparisons of this thesis, popular tests for two-samples include the independent two-sample t-test for parametric data, based on the t-statistic introduced by William Gosset / 'Student' (228,229), and the Man-Whitney U test for non-parametric data, based on 20th century work by the named mathematicians (230); multiple group comparison tests include one-way ANOVA (analysis of variance) for parametric data, a particular variant from a collection of statistical models attributed to the renowned statistician Ronald Fisher (229,231), and the Kruskal-Wallis test for non-parametric data, based on work by the named 20th century researchers (232,233).

Clustering or cluster analysis is widely recognised as the activity of partitioning data points into groups or classes, such that members of one group are more similar to each other than they are to members of other groups (234); this is a general guiding principle, and to date there is no universally acknowledged formal definition of 'cluster', a task which has been fraught with difficulty for millennia (235), and it has been argued that perhaps the attempt is misguided (236). Clustering has been applied in phenotyping of asthma in numerous studies over recent decades. In 2008 Haldar et al presented an application of k-means, one of the oldest clustering algorithms originating at least in the mid-20th century (237,238), to the identification of clinical phenotypes of asthma discriminated predominantly by inflammation (eosinophilic) and symptom expression (39). Latent class analysis (LCA), a subset of structural equation modelling (SEM) which has gained significant popularity (239), was applied in 2011 to phenotyping adult asthma in epidemiological studies (240); this identified four phenotypes discriminating quality of life, blood eosinophil and neutrophil counts within the population. In 2014 Ghebre et al applied k-means with linear discriminant analysis (LDA), a method for finding optimal projection for class separation, proposed by the aforementioned Ronald Fisher in 1936 (241,242), to present evidence supporting both the 'Dutch' and 'British' hypotheses of overlap with respect to asthma and COPD (243,244).

In 1901 English mathematician and biostatistician Karl Pearson published '*On lines and planes of closest fit to systems of points in space*' (245), a landmark paper that described what is now widely known and utilized as principal component analysis (PCA). This method is primarily seen as useful for dimensionality reduction, as points in higher dimensions are projected onto linear objects that best fit the point distribution. Principal

components are selected in a stepwise fashion such that each component maximises variance of the projected points whilst being orthogonal to all previously determined components; studying the fraction of variance explained by successive principal components provides insight into the maximal dimension $m < n$ of a data set of n features, see **Figure 1-18**. PCA and its close cousin factor analysis are now standard practise in biomedical studies, including asthma and COPD (246,247). A significant concern in dimensionality reduction is information loss due to projection to a lower dimensional space (47). Insight using PCA may be limited due to the fitting of linear objects to non-linear point clouds, causing ‘projection losses’. Many methods have sought to solve this problem, such as the ‘elastic map’ developed by Gorban and Zinovyev (248), allowing for non-linear fitting to point clouds in data sets using objects of controllable complexity / variable elasticity.

Figure 1-18: Dimension of a set of points, association with principal components.

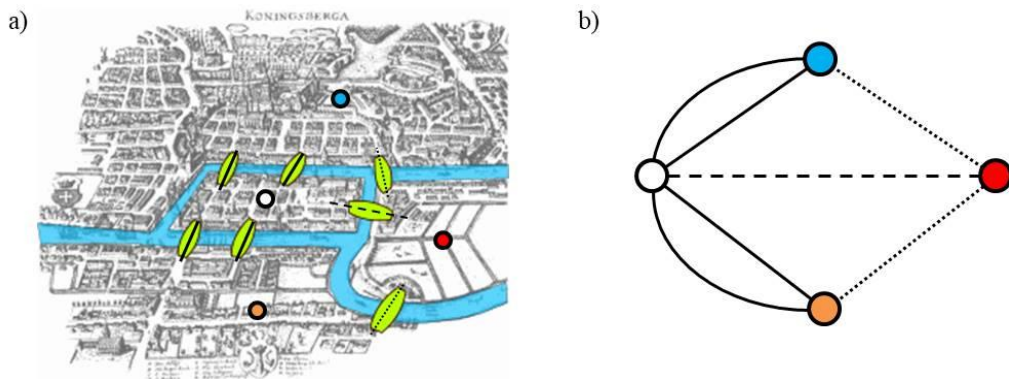


PCA applied to 2D plane with equation: $z = (x + y)/5 + \varepsilon$. A, $\varepsilon = 0$, variance explained completely by first two principal components, indicates maximum of 2 dimensions in data. B, ε is randomly generated from (0,10); significant variance explained from all three principal components, implies maximum of 3 dimensions.

1.5.3 Topological Data Analysis

Topology is a branch of mathematics that can trace its origins back to the early 18th century, when in 1736 the famed mathematician Leonhard Euler published a paper on a solution to the ‘Seven Bridges of Königsberg’ problem (249), see **Figure 1-19** for illustration; a key feature of his method was the replacement of the complex geometry of the city’s land masses and rivers with an abstraction made simply of points connected by lines, arguably embodying the essence of mathematical problem solving in the simplification of the problem down to its essential components. The solution followed from analysis of the properties of the resulting graph, in particular the conditions under which the graph could be traversed starting from and ending at the same vertex. It is argued that Euler thus established the basis of both topology and graph theory with this work (250). Another landmark development came in 1750 with ‘Euler’s gem’ or ‘polyhedron formula’, an equation classifying polyhedra based on an alternating sum of counts of their faces, edges and vertices (251); better known in general form as the Euler characteristic, commonly denoted using the Greek letter chi: $\chi = V - E + F$. Many consider this to mark the birth place of topology.

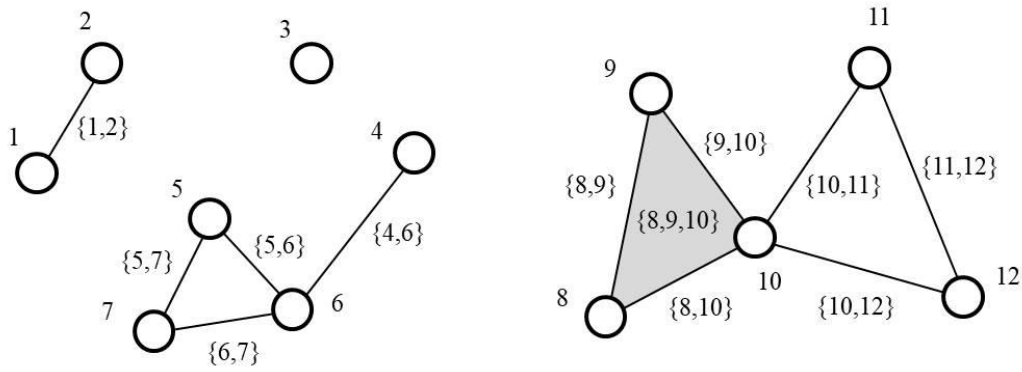
Figure 1-19: The seven bridges of Königsberg and associated representative graph.



The seven bridges of Königsberg with associated graph. A, a map of the seven bridges of Königsberg, mid-18th century, with annotations on rivers, bridges and land masses. B, graph providing a compressed representation of the geographical information. The problem was to determine whether or not it is possible to cross every bridge only once on a single walk. Euler proved it was impossible, solving the equivalent problem of moving between vertices of the graph (B) using each edge only once. Map (A) adapted from: https://en.wikipedia.org/wiki/Seven_Bridges_of_Königsberg, under creative commons license CC BY-SA 3.0: <https://creativecommons.org/licenses/by-sa/3.0/>.

In 1895 another renowned mathematician, Henri Poincaré, published ‘Analysis Situs’, a seminal work building on advancements made by other notable mathematicians such as Cauchy, Riemann and Betti, introducing the key concept of homology to the field and considered the birth of algebraic topology (252). Homology roughly speaking aims to count ‘connected components’ and ‘holes’ present within an object, see **Figure 1-20**. Betti numbers summarise homology over varying dimensions, and the sensitivity of these numbers to changes gives rise to the idea of persistent homology, an important component in topological data analysis (TDA) (253). In 2016 researchers studied persistent homology in the 1000 Genomes Project (1KGP) data for mapping recombination in human populations (254).

Figure 1-20: Simplicial complex with associated Betti numbers β_0 and β_1 .

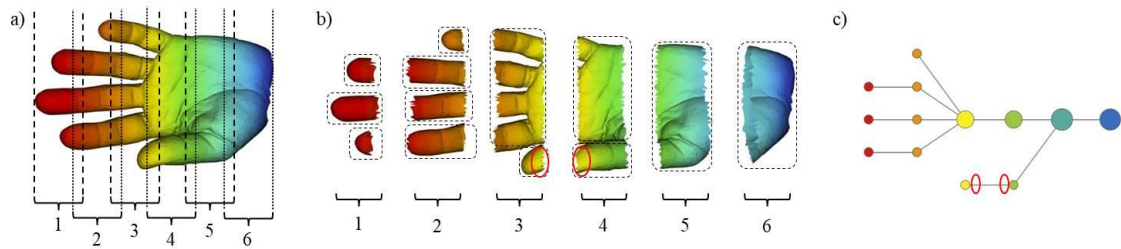


A simplicial complex with typical notation illustrating homology with associated Betti numbers β_0 and β_1 counting distinct sets of connected components and holes respectively. Vertices are indicated with numbers 1-12, edges as sets of vertices with exactly two members, and a face as a set of three vertices ($\{8,9,10\}$). There are four distinct sets of connected vertices: 1-2, 3, 4-7 and 8-12; $\beta_0 = 4$. A ‘hole’ is more precisely defined in this context as a cycle of edges with empty interior, there are two: $\{5,7\} \rightarrow \{5,6\} \rightarrow \{6,7\}$ and $\{10,11\} \rightarrow \{11,12\} \rightarrow \{10,12\}$; $\beta_1 = 2$.

In 2007 Singh, Mémoli and Carlsson published a paper presenting a novel method called ‘mapper’ (255), a workflow inspired by ideas from topology, in particular the Reeb graph (256). The primary purpose of this TDA approach is to create useful visualisations presenting a compressed representation of the ‘shape of data’. The algorithm achieves this by applying clustering within overlapping subsets of the data points, and assigning points within the same cluster to a single vertex (0-simplex); then vertices with points in common are joined by a line (1-simplex), resulting in a

simplicial complex (list of simplexes) that may be mapped into the plane as points joined by lines presenting a visualisation based on similarity; see **Figure 1-21**. The metric used to determine overlap, and the degree to which overlapping occurs, are significant parameters in determining the resulting visualisation.

Figure 1-21: Data compression using TDA mapper algorithm.



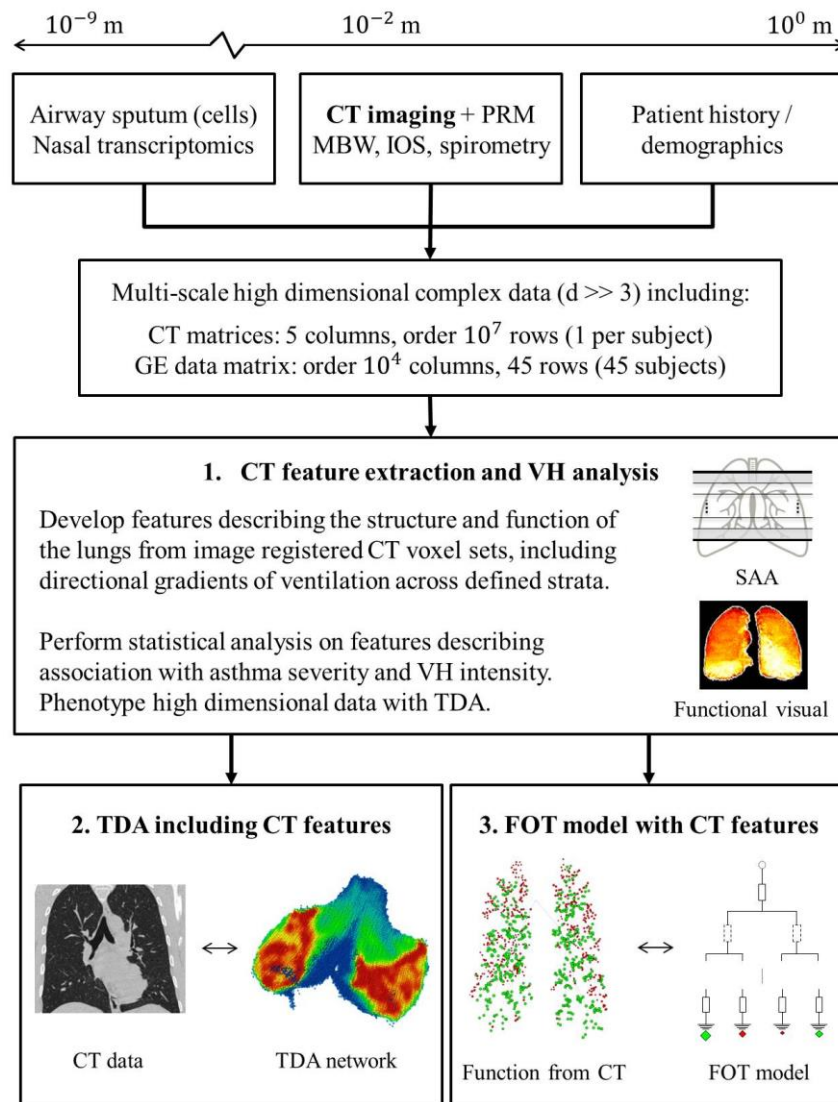
Demonstration of data compression retaining key topological traits of a set of points using TDA mapper. A, a set of points is sampled approximating the shape of a human hand, overlapping subsets are defined with respect to the horizontal coordinate function. B, clustering is applied in each subset, clusters indicated by dashed outline; case of clusters with points in common outlined in red. C, graph representing the hand, example case of edge between vertices from clusters with points in common annotated. Image adapted from <https://www.nature.com/articles/srep01236> (257), under creative commons license CC BY-SA 3.0: <https://creativecommons.org/licenses/by-sa/3.0/>.

Ayasdi is a company founded by Carlsson et al in 2008 to provide a computational platform capable of producing visualisations of data using the mapper algorithm (258). Ayasdi's potential for application to big data, one of its driving goals, has been evaluated (259); the thesis advocated the strength of this tool being in providing rapid delivery of qualitative insight into data with minimal effort by the user. In 2015 a paper on phenotyping type 2 diabetes subgroups achieved the front page of science translational medicine (50), presenting three subgroups of the disease in a generated network; in the same year another group applied TDA in the study of spinal cord and traumatic brain injury in rodents, presenting proof of concept in applying TDA to high volume neurotrauma data sets (260). In 2018 Soares et al applied TDA to identify clinical phenotypes in asthma incorporating the study of particles in exhaled air (PExA) data (52), and later that year Siddiqui et al employed TDA to evaluate pathological heterogeneity in asthma, demonstrating replication of phenotype discovery through persistent homology (261).

1.6 Aims and hypotheses

The primary aim of this thesis was to extract information from CT imaging data to incorporate into high dimensional multi-scale data sets to analyse for insight into mechanisms driving VH in asthma, including application of TDA for phenotyping and compressed visualisation of the lung, to study spatial distribution of disease and lung function. Furthermore the resulting methods extracting lung function information from CT data would be linked with airway tree structures used in FOT modelling, supporting future insight discovery through improved computational modelling. See **Figure 1-22**.

Figure 1-22: Diagrammatic overview of study aims and work flow.



CT, computed tomography; PRM, parametric response map; MBW, multiple breath washout; IOS, impulse oscillometry; GE, gene expression; VH, ventilation heterogeneity; TDA, topological data analysis; SAA, stratified axial analysis. Data from disparate sources combined to study VH in asthma.

Specific aims were as follows (1.6.x):

1.6.1 *Perform feature extraction and statistical analyses on PRM registered CT data to study association with clinical markers of asthma and VH.*

The first task was to develop and implement an array of algorithms mapping PRM inspiration-to-expiration registered CT data to a number of descriptive attributes, including approaches to quantify gradients of lung function surrogated by density changes reflected in HU. These descriptors could then be cross analysed against features characterising asthma severity and VH magnitude, derived from clinical data sets including spirometry, MBW and IOS testing, nasal transcriptomics and induced sputum tests for inflammatory markers.

Binary and multiple group comparison testing would form the basic statistical analyses, followed by linear regression modelling, principal component analysis and linear discriminant analysis, to study relationships between CT and non-CT features, and in particular determine CT markers most significantly associated with asthma and VH, using GINA, FEV₁ and FEV₁/FVC, and MBW and IOS attributes. TDA would be applied to and integrated data set to conduct phenotyping to support observed associations. Hierarchical clustering applied to transcriptomics would identify groups of subjects with similar genetic profiles, with respect to the normalized correlation metric, which would be studied for discrimination of CT markers associated with VH.

I hypothesised that:

- PRM registered CT data would contain information on VH and be able to discriminate between extremes of subjects with respect to MBW and IOS.
- CT based features quantifying spatial gradients in lung function would associate with global markers of VH at the mouth.
- TDA would identify phenotypes in the given population that could be discriminated based on VH and VH-associated CT features.

1.6.2 Apply TDA to CT based data sets for visualisation of the lungs and lung function in asthma.

The second task was to leverage TDA to visualise patterns of ventilation and disease in the small airways through application of TDA to imaging data, with lung function features extracted, combined with simulated extended airway centrelines. To effectively visualise high dimensional imaging data sets of multiple spatial, anatomical and functional features across order 10^7 voxels, TDA would be applied directly to the associated data matrices, in subjects selected to represent varying levels of VH.

Complete conducting airway tree models based on major airway and lobar segmentations, and an anatomically informed simulated growth algorithm, would be utilized to assign tree depth scores to voxels based on proximity to tree branches and their associated Strahler order (SO); this would be used to allow TDA networks to isolate the smaller airways. The creation and implementation of this workflow, including the SO assignment algorithm, formed the initial workload for this aim.

Resultant lung-specific networks would then be interrogated for patterns of disease and function, represented by PRM^{fSAD} and ΔHU respectively, relative to the apparent VH magnitude and severity of asthma in subjects, as indicated by GINA score and established VH markers from MBW (LCI and S_{acin}).

I hypothesised that:

- A k-NN (k-nearest neighbour) assignment of SO to voxels would provide a method to isolate small airway regions in TDA visualisations.
- ΔHU and PRM^{fSAD} would discriminate between asthmatic and non-asthmatic subjects visibly in the small airway regions of TDA networks.

1.6.3 *Link functional CT imaging data with airway tree structures to support FOT modelling with regional compliance parameterisation.*

The final task was to again combine extended airway centrelines and imaging voxel sets, though this time mapping values of ΔHU onto the terminal branches of airway trees. The idea was to acquire tree structures with measures of regional compliance linked to every branch, which could then parameterise branch radii to increase FOT modelling accuracy of resistance in the small airways (R5-R20). An additional output would be the visualisation of branch ventilation, a significant compression from order 10^7 voxels to say 10^1 branches of the first few generations from the trachea (e.g. 2^5).

To complete this aim, FOT modelling would be carried out in a cohort of subjects with QC/QA passed tree structures, with ΔHU assignments, alongside IOS measurements of resistance; R5, R20 and R5-R20. Schemes for altering airway radii as a function of ΔHU would be developed and tested across a range of weighting values, with testing of correlation between clinically measured and computationally simulated values.

I hypothesised that:

- Adjustment of airway radii using ΔHU would significantly improve the linear correlation between measured and simulated R5-R20.
- The distribution of ΔHU in voxels proximal to terminal branch points would be shifted right of zero, relative to the global distribution; that is SO1 proximal voxels would better exhibit functional behaviour (parenchyma): $\Delta HU > 0$.
- The visualisation of airway trees with branch ΔHU indicated would clearly show the distribution of ventilation in the lungs.

1.7 Structure of thesis

To conclude this introductory section, a summary of the remaining sections, set out to address the given aims and hypotheses, is given.

The methods section anon (2.x) will cover:

- The clinical trial from which data presented in this thesis is derived.
- Procedures and protocols for data acquisition from all sources.
- CT feature extraction methods including segmentation and PRM.
- Analytical algorithms developed for outcomes presented in this thesis.

In general methods will be described technically, referencing code in the appendices where appropriate demonstrating implementation of developed software enacting described algorithms. Matlab was the primary language used for software development in the thesis, with Python supporting TDA network discovery and homology persistence study, and Bash (bourne again shell) automating PRM work.

Three studies are then presented over the following three sections (3.1-3.3), across which the described methods are applied to data from subsets of subjects from the given clinical trial, to fulfil the aims and test hypotheses set out in the previous subsection (1.6.x). Each study is adapted from a manuscript and presented in a style close to the standard ‘original research article’ format followed by tables and figures.

1. Functional CT imaging for the identification of the spatial determinants of small airways disease in adults with asthma.

This first study applied multiple standard statistical methods to evaluate the relationships between extracted CT markers and non-CT features, in a set of 52 subjects, 41 asthmatic and 11 healthy volunteers. Stratified axial analysis (SAA) was the most notable outcome in methods development here, describing HU changes in strata along inferior-to-superior and anterior-to-posterior directed vectors.

SAA imaging markers uniquely discriminated between population thirds for both MBW and IOS features, illustrating a potential link between ventilation gradients in the lung and ventilation heterogeneity measurements made at the mouth.

2. Topological data analysis to visualise the lungs and small airways in adult asthma.

In this study a 1-NN SO assignment algorithm was applied to lung voxels to provide a measure of depth in the lung with respect to the airway tree; then data sets of lung voxels with attributes describing spatial, anatomical, functional information, PRM class and SO proximity, were subjected to TDA mapper using Ayasdi culminating in a workflow for producing highly compressed multidimensional visualisation of the lungs.

The workflow was applied to 4 subjects selected to represent lungs with no VH, low VH, mild VH and high VH, based on MBW markers of VH. TDA demonstrated airway and vascular tissue segmentation, as well as discrimination of asthmatic (n=2) and healthy (n=2) subjects through spatial distribution patterns of disease and ventilation.

3. Quantitative CT image analysis to support computational modelling of small airway resistance in asthma.

In the final study, the airway tree to voxel linkage process was inversed; voxel ΔHU values are used to assign mean ventilation scores to terminal branches. Scores for all other branches were calculated by averaging. Visualisations of branch ΔHU across specified ranges of SO were generated. Branch radii were adjusted in direct proportion to assigned ΔHU , and the resulting correlation between simulated and measure resistance in FOT modelling was evaluated across a range of constriction magnitudes.

Visualisations of ΔHU in the upper airways of trees provided proof of concept of the methods developed, as expected ventilation gradients due to subject posture and lung health were strongly visible. The distributions of ΔHU about the deepest parts of the airway tree were seen to shift significantly right of zero relative to the global voxel set. An optimal weighting for radii adjustment was identified and presented a marked improvement in simulated and measured R5-R20 correlation.

The final section of the thesis (4.x), entitled conclusions, gives a discussion of the whole work, including limitations and significant points of controversy, comments on future direction for research and provides a closing statement on the underlying project.

2 Methods

2.1 Overview

In this section of the thesis all methods pertinent to data collection and analysis are described, with specific protocols and technical details where appropriate. Code snippets are provided in simplified form in the appendices from source scripts developed and implemented in this thesis, to elucidate application of designed algorithms, and are referenced as required in the text.

Matlab R2015a (Matlab Release 2015a, The MathWorks, Inc., Natick, Massachusetts, United States) was used to perform all statistics and implement algorithms developed for feature selection, extraction and analysis. Topological data analysis was performed using the Ayasdi machine intelligence platform (release 6.x) and associated Ayasdi Python software development kit (SDK); Ayasdi, 4400 Bohannon Dr #200, Menlo Park, CA 94025, United States. Imbio lung density analysis software (Imbio, LLC, Minneapolis, MN) was used for initial CT image segmentation and registration (TLC to FRC). Independently developed software Elastix (262,263) was utilized with Matlab code provided by the University of Michigan centre for molecular imaging (CMI) lab for FRC to TLC image registration, on segmentations performed using the Yacta (yet another CT analyser) software application (264), v2.8.5.11. Airway volume and lobar surface calculation, for complete conducting tree modelling, was performed a priori using ‘mimics’ software by Materialise NV (Leuven, Belgium).

The first section gives an overview of the subjects and clinical trial procedures providing the source data on which this thesis is based. This covers: collection of self-reported and measured clinical and demographical characteristics, asthma diagnosis and inclusion / exclusion criteria, asthma control and quality of life assessed by questionnaire, atopy classification assigned using allergen skin prick testing, eosinophil and neutrophil counting from induced sputum samples, and nasal transcriptomics sampling and processing from scraping of the nasal epithelium.

Then lung function test protocols are described for body plethysmography, spirometry, multiple breath gas washout and impulse oscillometry methods. In each case test equipment specs and maintenance are addressed where appropriate, along with subject

posture (sedentary), test accessories and instruction on test manoeuvres such as breath holding / control, cheek support and equipment use are specified where appropriate. Statistical requirements for recording test data are stated.

CT imaging is the focus of the third section, covering image acquisition, i.e. raw data generation using CT scanner, image segmentation and registration, FOT modelling using airway centrelines and lobar surfaces determined from segmentation, and alignment of point sets from similar CT volumes, specifically points from the same underlying inspiration image series but generated using different software and workflows; one using original raw inspiration positional vectors, and the other output from the combination of simulated centrelines and lobar boundaries.

The forth section covers all major CT feature extraction exercises performed in this thesis, including description of developed algorithms with formulae and practical implementation in Matlab. Basic features, including fundamental PRM indices, are discussed first from the perspective of the joint density histogram (JDH) of expiration and inspiration voxel distributions. Global features based on voxel counting and HU change (ΔHU) statistics are covered. Study of ΔHU variance in a given direction across the lung is discussed under the term stratified axial analysis (SAA). Inter-lung features are briefly discussed, i.e. features arising from comparison of left and right lung volumes. Finally, assignment of SO score to voxels, and in the other direction conceptually, ΔHU score to branches of an aligned airway tree, are described.

In the last section, algorithms for the analysis of data arising from the aforementioned methods are covered. Statistical methods: binary and multiple group comparisons, feature correlation analysis, tertile polar analysis, feature and feature set relational strength evaluation using mean correlation, linear regression, both raw and multicollinearity reduced, principal component analysis with Kaiser rule, and Fisher's linear discriminant analysis (LDA). JDH generation and visualisation algorithms are briefly described, with min-max projection method for spatial visualisation of voxel clouds. Hierarchical clustering applied to gene association work using nasal transcriptomics is described. Topological data analysis (TDA) with Ayasdi is covered with application to both data driven phenotyping and pulmonary visualisation, followed finally by operations to visualise ΔHU in simulated airway trees and methods for affectation and analysis of airway constriction in FOT modelling.

2.2 Clinical Procedures

This section covers subject recruitment and clinical / demographical features including asthma control and quality of life questionnaires (ACQ/AQLQ), atopy classification, sputum induction and nasal scraping for transcriptomic microarray analysis. Clinical trials were predominantly funded by Roche Pharmaceuticals (Basel, Switzerland) and the European Union Airways Disease Predicting Outcomes in Patient Specific Computational Models (AirPROM-FP7) consortium, and carried out by staff and equipment based on site at Glenfield Hospital (Leicester, UK) biomedical research centre (BRC) respiratory theme.

2.2.1 Subjects

Asthmatic subjects were recruited from Glenfield Hospital in Leicester, UK; Aged matched healthy volunteers were recruited via local advertising and staff with normal airway physiology and no features of respiratory disease. All subjects were non-current smokers, however due to the known association of smoking and small airways disease, smoking exposure was not an exclusion criterion.

Asthma had been defined by a clinician diagnosis with one or more of the following criterion (i) bronchodilator reversibility of FEV_1 to 400 mcg inhaled salbutamol of $\geq 12\%$ and 200 mls (17 of 41 asthmatics), (ii) Methacholine $PC_{20} \leq 16$ mg /ml (11 of 41 asthmatics) or (iii) peak flow variation of $\geq 20\%$ over a 2 week period (13/41). Asthma severity was classified according to the current Global Initiative for Asthma (GINA) treatment steps (143). Severe asthmatics within the cohort had similar lung function (post bronchodilator FEV_1/FVC) to previously reported severe asthma cohorts in Leicester, UK (265), but higher average post bronchodilator $FEV_1\%$ predicted values.

Clinical and physiological assessments were performed over 1-2 study visits, no more than 1 week apart. Asthma medication, such as prescribed inhaled corticosteroid (ICS) dosage, and any comorbidity were recorded. Subject age, sex and smoking history (pack years) were recorded. Exacerbation frequency was quantified according to given ATS/ERS criteria. Asthma duration was estimated from patient reported age of onset.

Carbon monoxide uptake in the lung was determined using the single-breath method, according to standard guidelines (266). Alveolar volume (V_A) and the carbon monoxide transfer coefficient (K_{CO}) were calculated.

2.2.2 Questionnaires (ACQ/AQLQ)

The Juniper asthma control questionnaire (ACQ) is a widely known method for characterising asthma control, developed and validated by Juniper et al (73). Questions require asthmatic participants to describe symptoms relating to asthma control on a 7-point scale marked 0 to 6 in increasing severity of response. In this study a shortened variant, ACQ-6, was used to assess asthma control in participating subjects; studies have shown ACQ-6 performs sufficiently in comparison to the complete ACQ (75).

The Juniper asthma quality of life questionnaire (AQLQ) is a well-known method for assessing quality of life impairment due to asthma, developed also by Juniper et al (72). It presents asthmatic participants with 32 questions covering recent experience with respect to symptoms, activity, emotions and environment. A standardised and validated version of this questionnaire was used in this study (74).

2.2.3 Atopy Testing

An allergen skin prick test was used across all participants to assign atopy classification (yes/no). All participants were instructed to avoid using non-vital anti-histamine medication for at least 72 hours (3 days) prior to testing. Common allergens tested included feline fur, canine dander, grass pollen and *Dermatophagoides pteronyssinus*. A wheal, a small raised swelling area about the prick site, with a diameter $d \geq 2\text{mm}$ larger than a control (saline solution) prick after 15 minutes indicated a positive response to a particular allergen. One or more positive responses determined positive atopy classification (yes) in a participant.

2.2.4 Induced Sputum Tests

Airway inflammation was characterised by eosinophil and neutrophil counting in sputum, following procedure previously described (267). Subjects inhaled bronchodilator (salbutamol) through a spacer 15 minutes prior to testing. Optimal FEV₁ was then recorded as a benchmark for comparison during the induction method effecting inhalation of a saline solution at increasing concentrations of 3%, 4% and 5% to induce sputum production from the airways; concentration increment was stalled for intermediate FEV₁ between 80% and 90 of optimum, and testing halted if it fell below 80%. Cell counts were performed on successful sample collections using a Neubauer haemocytometer with trypan blue exclusion method.

2.2.5 Nasal Transcriptomics

Nasal epithelial cell samples were obtained using a nasal scraping. Subjects underwent nasal scraping to collect airway epithelial cells. Scrapings were performed behind the inferior turbinate using a nasal cytology curette to obtain a cellular sample, which was then stored in 500µl of RNA protect cell reagent, in an RNase/DNase-free cryovial, and kept at -80°C until shipment to central laboratory. RNA was extracted and processed using a Human Genome U133 Plus 2.0 Array from Affymetrix. Gene expression levels were derived and normalised.

2.3 Lung Function Tests

This section covers lung function testing protocol for body plethysmography, spirometry, multiple breath washouts and impulse oscillometry. All physiological tests were performed in the seated position by individuals with appropriate training and accreditation. Physiological tests were performed 15 mins after administration of a short-acting bronchodilator (salbutamol 400 µg). This was administered via a metered dose inhaler and spacer, with each 100 microgram actuation being inhaled in a separate inhalation to TLC, followed by a 5- to 10-s breath-hold.

2.3.1 Body Plethysmography

Lung volumes such as total lung capacity (TLC), functional residual capacity (FRC) and reserve volume (RV) were calculated using the whole body plethysmography method (268), using a constant volume plethysmograph and following ATS/ERS guidelines (269). Prior to testing subjects were screened for contraindications. Subjects were seated in a ‘body box’ with an isolated environment and left for a minimum of two minutes for temperature equalisation. Subjects were then instructed to support their cheeks with their hands, wearing a nose clip to block the nasal passages, and undergo a minimum of four repeated breathing manoeuvres. A shutter was manually activated and subjects instructed to inspire and exhale against the shutter to obtain pressure measurements; shutter automatically engaged at end expiration once active, and released after one inspiratory and expiratory breath.

2.3.2 Spirometry

Spirometry was performed according to ATS/ERS standards (109). Tests were performed using a Vitalograph spirometer, with subjects in a sedentary position

instructed to inspire to full inspiration (TLC) followed by a forced expiration, that is breathing out as fast and for as long as possible, without pain or discomfort. The breathing manoeuvre was repeated at least 3 times, with minimum 1 minute rest between manoeuvres, until consistent measurement of forced expiration volume in the first second, FEV₁, and functional vital capacity, FVC, were recorded to within a tolerance of 5% of each other; tolerance was as advised by the afore-cited standards.

2.3.3 Multiple Breath Washout Testing

Multiple breath inert gas washout (MBW) was performed in triplicate according to current guidelines (20), using the method described by Horsley et al. (270). Volume calibration of the pneumotachograph was performed daily using a 1-L syringe. Participants wore a nose clip and breathed an air mixture containing 0.2% SF₆, while respiratory flows and exhaled breath SF₆ concentrations were monitored by an Innocor photoacoustic gas analyser (Innovision A/S, Odense, Denmark). Participants maintained a steady respiratory rate of approximately 12 breaths per minute, and a constant tidal volume of 1L, using a real-time visual display of inspired volume as a guide. Once inhaled and exhaled SF₆ concentrations had equalized, participants were switched to breathing room air during expiration and asked to continue breathing at the same respiratory rate and tidal volume. The test was terminated once the end-tidal concentration of SF₆ in exhaled breath reached less than 1/40th of the original concentration (0.005%) for three consecutive breaths. The parameters S_{cond} and S_{acin} (120) were calculated using custom software written with TestPoint (Measurement Computing Corporation, Norton, MA, USA).

2.3.4 Impulse Oscillometry

Impulse oscillometry (IOS) was performed in triplicate according to standard guidelines (133). A volume calibration was performed daily using a 3-L syringe, and the accuracy of resistance measurements was confirmed daily using a standard 0.2 kPa·s·L⁻¹ resistance mesh. Participants wore a nose clip and supported their cheeks, while an impulse waveform was delivered to their respiratory system via a loudspeaker connected to a mouthpiece, during 60 seconds of tidal breathing. Resistance at 5 Hz (R5), resistance at 20 Hz (R20), and thus small airway resistance measure R5-R20, as well as reactance at 5 Hz (X5) and area of reactance AX were derived from pressure and flow measurements recorded throughout the 60-second period.

2.4 CT Imaging

This section covers CT image acquisition, segmentation and registration, FOT modelling and imaging set alignment procedures. Graphics are provided illustrating key functions: Imbio lung density analysis software automating PRM (segmentation, registration and tissue characterisation), lung volume segmentation and centreline calculation with mimics, Yacta segmentation and CMI scripts for FRC to TLC registration, extended centreline tree simulation and voxel point set alignment.

2.4.1 Acquisition

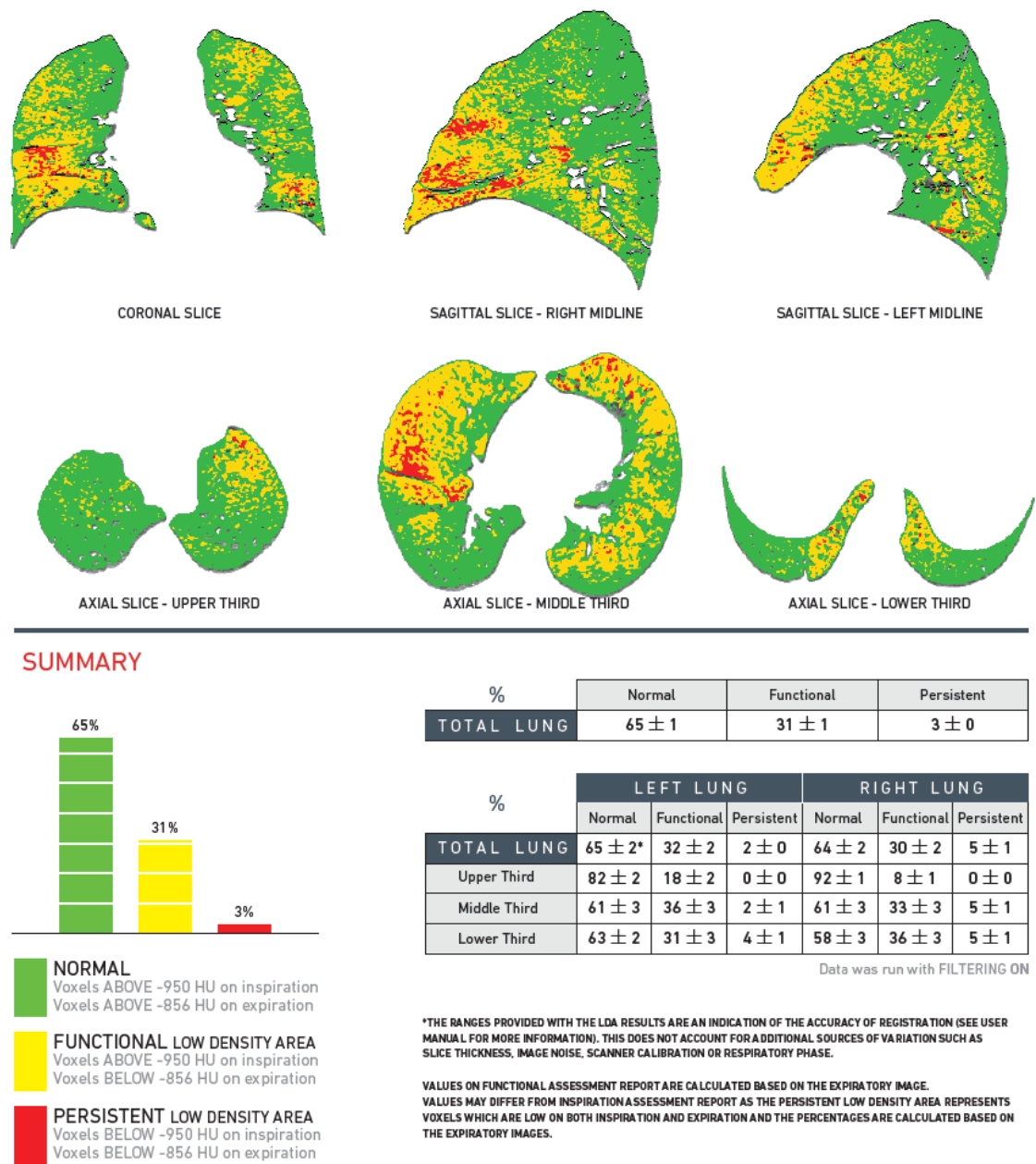
Volumetric whole-lung scans were obtained with a Siemens Sensation 16 scanner with the following low-dose protocol: $16 \times 0.75\text{mm}$ collimation, 1.5-mm pitch, 120 kVp, 40 mA, 0.5-second rotation time, and scanning field of view of 500 mm, with dose modulation off. Scans were obtained at full inspiration and at functional residual capacity, following the administration of 400 mcg of inhaled Salbutamol. Images were reconstructed with a slice thickness of 0.75mm at a 0.5mm interval by using B35f kernel, stored in Digital Imaging and Communications in Medicine (DICOM, .dcm) format on DVD, and shared electronically with the University of Michigan through a data transfer agreement.

2.4.2 Segmentation and Registration

Registration of the inspiratory (at total lung capacity, TLC) and expiratory (at functional residual capacity, FRC) DICOM series was initially performed at the University of Michigan (USA), using Imbio's Lung Density Analysis (LDA™) software application (Imbio, LLC, Minneapolis, MN). Sample of Imbio visual output with summary statistics for PRM is provided in **Figure 2-1**.

Imbio software automatically segmented lung volumes (excluding the major airways: approximately 3-4 generations from the trachea), the series with least volume was taken to be the expiration set, and a warping function T approximated regional geometric deformation between expiration and inspiration states, minimizing residual complexity. Registration using Imbio was performed inspiration-to-expiration (I2E), that is, a voxel v in the expiration image is linked to a set of voxels in the inspiration image based on $T(v)$. This provided an assignment of both expiratory Hounsfield unit (HU), i.e. HU of v , and inspiratory HU, i.e. mean HU of the voxel set captured by $T(v)$.

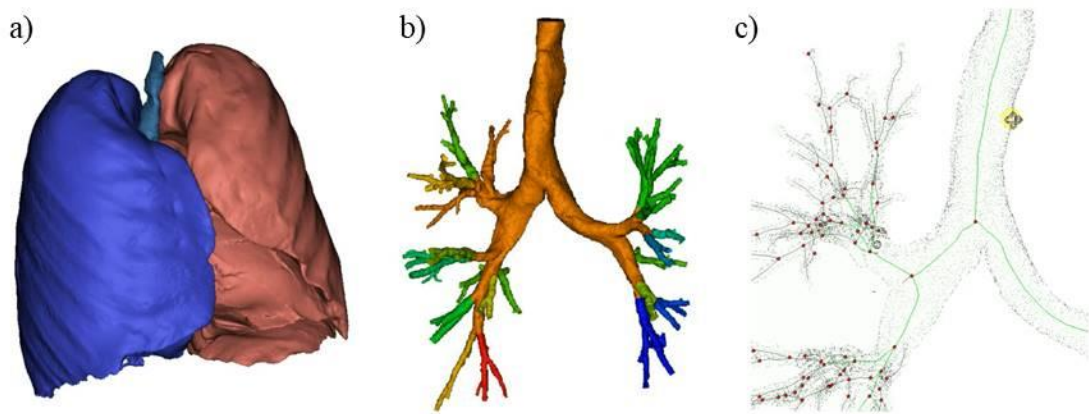
Figure 2-1: Imbio lung density analysis implementing PRM (I2E).



Imbio lung density analysisTM functional assessment report for a subject with significant functional small airways disease (fSAD) and emphysema as evaluated by PRM markers. Imbio automatically performed inspiration-to-expiration (I2E) image registration and voxel classification based on the point (-856,-950) in the expiration-inspiration HU plane as illustrated (bottom left component). I2E registered output was used in the study presented in section 3.1 of the thesis. N.B. ‘LDA’ here refers to ‘lung density analysis’, not to be confused with linear discriminant analysis.

Lobar and airway segmentation, of the trachea and major airways, up to the 6th generation on average, was historically performed using software application ‘mimics’, developed by Materialise NV (Leuven, Belgium), providing reliable 3D modelling with imaging data (271). Lobe surfaces were calculated as triangular meshes and saved in .stl format. Centrelines and branch bifurcation points were identified from upper airway volumes. See **Figure 2-2** for exemplars of lung unit and airway segmentation with centreline and branching point identification. The resultant airway centreline and surface mesh of the lobes were required inputs for extended centreline modelling (section 2.4.4). N.B. this work was performed in previous studies, the resultant data and in particular extended centreline models generated using major airway segmentations and lobar surfaces was used as input for workflows developed in this thesis.

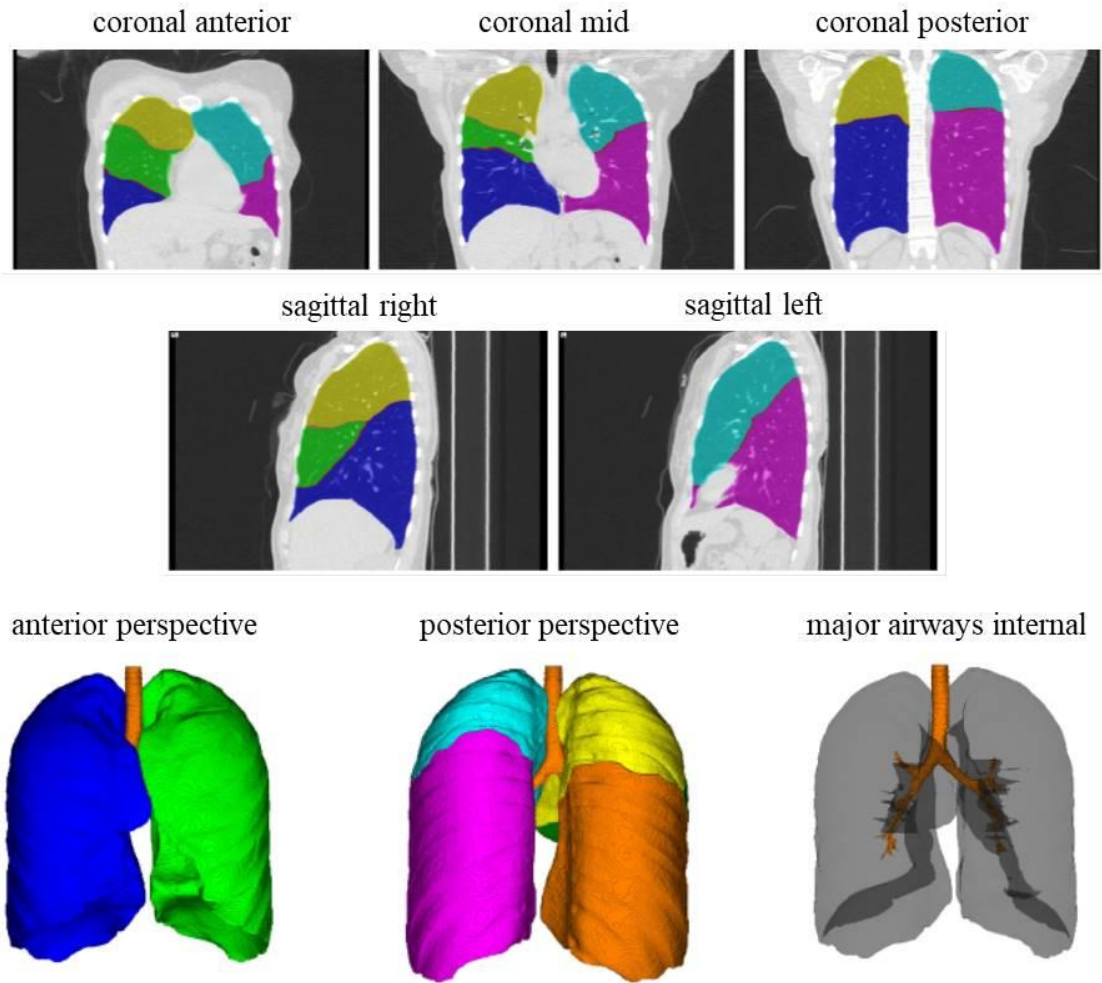
Figure 2-2: Lung volume segmentation and airway centreline computation using Mimics.



Graphical examples of VOI identification and segmentation from CT imaging data using Materialise Mimics software, used to calculate major volumes and lobar surfaces (a), major airways and terminal segmented branches (b) and airway centrelines (c) as inputs for extended airway centreline modelling that formed part of the basis for work presented in this thesis (sections 3.2 and 3.3).

To calculate expiration-to-inspiration (E2I) registrations (E2I not being facilitated by Imbio), independently developed CT image analysis software Yacta (yet another CT analyser) was used (264). Directories were batch processed for automated airway and lung volume segmentation. Sample of volume identification and visual output using Yacta is provided in **Figure 2-3**. Inspiration and expiration segmentations were calculated for all subjects with successfully simulated extended centreline structures.

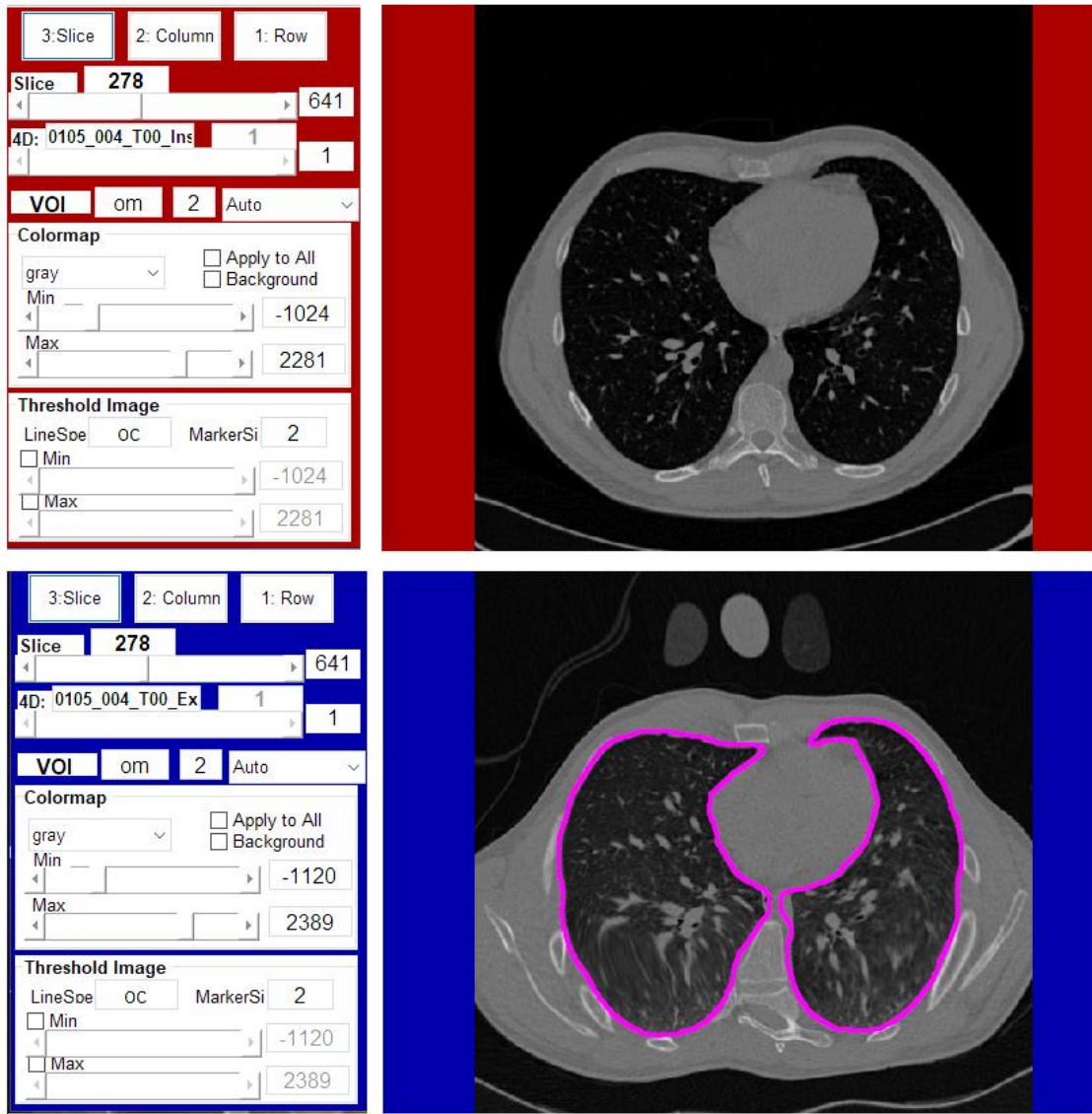
Figure 2-3: Yacta volume identification and segmentation with CT data.



Graphical output demonstrating lobar identification and volume segmentation, from CT data, using Yacta software. Lobes are colour coded in first two rows: yellow right upper lobe (RUL), green right middle lobe (RML), blue right lower lobe (RLL), cyan left upper lobe (LUL) and magenta left lower lobe (LLL). Output voxel lobe ID using Yacta was utilised in expiration-to-inspiration (E2I) image registration in section 3.3.

E2I image registration of segmented lung volumes was performed using custom Matlab software, capable of more general image registration tasks, developed by the centre for molecular modelling (CMI) group at the University of Michigan. Inspiration series were selected as the ‘fixed image’, and the corresponding expiration series as the ‘moving image’. Lobe labels were used to identify VOI and an initial affine transformation matrix was calculated. Registration was computed using Elastix (262). Post-registration QC/QA using the CMI application is exemplified in **Figure 2-4**.

Figure 2-4: Image registration using Matlab with CMI libraries (Michigan, USA).



*Demonstration of a typical workspace active whilst performing image registration, specifically post-registration QC/QA. Red interface (upper region) concerns the **fixed image** that is the VOI being registered to, inspiration (TLC) in this case. The corresponding blue interface (lower region) depicts the **moving image**, expiration (FRC) in this case; observe the warped profile, as this image has been geometrically mapped to fit the surface bounding the inspiration lung volumes (bright pink outline). It is required to check that the deformation computed matches the shape of the lungs in the fixed image (appears satisfactory in this example). The interface on the left side of both VOI controls visualisation of data, including viewing by slice (transverse plane), column (sagittal plane) and row (coronal plane); the default (slice) view is depicted.*

2.4.3 FOT Modelling

Computational calculations of lung impedance were performed through simulation of an electrical-circuit analogous model on patient-specific virtual conducting zone lung structures. Virtual structures were created in a prior study (207) through processing of inspiratory-expiratory CT scan images. From each scan, centrelines of the central airways (typically up to generation 6) were extracted, and lobar boundaries identified. An algorithmic airway generation process (201) was used to grow the remainder of the conducting zone (to an average generation of 16) within the defined lobar boundaries. See **Figure 2-5** illustrating simulation from major airway centreline to a complete conducting tree model in a test subject, using Lung Chaste data.

Figure 2-5: Extended centreline modelling from CT derived airways.

a) centreline of major airways.

b) complete conducting tree model.

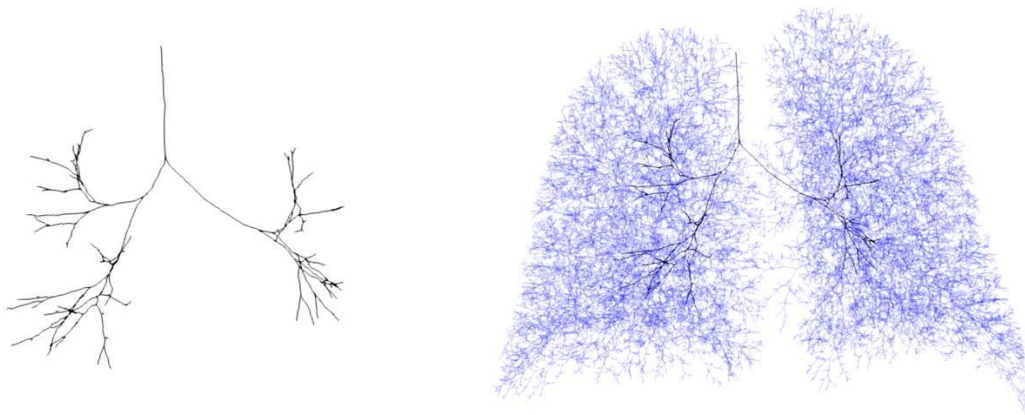


Illustration of airway centreline structures using ParaView (version 5.6.0-RC2). A, an initial centreline accurately derived from CT data up to an extent limited by image resolution (voxel size); typically of order 10^2 branches. B, overlay of simulated extended airway centrelines branching out from the initial structure, simulated branches are blue with 50% transparency; typically of order 10^4 branches.

Total lung impedance in complete airway tree structures was calculated by assuming each branch had associated impedance due to oscillatory flow, with each terminal bronchiole being subtended by a viscoelastic acinar unit. The impedance of each individual branch was based off mid-twentieth century works by Benade and Thurston (203,204), with the impedance of a branch j , denoted as Z_j , experiencing flow at a frequency f , was defined by the following equation:

$$Z_j = \left(i \frac{\omega \rho L}{\pi r_j^2} \right) (1 - F_v e^{i\phi r_j})^{-1}$$

where ρ is the air density, $\omega=2\pi f$, L and r are the branch length and radius respectively, and $i = \sqrt{-1}$ (imaginary unit). The exponential contribution is defined by:

$$F_v e^{i\phi r_j} = \frac{2}{r_v \sqrt{-i}} \frac{J_1(r_v \sqrt{-i})}{J_0(r_v \sqrt{-i})}, \quad r_v = \sqrt{\frac{\omega \rho}{\mu}} r$$

where r_v is the boundary layer thickness, μ is the air viscosity, and J_0 and J_1 are the zeroth and first Bessel functions respectively. The viscoelastic acinar units were described by a homogeneous constant-phase model of the form:

$$Z_{acin} = \frac{G - iH}{N \omega^\alpha}$$

G and H are coefficients for tissue damping and elastance respectively, taken as 0.12, and 0.57 kPa·s·L⁻¹ (207), N is the number of terminal bronchioles, and

$$\alpha = \frac{2}{\pi} \tan^{-1} \left(\frac{H}{G} \right)$$

Total impedance of the lung was calculated by adding series and parallel contributions from each acinar region and airway over the entire airway tree. Following the work of Bhatawedakar et al (272) this value was then added in series to chest wall, tracheal and glottal resistances and (all taken as 0.049 kPa·s·L⁻¹) and chest wall elastances (taken as 1.04 kPa·s·L⁻¹), and in parallel to a non-specific shunt impedance.

2.4.4 Set Alignment

It was required to sufficiently align two sets of points in \mathbb{R}^3 , a PRM registered CT voxel set of typical order 10^7 , and extended centreline point set of typical order 10^4 , to accurately associate SO proximity of CT voxels based on the computed complete conducting model tree. The following algorithm was developed to achieve this:

Let the CT voxel set be C and the model tree node set be N .

The first stage of alignment was deterministic and implemented as follows:

- 1) Mean centering. $N \rightarrow N - \bar{N}$, $C \rightarrow C - \bar{C}$, i.e. zero mean for both sets.
- 2) Rotation of $\pi/2$ applied to N . $(x, y, z) \rightarrow (y, -x, z)$.
- 3) Reflection of N in the plane $y = 0$. $(x, y, z) \rightarrow (x, -y, z)$.
- 4) Rescaling N to **match ranges** with C :

$$x \rightarrow \alpha_x \frac{\text{range}(x(C))}{\text{range}(x(N))} x + \beta_x$$

$$y \rightarrow \alpha_y \frac{\text{range}(y(C))}{\text{range}(y(N))} x + \beta_y$$

$$z \rightarrow \alpha_z \frac{\text{range}(z(C))}{\text{range}(z(N))} x + \beta_x$$

Where the vectors $\alpha = (\alpha_x, \alpha_y, \alpha_z)$ and $\beta = (\beta_x, \beta_y, \beta_z)$ parameterize magnitude of additional stretch and translation, parallel to coordinate axes, respectively. Initially $\alpha = (1,1,1)$ and $\beta = (0,0,0)$. Rotation and reflection settings were determined experimentally, and necessary given difference in output of Imbio lung density analysis (LDA tm) software, and materialize mimics segmentation software; i.e. post-hoc alignment using output from disparate software applications using distinct coordinate systems and linear transformations, but the same geometric input data.

The second stage of set alignment was determined by manual configuration of α and β using visualisation of the output from the first stage. Subsets of points from both sets were plotted with different colours, black and red, to appreciate accuracy of alignment from multiple perspectives (rotating 3D overlapping sets of points), particularly with respect to xy, xz and yz planes (informing choices for altering enlargement and translation of N). Alignment was deemed complete by the author when any point set misalignment appeared negligible. See **Figure 2-6** for example.

Figure 2-6: Manual alignment with visualisation of point sets.

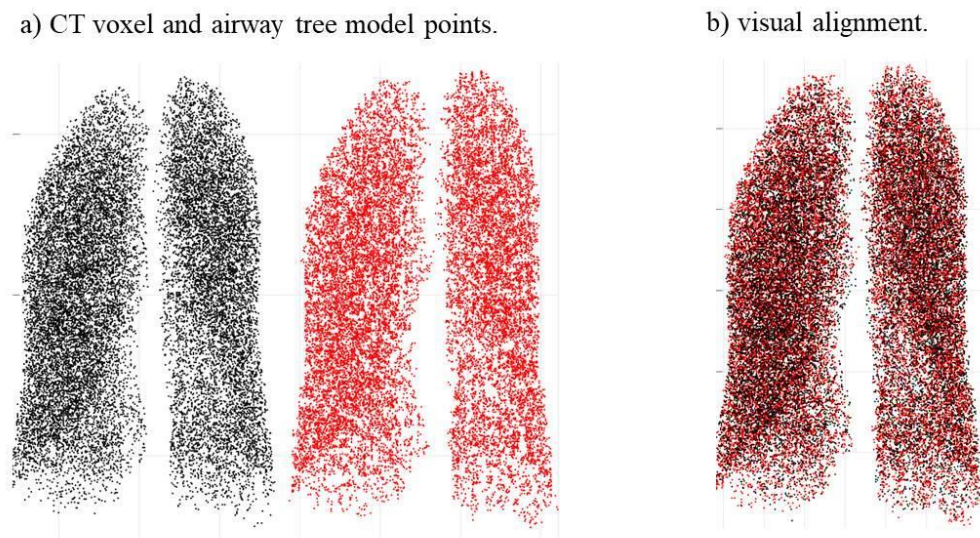


Illustration from the coronal perspective (y-z plane) of aligning disparate point sets generated from the same inspiration image. The algorithm randomly selects, without replacement, $2 \cdot 10^4$ points from a given set to plot. The black points are drawn from a segmented lung volume set (typical order 10^7) and the red from the associated extended airway centreline node set (typical order 10^4). Visualisation generated through varying β_y as described above (translating in the left-right direction).

2.5 Feature Extraction

In this section CT feature extraction and key data processing algorithms are described.

To discuss feature extraction, the following notation will be used:

- $L = \{v_i \mid i = 1, \dots, n\}$, denotes a ‘lung set’ L , of n voxels v_i .
- $iHU: L \rightarrow [-1000, -500] \subset \mathbb{Z}$, maps a voxel to its inspiration HU value.
- $eHU: L \rightarrow [-1000, -500] \subset \mathbb{Z}$, maps a voxel to its expiration HU value.
- $c: L \rightarrow \mathbb{R}^3$, maps a voxel v_i to its Cartesian coordinates (x_i, y_i, z_i) .
- $\Delta HU: L \rightarrow [-500, 500] \subset \mathbb{Z}$. $\Delta HU(v) = eHU(v) - iHU(v)$.

A total of 22 features summarising image registered CT data were developed and are listed in **Table 3.1-3** in section 3.1 for reference.

In addition to the listed features, processes for linking information between image registered voxels sets and simulated complete conducting airway tree structures are described, in order to assess SO proximity of lung voxels, i.e. indication of depth in airway tree, and functional evaluation of airway branches using ΔHU .

2.5.1 Features from Joint Density Histogram

PRM registered inspiration expiration CT features reported by Galbán et al (32) were derived from consideration of the inspiration and expiration paired voxel HU distributions. The typical Gaussian distribution can be well visualized in a joint density histogram (JDH); summarizing voxel concentrations in the eHU iHU plane, see **Figure 2-7** illustrating eHU iHU distribution with PRM quadrants and elliptical features. In this thesis, PRM voxel classification was determined by the following rule set:

IF $[iHU(v) \geq -950]$ AND $[eHU(v) \geq -856]$ THEN $[class(v) = PRM^{\text{Norm}}]$

IF $[iHU(v) \geq -950]$ AND $[eHU(v) < -856]$ THEN $[class(v) = PRM^{\text{fSAD}}]$

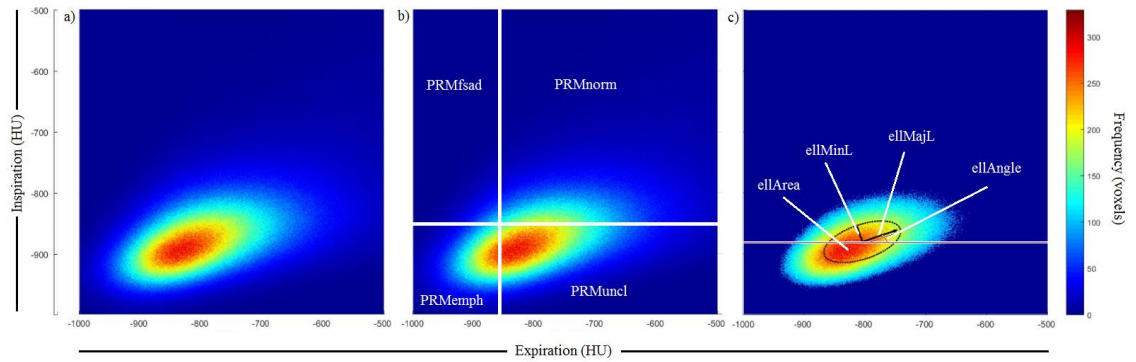
IF $[iHU(v) < -950]$ AND $[eHU(v) < -856]$ THEN $[class(v) = PRM^{\text{Emph}}]$

IF $[iHU(v) < -950]$ AND $[eHU(v) \geq -856]$ THEN $[class(v) = PRM^{\text{Uncl}}]$

Taking eHU as the horizontal axis, and iHU the vertical, then PRM^{Norm} , PRM^{fSAD} , PRM^{Emph} and PRM^{Uncl} classification relates to the 1st, 2nd, 3rd and 4th quadrants respectively of axes centered at (-856,-950) (see **Figure 2-7 B**), and superscripts abbreviate “normal”, “functional small airways disease”, “emphysema” and “unclassified” respectively. In this study, the features $\% \text{PRM}^s$ were defined to be the percentage of voxels classified as PRM^s , where $s \in \{\text{"Norm"}, \text{"fSAD"}, \text{"Emph"}, \text{"Uncl"}\}$. E.g. given a set $\{v_i | i=1, \dots, n\}$ of n voxels, $\% \text{PRM}^{\text{fSAD}} = \#\{v_i | \text{class}(v_i) = \text{PRM}^{\text{fSAD}}\} / n$, where $\#$ is the cardinality operator. $\% \text{PRM}^s$ was reported in decimal format.

JDH visualization typically demonstrates an approximate 2-dimensional Gaussian distribution, in which one may perceive an ellipse. The geometrical properties of this ellipse may thus form a means of describing the eHU iHU distribution. In this study, the properties chosen to approximate were minor axis length (ellMinL), major axis length (ellMajL), area (ellArea = $\pi \cdot \text{ellMinL} \cdot \text{ellMajL}$) and acute angle between the ellipse and the horizontal (ellAngle); see **Figure 2-7 C**.

Figure 2-7: Joint density histogram (JDH) of voxel HU with associated feature extraction.



Typical Gaussian distribution (2D) of a registered voxel set. A, raw distribution, elliptical shape indicates a wider expiration profile and narrower inspiration profile, indicative of tissue density increase on expiration (expected as gas vacates the volume). B, PRM classification based on quadrants centred on the point (-856,-950) as annotated. C, ellipse plotting and annotation of features extracted based on analysis of eigenvalues of the associated co-variance matrix; N.B. can see the background is deep blue (voxel frequency = 0) relative to A and B, as the algorithm removes voxels below a threshold of frequency to better isolate the ellipse arising from higher frequency cells.

To obtain ellipse based properties, the ellipse is first isolated by setting all cells in the JDH, with a voxel count of below $0.25 \cdot [\text{maximum cell voxel count}]$, to zero. The covariance matrix of the resulting voxel count distribution thus provides a description more localized to the ellipse perceived. Eigenvectors of this matrix provide representations of the major and minor axes, and so the area. To calculate ellAngle, the longer Eigenvector is identified (directed along major axis), then if necessary it is negated to obtain vector directed at positive quadrant. Then ellAngle is calculated from using the dot product theorem with the vector (1, 0), i.e. calculating the typically acute angle between the major axis and the horizontal. In this thesis angles are reported in radians, e.g. ellAngle = **0.13** radians $\leftrightarrow 0.13 \cdot 180/\pi = \mathbf{7.45^\circ}$ (to 2 d.p.). Code used in calculating ellipse vectors and plotting is provided in the appendices (**Appendix 1**).

2.5.2 Global Lung Features

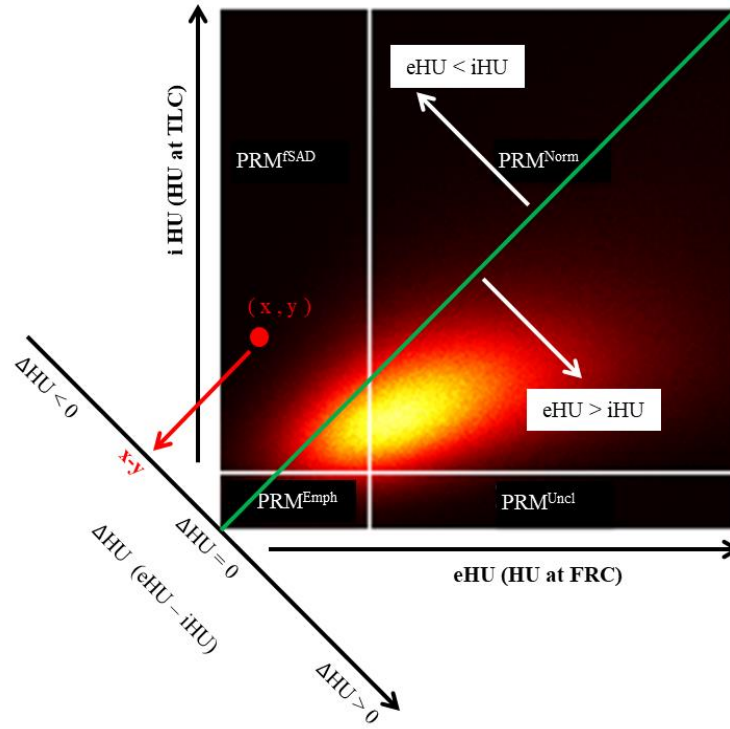
A number of simple features may be defined using just voxel counts and ranges.

- $vCnt(L) = n$, the number of voxels, a measure of total ‘lung tissue’ volume.
- $vCntX(L) = \text{range}(\{x_i\})$. $vCntY(L) = \text{range}(\{y_i\})$. $vCntZ(L) = \text{range}(\{z_i\})$.

Ranges of x, y and z coordinates provide some measure of anterior-posterior, lateral and inferior-superior lung (pair) dimensions respectively.

ΔHU as defined may be proportional to ventilation in a voxel, and has been previously studied as an immediate quantifier of ventilation behaviour (30,273,274). Mathematical relation between defined PRM voxels classifications and ΔHU function is elucidated in **Figure 2-8**. If a voxel at inspiration contains air, and at expiration has released this air, then the density, reported by HU (associated radiation absorption) should increase, i.e. $\Delta HU > 0$. It has been seen that this is true for most voxels in all subjects, and $\overline{\Delta HU} \gg 0$ on average (arithmetic mean). It should be noted however that perfusion and imaging artefacts are likely to affect ΔHU and add noise to the ventilation signal.

Figure 2-8: ΔHU as a projection from eHU iHU plane, and relation to PRM.



*Heatmap of JDH annotated with axes illustrating eHU iHU distribution, PRM classification (white lines), $\Delta HU = eHU - iHU$ (oblique line, $y = -x$) and line of no change ($y = x$) in green. A point in the eHU-iHU plane is mapped to ΔHU value, indicated in red. It can be seen that for most lung voxels $\Delta HU > 0$, and PRM^{fSAD} is most effected by voxels with negative ΔHU value. Image is taken from part of **Figure 3.1-11** in section 3.1 which further elucidates relationship between PRM and ΔHU with boxplots on upper and lower thirds of the study population based on SAA (section 2.5.3).*

In statistics, measures of central tendency and spread are two key characteristics that are of immediate interest in any distribution, and so naturally we define:

- $\overline{\Delta HU} : L \rightarrow \mathbb{R}. \quad \overline{\Delta HU}(L) = \frac{1}{n} \sum_1^n \Delta HU(v_i \in L).$
- $\text{std}(\Delta HU) : L \rightarrow \mathbb{R}. \quad \text{std}(\Delta HU)(L) = \sqrt{\frac{\sum_1^n (\Delta HU(v_i) - \overline{\Delta HU})^2}{n-1}}.$

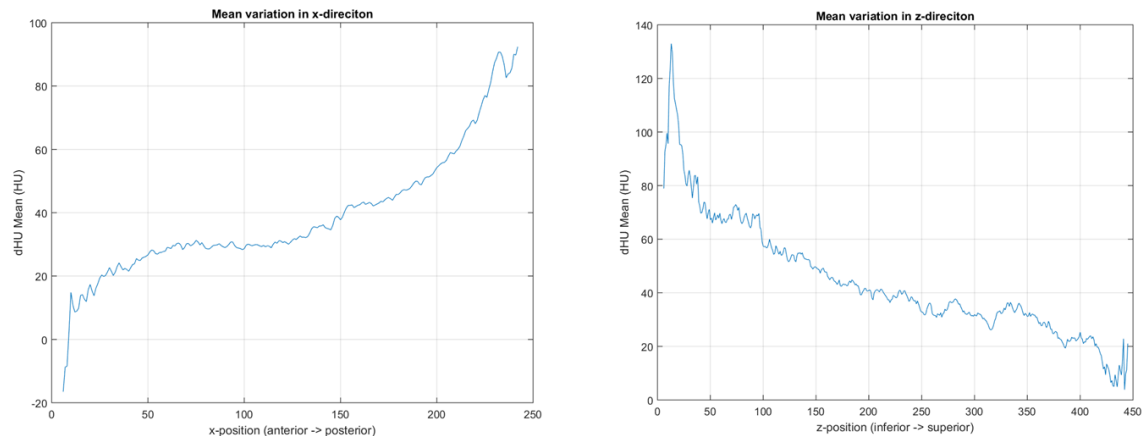
That is the mean and standard deviation (std) of change in HU, intended as a measure of average ventilation and variation in ventilation.

2.5.3 Stratified Axial Analysis

Stratified axial analysis (SAA) was developed to provide a basic tool for quantifying functional behaviour, approximated by ΔHU , as the lung is traversed along a cardinal (x, y or z) axis (see **Figures 3.1-7** and **3.1-8** in section 3.1). Voxels are stratified into groups using a co-ordinate distribution (x, y or z) and contiguous uniform intervals, then computing a functional average (e.g. $\overline{\Delta HU}$) for each stratified voxel group individually, and finally presenting a summary of the gradient for this distribution, i.e. on average, how the functional output changes across the defined strata.

This method was developed after observing ΔHU by slice, see **Figure 2-9**, where strong recurrent trends were apparent. Generalisation to n intervals, and n=10 in particular, followed to smooth the functional signal reducing sensitivity to noise; likewise the trimming of the ends of the distribution, particularly 15% and 10% of voxel ranges in the inferior-superior and anterior-posterior directions respectively, was implemented to control for sensitivity to sharp dips / peaks typically occurring at the extremes; this was likely due to low voxel counts and geometric variation at the apex and base of the lung.

Figure 2-9: Mean ΔHU in slices, anterior-posterior (x) and inferior-superior (z) directions.



Early development of SAA as plotting of slice by slice mean ΔHU in the indicated axes and directions. N.B. non-continuous behaviour prominent at the extremities, particularly in anterior slices of mean variation in the x-direction; this was the primary reason for excluding a defined portion of the poles in the final version of SAA implemented in this thesis. N.B. $dHU = \Delta HU$.

To describe the process in practice using 10 intervals, with some operational intricacies, consider L the entire set of voxels for a pair of lungs. Then to stratify in the z axis (see **Figure 3.1-7**), we used the range = $\max(z_i) - \min(z_i)$. A portion of the strata at the poles is eliminated, to reduce noise at the ends of the distribution of interest. 15% of r , of the extremes, are trimmed, retaining the middle 70% r for the remaining steps. This trimming helps deal with low voxel count at the extremities, difference in lung heights, and beam hardening effects in the CT image, common to the conical base and apex of the lungs. The remaining voxels are then split into 10 intervals, of roughly 7% r each, with 9 intermediary points termed ‘deciles’. One may consider 10 values determined by the function ΔHU on each interval, providing a smoothed indicator of the average ‘ventilation’ at each level. Then we take the differences at the deciles, with direction chosen to be ‘inferior-to-superior’ (by ordering in subtraction).

To describe how $\overline{\Delta HU}$ varies across the intervals, we initially used fitting of a 1 degree polynomial, i.e. straight line, using ordinary least squares (OLS) criterion. The gradient of the resultant line provides a natural measure of change, so we define 4 SAA based attributes as follows:

- $\overline{\Delta HU}^{AP}$: gradient measuring $\overline{\Delta HU}$ change in the **anterior-posterior** direction.
- $\text{std}(\Delta HU)^{AP}$: gradient of $\text{std}(\Delta HU)$ change in the **anterior-posterior** direction.
- $\overline{\Delta HU}^{IS}$: gradient measuring $\overline{\Delta HU}$ change, in the **inferior-superior** direction.
- $\text{std}(\Delta HU)^{IS}$: gradient of $\text{std}(\Delta HU)$ change, in the **inferior-superior** direction.

Lateral $\overline{\Delta HU}$ gradient across the lungs was not investigated. As it became clear that inferior-superior ventilation gradient was strongly linked to VH, another method of measuring gradient was applied to this axis, focusing on $\overline{\Delta HU}$. Termed $\overline{\Delta HU}^{IS*}$, this simply takes the mean of the decile changes (see **Figure 3.1-7 B**), as follows:

$$\bullet \quad \overline{\Delta HU}^{IS*} = \frac{1}{9} \sum_{i=1}^9 (\overline{\Delta HU}(I_{i+1}) - \overline{\Delta HU}(I_i)) = \frac{\overline{\Delta HU}(I_{10}) - \overline{\Delta HU}(I_1)}{9}$$

Where I_i is the i^{th} interval, travelling from inferior to superior as i ranges from 1 to 10. This happens to be equivalent to a scaled difference between the extreme intervals. This measurement was found to have marginally higher correlational strength, and so

possibly a cleaner signal (smoothing measurement to extremes of post-trimmed intervals), relative to $\overline{\Delta HU}^{IS}$.

The described algorithm was implemented in Matlab for the inferior to superior and anterior to posterior axes. Matrices of eHU, iHU, x, y and z were permuted using sortrows.m on the column for the chosen axis. Start and end points were calculated using the floor of the exact values calculated for trimming, i.e. minimum coordinate value plus trim length and maximum coordinate value minus trim length. Interval dividing points were calculated using the floor of points generated by linspace.m using the given start and end points, and 11 equally spaced separating points i.e. 9 interior points dividing the intervals. $\overline{\Delta HU}$ and $\text{std}(\Delta HU)$ averages over the 10 intervals were fitted using polyfit.m, and $\overline{\Delta HU}^{IS*}$ as the mean of the mean-differences. Code implementing SAA is provided in the appendices (**Appendix 2**).

2.5.4 Inter-Lung Features

Let the voxels belonging to the left lung be the voxel set L_{left} , and likewise for the right lung, L_{right} , such that $L_{\text{left}} \cap L_{\text{right}} = \emptyset$, $L_{\text{left}} \cup L_{\text{right}} = L$. In this thesis 2-means clustering was applied with centroids initiated at $(\bar{x}, \bar{y} \pm 2\text{std}(y), \bar{z})$, to approximate L_{left} and L_{right} from the given raw data (all points for both lungs in one data matrix).

Using the described ‘ ΔHU ’ functions, we define:

- $\text{RLmeanDiff} : L \rightarrow \mathbb{R}$, $\text{RLmeanDiff}(L) = |\overline{\Delta HU}(L_{\text{left}}) - \overline{\Delta HU}(L_{\text{right}})|$.
- $\text{RLstdDiff} : L \rightarrow \mathbb{R}$, $\text{RLstdDiff}(L) = |\text{std}(\Delta HU)(L_{\text{left}}) - \text{std}(\Delta HU)(L_{\text{right}})|$.

That is, measuring the difference in average ventilation, and difference in variation of ventilation, between the two lung voxel sets. An additional feature is provided through considering the difference in lung sizes, RLsizeRat , read as ‘right-left size ratio’.

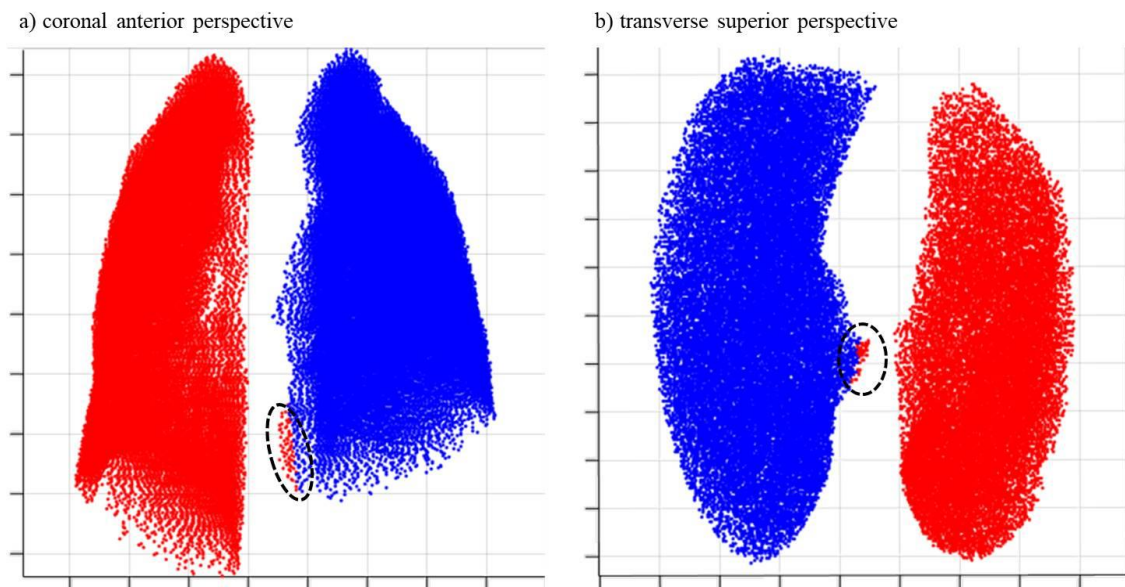
- $\text{RLsizeRat} : L \rightarrow [0,1) \subset \mathbb{R}_{\geq 0}$.

$$\text{RLsizeRat}(L) = 1 - \min \left\{ \frac{vCnt(L_{\text{left}})}{vCnt(L_{\text{right}})}, \frac{vCnt(L_{\text{right}})}{vCnt(L_{\text{left}})} \right\}.$$

As defined it constitutes a metric function meant to measure the extent to which lungs differ in size. A value of 0 can be achieved only if the lungs are equal in size (biologically abnormal); else the value tends to 1 as the difference in size increases. Code implementing the described algorithm on lungs partitioned into voxel sets L1 and L2 by the described 2-means process is included in the appendices (**Appendix 3**).

The heuristic approach of splitting by 2-means with centroids separated about the mean point in the y (left-right) direction by two standard deviations was implemented and monitored using a basic plotting script that used the kmeans.m function, see code provide in appendices (**Appendix 4**), and graphically displayed results whilst returning L1 and L2 matrices to the above script. Imperfect splitting was observed occasionally, see plots in **Figure 2-10**, however appeared relatively negligible in QC/QA work.

Figure 2-10: Visual evaluation of 2-means lung identification.



Visual plotting of points on coronal (a) and transverse (b) planes from the direction indicated, colour coded based on cluster ID assignment from k-means clustering with $k=2$ and centroids initiated from the mean point $(\bar{x}, \bar{y}, \bar{z})$ plus or minus two standard deviations of the y distribution in the y-direction (left-right axis); configured to optimise probability of non-erroneous left-right lung identification. This heuristic approach included some error as annotated by the dashed ovals.

2.5.5 Voxel SO Score

To assign an SO score to a voxel, a k-NN based algorithm was implemented with $k = 1, 5$ and 10 . Given a voxel v , and k nearest (Euclidean distance) nodes in the aligned simulated tree node set, $N_k(v) = \{n_1, \dots, n_k\}$, the SO score of v was defined to be the arithmetic mean of the associated SOs of N , i.e.:

$$SO(v) = SO(N_k(v)) = \frac{1}{k} \sum_{i=1}^k SO(n_i)$$

Where $SO(n_i)$ is the SO of the branch for which n_i is the **distal node**. N.B. when $k=1$ this is simply the Strahler order associated with the nearest node. This function is meant to provide a measure of proximity of a voxel to the smaller airways and parenchyma. The idea is illustrated in **Figure 3.2-9** in section 3.2, where the score for a particular voxel (highlighted yellow) is calculated for $k=1$ and $k=3$. The algorithm was implemented in Matlab using `knnsearch.m`, and illustrated in code provided in the appendices (**Appendix 5**).

2.5.6 Branch ΔHU assignment

In the previous section information from the tree structures, Strahler order, was used to assign a meaningful value to CT voxels; the reverse is now described, where information from the set of registered voxels, in particular ΔHU , is used to assign a meaningful value to branches in the tree structure. The idea is illustrated in **Figure 3.3-2** in section 3.3; voxels within a certain distance, r , of the terminal branch nodes are identified and the average of their ΔHU is assigned to that terminal branch. This gives a measure of lung function, inferred from HU changes, to every distal branch in the tree structure. Upper branches are then assigned a ΔHU score by averaging the ΔHU of their connected lower branches in consecutive SO from 2 to the maximum assigned value (variable between subjects).

The value of r used in this study was $r = 5 \cdot \sqrt{3}/2 + 0.1$. The $\sqrt{3}/2$ is used as a natural factor given it is the distance from the centre of the unit cube to any of its corners, and

the set of CT voxel positional coordinates is given as an integer lattice, i.e. this is the smallest distance required to ensure at least one CT voxel in a filled grid will fall within the boundary of the scribed sphere of radius r . The 0.1 provides a buffer in favour of capturing points that fall on the edge of the sphere, and the 5 was determined adequate from experiment (see section 2.6.4).

A problem that arises from this direction of information assignment is that now a space of order 10^7 points must be searched for proximity, rather than the 10^4 in the previous section; a task that is computationally very difficult for any simple approach. To solve this problem a cell array method previously described (198) was implemented, see **Figure 2-11**. Points from the larger space of CT voxels were classified into cells, effectively cubes of side length r , with a function f mapping points to cell integral coordinates derived. This f could then be used on terminal branch points of the tree structure to identify the cell containing the distal point of a branch, and thus isolate at most 27 (3^3) cells to search for points lying within distance r . This greatly decreases computational cost of ΔHU assignment and was a key in making this algorithm feasible. Code for efficiently identifying proximal points to average over for ΔHU is illustrated in the appendices (**Appendix 6**).

Figure 2-11: Cell array assignment for efficient checking of point proximity sets.

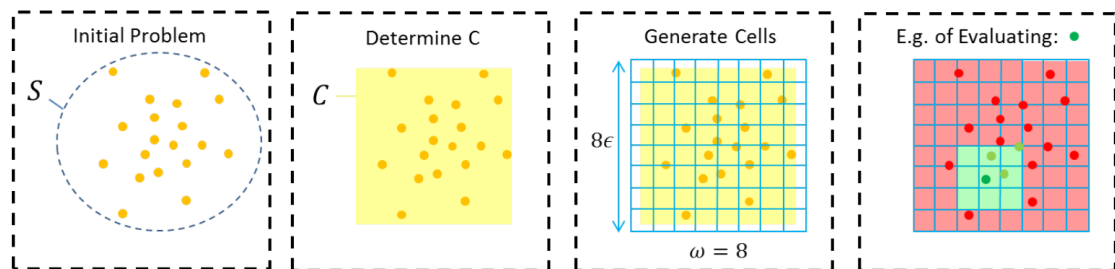


Illustration of cell array generation, point assignment and checking a particular point for neighbours within a given distance, where: S = set of points to evaluate, C = domain to cover with cell array (e.g. smallest cuboid containing all points), ϵ (epsilon) is the critical distance under analysis and ω is the number of cells on a side of the resulting cell array grid. This method allows for rapid checking of points proximal to a given point within a very large point set. N.B. operates at peak efficiency in a homogenous point distribution, which appears sufficiently satisfied in this study.

2.6 Analysis

In this final section, statistical methods are described with details of implementation and reference to code in the appendices, covering: group comparisons, correlation analysis, box plotting, tertile polar analysis, feature and feature set relational strength evaluation through i) correlation, ii) linear regression including PCA, and iii) linear discriminant analysis (LDA), joint density histogram definition and implementation, and clustering applied to nasal transcriptomics. Then following statistical methods: min-max projection of voxel point sets, topological data analysis using Ayasdi software, complete conducting tree structure function visualisation and FOT modelling analysis with respect to airway adjustment in direct proportion to HU changes (Δ HU).

2.6.1 Statistical Methods

Standard statistical testing was implemented using custom scripts written in Matlab with built in functions and structural elements to automate analysis and output. In all tests with an associated p-value, significance was defined as $p < 0.05$; i.e. the standard 5% significance level was used in all cases.

2.6.1.1 Group Comparisons

Binary group comparisons were performed using built-in function 'kstest.m' (Kolmogorov-Smirnov), followed by 'ttest2.m' (two-sample t-test) if attribute was determined parametric, and 'ranksum.m' (equivalent to Mann-Whitney U-test) otherwise. Multiple group comparison (for 3 or more groups) was computed using built-in function 'kstest.m' (Kolmogorov-Smirnov), followed by 'anova1.m' (one-way ANOVA) if attribute was determined parametric, and 'kruskalwallis.m' (Kruskal-Wallis) otherwise. The stats output variable from multiple group comparison testing was passed to the 'multcompare.m' function, with default Tukey's honest significant difference criterion. Output from 'multcompare.m' was used to define significance in tables and text presenting group comparisons. All multiple group comparisons were performed on groups with empty intersections (none overlapping).

Code automating the described statistical analysis scheme was developed in Matlab and exemplified concisely in the appendices (**Appendix 7**). To compare ratios of a boolean

variable across groups, a 3rd party script ‘prop_test.m’¹ was utilized to implement a simple two-sample Chi-squared test of proportions.

Tertile (dividing into three) polar (extreme thirds) analysis (see section 2.6.1.4 anon) was performed on intervals in SAA, involving by design the formation of groups with an artificial distribution (extreme tertiles taken from a given natural distribution). In this instance, comparisons were performed pairwise on the two groups formed from extremes, across all intervals or deciles, using the ‘ranksum.m’ function. A custom script was written to generate plots of this statistical analysis of SAA and incorporated alongside ranksum.m testing, code exemplified in appendices (**Appendix 8**).

Resultant plots were the basis for graphics such as **Figure 3.1-3 A** and **Figure 3.1-10** in section 3.1, providing a detailed visual and statistical analysis of SAA outcome across multiple strata of the lung and subjects within the study population.

2.6.1.2 Correlation Analysis

Pearson correlation coefficient was calculated using the ‘corrcoef.m’ and ‘corr.m’ functions. A script was written to analyse absolute correlation strength above a given threshold, over all pairs of a given set of attributes. In cases of missing data, correlation calculation was restricted to subjects for which data was present. Results for attribute pairs above the threshold were studied by way of scatter plot with graphed line of best fit, and table listing signed correlation coefficient, p-value, and index for attributes.

2.6.1.3 Box plots

Box plots and annotations of healthy and asthmatic subjects were created using custom script including the ‘boxplot.m’ function. In all cases boxplots present differentiation of median split groups for feature included with approximate median value on the horizontal axis. E.g. see **Figure 3.1-1 C**. Outliers with respect to split groups are highlighted with black spots, e.g. see **Figure 3.1-9**.

¹ <https://uk.mathworks.com/matlabcentral/fileexchange/45966-compare-two-proportions--chi-square-?focused=3813016&tab=function>

2.6.1.4 Tertile Polar Analysis

Given noisy data it can be informative to investigate attribute extremes i.e. compare subjects from the tails of a distribution for a significant difference in another feature. In the first study presented in this thesis, section 3.1, the highest and lowest 16 values from an attribute distribution were selected for comparison, being approximately distribution tertile poles, since $16/52 \approx 1/3$, balancing isolation of signal at the extremes with preservation of numbers for statistical significance. This analysis was applied to relevant attributes, particularly VH markers S_{acin} and R5-R20, to generate groups for comparison in $\overline{\Delta\text{HU}}^{\text{IS}*}$ and $\overline{\Delta\text{HU}}^{\text{AP}}$ interval-wise and decile-wise analysis, i.e. comparing mean ΔHU and inter-interval ΔHU differences respectively.

2.6.1.5 Feature and Feature Set Relational Strength Analysis

In order to determine if two variables, or more generally two sets of variables, are ‘related’, a well understood and standard approach is to look at linear correlation and discriminatory properties. This approach is limited in the sense that not all relationships may be linear, though many important relationships between variables in nature are. Correlation and linear regression modelling was utilized to study how CT data, represented by a set of 22 features, related to single non-CT attributes, and linear discriminant analysis (LDA), see section 2.6.1.9, to test the strength of different feature sets in VH discrimination, given an optimal projection to one dimension.

2.6.1.6 Average Correlation

Given a data matrix $X_{n \times m}$ of m attributes over n subjects, it was desired to see how some submatrix X_s , formed from a selection of columns from X in some order, compares to single attributes from $X \setminus X_s$. Specifically, the 22 CT attributes defined in this thesis formed the submatrix X_{CT} , and attributes from spirometry, IOS, MBW and GINA were chosen as target variables, single attributes in $X \setminus X_{CT}$ (X being the complete data set analysed in section 3.1). Then the average absolute correlation is defined:

$$\bar{r} = \frac{1}{22} \sum_1^{22} |\text{corr}(x, y)|$$

where $x \in X_{CT}$, y is a target variable and ‘corr’ is the Pearson correlation coefficient taken over all subjects. This provides one of the simplest though apparently prognostic measures of linear relation. Simple rationale is that if there exists a large number of highly correlating (absolute value) CT variables with a variable y , then y is strongly related (linearly) to ‘CT data’ (as represented by the given feature set). Results of this approach are presented in the first column of **Table 3.1-7** in section 3.1.

2.6.1.7 Multicollinearity Limited Subset Linear Regression

Linear regression was utilized to test likelihood (F-statistic) and strength (R^2 , variance explained) of linear relation between predictor variables and some outcome variable. In this study CT based predictor variables were used to predict an outcome variable from non-CT attributes. However, the problem of multicollinearity, that is the existence of strong absolute correlation between input variables begetting an ill posed prediction model, needed to be handled to reduce likelihood of spurious results.

One method supposed to limit multicollinearity impact, whilst preserving original (unaltered) data values for predictor variables, is to select a ‘representative’ subset of features, eliminating the occurrence of absolute pairwise correlation above some threshold. An ad hoc approach to feature selection was utilized, whereby data analyst experience and consistency lead decision produced a 14 feature subset (from 22 features) in which no absolute pairwise correlation exceeded 0.7, see correlation matrices in **Figure 3.1-12**. Specifically: **ellMajL** summarized itself, $\overline{\Delta HU}$ and $\text{std}(\Delta HU)$; **ellMinL** summarized itself and **ellArea**; **RLmeanDiff**, **RLstdDiff** and **RLsizeRat** lacked intra-CT correlation (thus were all included); **PRM^{Emph}** summarized **PRM_s** attributes; **vCntX**, **vCntY** and **vCntZ** all appeared to lack intra-CT correlation, whilst **vCnt** strongly correlated with **vCntZ**; All OLS based SAA features, $\overline{\Delta HU}^{AP}$, $\text{std}(\Delta HU)^{AP}$, $\overline{\Delta HU}^{IS}$ and $\text{std}(\Delta HU)^{IS}$, lacked intra-CT correlation, and $\overline{\Delta HU}^{IS}$ was chosen to summarize $\overline{\Delta HU}^{IS*}$ for consistency. Results of the collinearity reduced linear regression are presented in **Table 3.1-7**, second column.

2.6.1.8 PCA Kaiser Rule Linear Regression

Though given some rationality, and the arguably preferable aspect of using original data values, the described multicollinearity limited subset method can be highly sensitive to chosen ‘representatives’, and loss of information from eliminating attributes. PCA provides another approach to eliminating multicollinearity within data, since by design it produces orthogonal representative coordinates, principal components that have correlation strength zero between any pair.

In this study PCA was used to provide a low dimensional approximation, by principal components, to X_{CT} , the data matrix of defined CT features. Applying the ‘Kaiser rule’, the first principal components were selected, as these had a variance explained greater than the average variance explained across all components. These 6 components were then used as predictor variables in linear regression predicting the outcome variables described. Results presented in **Table 3.1-7**, third column. One may consider the PCA Kaiser rule (PCA-KR) approach as being a relatively less biased representation of linear strength. Results of both approaches are presented for observation.

2.6.1.9 Linear Discriminant Analysis

In order to assess ‘linear relational strength’ between sets of features, representing a more general object (e.g. ‘CT data’, ‘spirometry’), linear discriminant analysis (LDA) was used to determine the best linear discriminator between two sets of points, according to Fisher’s criteria of maximal inter-class mean separation with minimal intra-class variance. Specifically a projection vector \mathbf{p} , mapping data points to a line with optimal class separation, is sought which minimises the ‘Rayleigh quotient’:

$$R = \frac{\mathbf{p}^T S_B \mathbf{p}}{\mathbf{p}^T S_W \mathbf{p}}$$

Where $S_B = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T$ is the ‘between-class scatter matrix’, μ_i the arithmetic mean of the i^{th} class, $S_W = \sum_{i \in C_1} (\mathbf{x}_i - \boldsymbol{\mu}_1)(\mathbf{x}_i - \boldsymbol{\mu}_1)^T + \sum_{i \in C_2} (\mathbf{x}_i - \boldsymbol{\mu}_2)(\mathbf{x}_i - \boldsymbol{\mu}_2)^T$ is the ‘within-class scatter matrix’, \mathbf{x}_i the i^{th} data point, C_i the set of indexes for the i^{th} class, and T is vector transposition. The solution for \mathbf{p} is obtained by matrix inversion: $\mathbf{p} = S_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$.

LDA was implemented using custom Matlab script, written to output histograms of projected value distributions for two classes, determined from splitting a population on mean of VH marker S_{acin} or R5-R20, and determine classification accuracy from best (least error) of one dimensional point of discrimination. Coefficients of LDA across all selected features were reported (**Table 3.1-11**). Simplified code is illustrated in the appendices (**Appendix 9**).

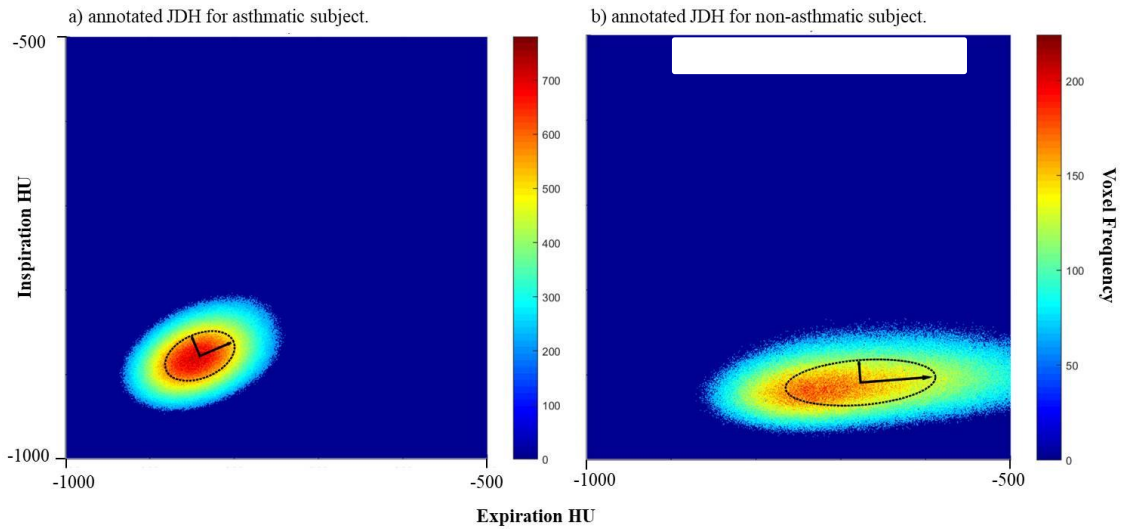
2.6.1.10 Joint Density Histogram (JDH) Visualisation

JDH visualization was generated in Matlab using the 'surf.m' function with re-specified tick locations and labels, applied to a 2-dimensional array with dimensions determined by the floor of the range of the associated distribution (i.e. bin number for binning point counts with respect to a variable set equal to rounded range of that variable). Colormaps 'hot' (e.g. **Figure 3.1-11**), to create black and white compatible presentation, and 'jet' (e.g. **Figure 2-12**) for higher resolution overview of variability, were applied.

Eigenvectors and Eigenvalues of the 2D array were utilized to determine ellipse axes fitting dual Gaussian distribution, plotted with vectors directed outward from mean point using the square root of the associated Eigenvalue to determine length. See **Figure 2-12** below for illustration of disparate JDH plots, in an asthmatic and a non-asthmatic subject, with elliptical features annotated.

Plotting of the ellipse (black) annotating the JDH plot was achieved using the third party function 'ellipse.m' written by D.G. Long, Brigham Young University, based on the CIRCLES.m original written by Peter Blattner, Institute of Microtechnology, University of Neuchatel, Switzerland. The addition of vectors indicating minor and major axes with direction indicated by arrow head (QC/QA for ellipse features) was implemented using the built-in 'quiver3.m' function.

Figure 2-12: JDH ellipse approximation in disparate example cases.



*Examples of ellipse plotting using written algorithm with vectors indicating minor and major axes in two subjects, one asthmatic and one healthy control, showing significant difference in JDH profile. The distribution for the asthmatic subject (left) appears relatively **compact** and circular, indicative of little difference between inspiration and expiration states (i.e. limited density change due to poor ventilation); the distribution for the non-asthmatic subject (right) appears stretched in the horizontal (expiration) axis, indicative of a much greater density increase at expiration relative to inspiration (i.e. large density increase due to large volume of gas expired).*

2.6.1.11 Hierarchical Clustering / Transcriptomics Analysis

To investigate association between nasal gene expression and inferior-to-superior HU change gradient marker $\overline{\Delta HU}^{IS*}$, a three step approach was implemented in study 1: i) hierarchical single linkage agglomerative clustering was applied to subjects with successful nasal scraping samples (45 out of 52) based on correlation of z-score normalized set of 13,400 genes, ii) Wilcoxon ranksum testing to isolate genes with most significant difference in means between selected subject groups, and iii) 30 genes with largest absolute difference of group means listed for observation. Raw gene array data contained 25,000 attributes including artefacts from post-processing, and was subjected to a cleaning algorithm removing erroneous and duplicated data vectors, reducing the number of gene features to 13,400.

This algorithm was applied using Matlab, with built-in function ‘clustergram.m’, on the raw data matrix of 45 subjects and 13,400 gene attributes, generating an initial dendrogram and clusters. Clusters with subjects occupying extreme ends of the $\overline{\Delta H U}^{IS*}$ distribution were selected for gene-wise mean comparison using Wilcoxon ranksum testing; the lowest p-value(s) were then selected, satisfying the criterion that resulting set has $n > 29$. Finally, the 30 genes with largest absolute mean difference were selected for presentation and investigation (see **Table 3.1-12**). Step-by-step visualisation and the results of this process are provided in **Figure 3.1-13** of the first study (section 3.1).

2.6.2 *Min-max Projection*

A simple method to perceive information based on high volume point cloud distributions in three dimensions is ‘min-max projection’, that is to project the distribution to some 2D plane, binning 2D cells to count point frequency and visualize using a heatmap; i.e. essentially applying the JDH algorithm previously described to some 2D projection of a 3-dimensional distribution.

To visualize functional averages in space, replace ‘point frequency’ with $\sum f(x)$, where x varies over all points in a cell. Then mass bias is cancelled by averaging, dividing by the number of points in the cell. In this study, such mass normalized min-max projection was applied in the coronal plane and transverse planes to visualize $\overline{\Delta H U}$ and PRM^{fSAD} distributions across the entire lung. A sample of min-max projection, in one subject, with ‘jet’ colormap of PRM^{fSAD} is provided in **Figure 2-13**.

2.6.3 *Topological Data Analysis*

A significant component of this work is a ‘novel mathematical approach’ to data analysis using the topological data analysis (TDA) algorithm ‘mapper’. This project collaborated with the US company ‘Ayasdi’ which provides hardware and software for applying this methodology. It is inspired by the mantra that ‘data has shape’, and that ‘shape has meaning’. It is part of a wider mind-set that, given a set of points in space, one may infer a geometric structure on those points, with properties that may then provide useful information with respect to that data. E.g. inferring an ellipse on the JDH (**Figure 2-12**) and measuring its properties, though mathematically it can be stated no actual ellipse exists in the original data, being precisely just a cloud of points in space.

The mapper algorithm was designed for application in the analysis of high dimensional data sets and has been described as a generalisation of the Reeb graph (255). An illustration of the Reeb graph of the standard torus, using the ‘height function’, is illustrated in **Figure 2-14**. The idea is that though information is necessarily lost in the projection, important features such as homology (e.g. number of holes) are preserved and apparent in the compressed representation, i.e. you can see the ‘hole’ in the graph. This is of course dependent upon the selection of the function used in projection.

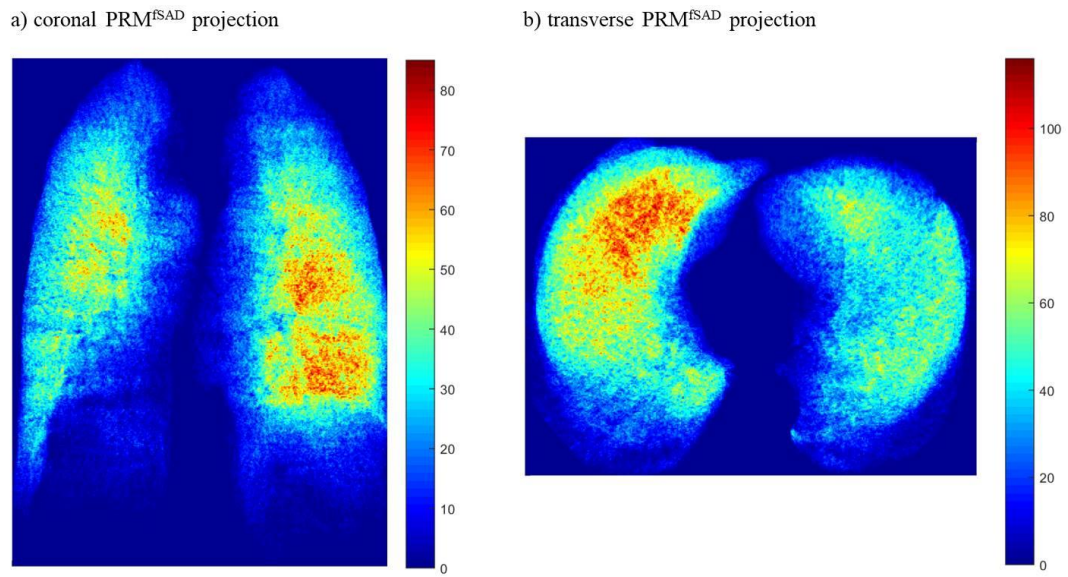
Typically TDA relies on the construction of a simplicial complex, by some method (mapper is essentially an algorithm to accomplish this). A simplicial complex, or just complex, is a combinatorial object listing simplexes. A simplex is, roughly speaking, a ‘simple component’ of some dimension, from which a more complex shape is constructed. A full description of simplicial complexes is beyond the scope of this thesis. A sufficient framework to understand the work is that 0-simplexes are ‘points’, and 1-simplexes are ‘lines’, and a collection of points and lines may provide a compressed representation of a more complex shape. E.g. any polygon may be decomposed into points and lines.

Figure 2-15 provides an exemplified overview of a process used in mapper.

Conceptually, it begins by associating a data matrix with a set of points in Euclidean space (common idea from early education). Though there is no shape in the sense of an infinite set of points equating to a polygon/polyhedron, the points may be thought to ‘approximate’ a geometric figure, for example, and by analogy, figures have historically been inferred in celestial constellations. This idea is illustrated in **Figure 2-15 A**.

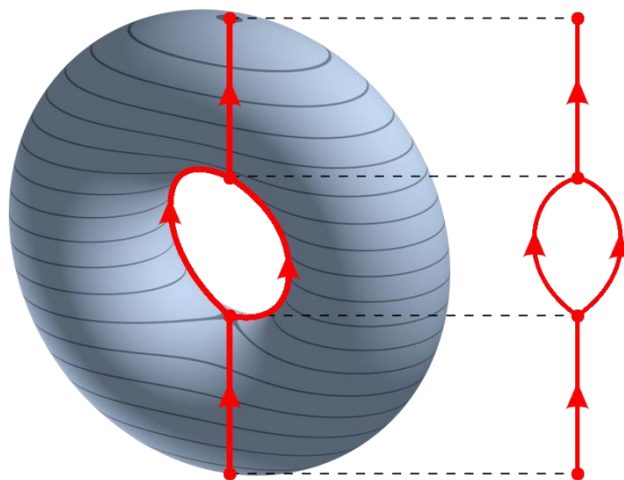
To generate a complex representative of the shape in space, a function is required to encode ‘**similarity**’. This is referred to as a ‘metric’ by Ayasdi, as similarity is closely related to the idea of ‘closeness’, and metric by definition attempts to capture this idea. Another function, and generally a set of functions, from points to the real numbers (for example), provides the framework for sensibly splitting up the space of points into groups, each group having an image of the real interval that may overlap with other images, and so distinct groups may contain common points (**Figure 3-15 B**).

Figure 2-13: Coronal and transverse plane min-max projection of PRM^{fSAD} in one subject.



*Min-max projections of voxels classified as PRM^{fSAD} across the entire voxel set onto the coronal (y-z) and transverse (x-y) planes in one subject, selected to illustrate apparent $fSAD$ concentration. Colormap 'jet' was applied in this case, whilst 'hot' was applied in section 3.1, e.g. **Figures 3.1-3 B C** and **3.1-4**. Colorbar annotated shows cell voxel frequency. This method illustrates a simple approach to visualising functional information from an entire 3D point set (projection from 3D to 2D).*

Figure 2-14: Graphical representation of Reeb graph calculation from the torus.



Reeb graph calculated from the torus with respect to the height function. Image reused from: https://en.wikipedia.org/wiki/Reeb_graph, from the public domain.

Figure 2-15: TDA mapper from first principles to k-NN based output.

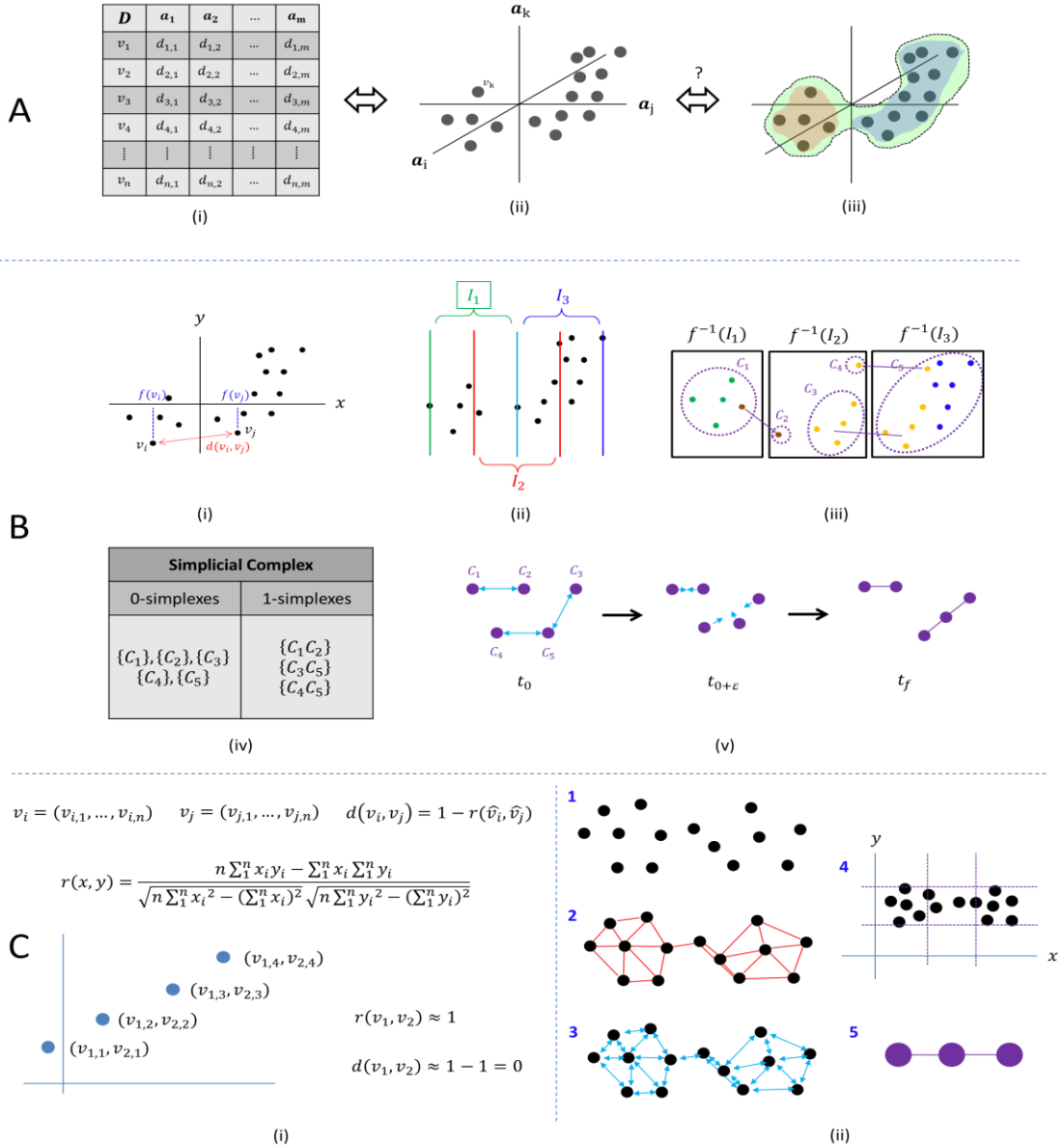


Illustration of TDA mapper motivation and ideology. A, a data matrix is associated with a set of points which in turn may be seen to approximate shape(s) within the data. B, mapper applied with x -coordinate function on exemplar set of points with three intervals (resolution=3) and an overlap of 50% (gain=2); clustering produces node groups, groups sharing at least one point are joined by lines. C, left (i), demonstration of Pearson correlation metric acting on a pair of vectors that are ‘close’ in this notion of distance (may not be close in Euclidean sense), and right (ii), a set of points being evaluated using a k -NN (nearest neighbour) lens set with $k=3$.

Then given this apparatus, each group undergoes hierarchical clustering (single-linkage) using the ascribed metric (**Figure 2-15 B iii**), and the resulting clusters form the 0-simplexes, or points of the complex. Essentially this compresses point clouds based on closeness / similarity with respect to the chosen metric. Clusters of points close in space are replaced by single ‘points’, reducing complexity. 1-simplexes are formed with respect to the second function (or the set of functions). A 1-simplex belongs to the complex if it’s 0-simplexes (of which there are two) contain common points, i.e. if those point clouds are close with respect to the defined set of functions that split up the original space. See **Figure 2-15 B iv**.

This provides a complex, specifically a set of simplexes, summarizing the geometry perceived. To then visualize this ‘shape’ is another problem, as there are infinitely many ways to ‘see’ this object. A common method is to use some pseudo-physical simulation, treating the points as atoms, and the 1-simplexes as springs, and allowing forces to resolve a suitable initialization of the points in space into an ideal visualization (see **Figure 2-15 B v**).

To exemplify the generality of this method, **Figure 2-15 C** illustrates a case where the metric is ‘norm correlation’, and the functions for 1-simplex formation are x and y coordinates for an embedded k-NN graph.

Given two points in Euclidean space, the immediate and natural interpretation of ‘distance’ from early education is with the Euclidean metric:

$$d(v_1, v_2) = \sqrt{\sum (v_{1,i} - v_{2,i})^2}$$

It may be advantageous however to consider other notions of closeness, here this is presented as positive correlation strength. Two points (vectors) may be distant with respect to the Euclidean metric, but close under this correlation metric, if the vectors are strongly correlated **Figure 2-15 C i**. The ‘norm’ refers to normalization using z-score.

K nearest neighbours (k-NN) is an algorithm that, for a given point, identifies the k points closest to it with respect to a given metric, and uses this to classify the given point. Mapper used k-NN, with k=20 (arbitrary choice by Ayasdi). To begin with, all

points in the data are 0-simplexes, and each point shares a 1-simplex with its 20 nearest neighbours. Then a physical simulation produces a 2D embedding, which may then be partitioned, i.e. using x and y coordinate functions to split the space into overlapping polygons. The 0-simplexes of the resulting complex are obtained through clustering as previously described, and the 1-simplexes are formed from common membership occurring at the overlapping sites of the polygonal mesh. The process is roughly illustrated in **Figure 2-15 C ii**, with $k=3$.

Ayasdi has implemented the mapper algorithm in a software platform that is accessible via internet browser and/or through a Python package. The function used to encode similarity is referred to simply as ‘metric’. The function(s) for mapping points to a space in which overlapping occurs is referred to as ‘lens’. The number of intervals used in overlap is termed ‘resolution’, and the extent to which subsets are overlapped is called ‘gain’, defined relative to percentage overlap k as: $g = \frac{1}{1-k}$. Bin generation is performed based on the range of the lens function with uniform bin width, or with ‘equalization’ active, where bin width is varied to ensure each bin contains an equal number of data points, useful for skewed data distributions.

2.6.3.1 Exploratory Phenotyping Workflow

The first application of TDA mapper in this thesis was in exploratory phenotyping of asthmatic and non-asthmatic subjects in data set ($n=52$) with features extracted from CT, clinical / lung function testing and demographical characteristics. Ayasdi’s auto generation facility was used to apply a variety of metrics and lenses to the data, based on a proprietary algorithm selecting functions, resolution and gain; the features used in the column set for this analysis were: S_{acin} , R5-R20, LCI, FEV_1 and FEV_1/FVC , i.e. five dimensional point cloud containing information on VH and lung function. Resulting networks were inspected for insight into mechanisms driving VH.

The phenotyping complex presented in this study was generated using the variance normalised Euclidean metric, with the L-infinity centrality lens computing a measure of how central a point is within a set of points:

$$L_{\infty}(y \in X) = \max_{x \in X}(d(y, X))$$

Parameter selection was as follows: resolution = 30 and gain = 3, i.e. 30 bins with an overlap of $\frac{2}{3}$. Bin equalisation was applied. Persistent homology was observed using multiple network generation facilitated by Python script interacting with the Ayasdi server to vary resolution and gain in networks. The network most strongly discriminated based on S_{acin} and is presented as untrained variables characterising asthma and CT variable $\overline{\Delta HU}^{\text{IS}*}$ were also found to be strongly related to centrality in this point cloud, i.e. presenting a case relating inferior-superior ventilation gradient to VH magnitude as measured by the described markers.

Lung Specific Voxel Based Visualisations

Mapper was applied extensively to the visualisation of lungs using a set of 12 voxel based features (see **Table 3.2-2**) was selected for analysis. Data matrices with n rows, n = number of voxels (order 10^7) in a lung pair set, and 12 columns (selected voxel features), were written to .csv Matlab and uploaded to the Ayasdi server. The entire set was then submitted to TDA using the neighbourhood lens functions with default settings of resolution = 80 and gain = 2.5, acting on a column set of the features X, Y, Z and iHU. This was termed ‘anatomically driven’ TDA due to incorporation of iHU with the spatial coordinates, combining information of tissue density with spatial location.

To increase nodal resolution of resultant networks, under constraint of limited network node count, the data was partitioned into left and right lung sets using the Y coordinate feature. The heat map of Y clearly showed the separate lungs in the combined networks, and selection about the peaks of the associated Y histogram provided a natural tool to isolate the left and right lungs. TDA was then applied to these subsets with identical settings (in lens, resolution, gain and column set), to produce lung specific networks.

Figure 3.2-8 in section 3.2 strategically illustrates the above process, leading on to an investigation of iHU across the networks. Lung networks appeared to be stratified by a smooth gradient of iHU with lower iHU nodes coinciding with the majority of voxels (high voxel count), a mid-range (yellow) border and a mass of high iHU low voxel count nodes. The latter group was investigated by lasso tool selection and exporting X, Y and Z of selected nodes using Python script. Then slices, with thickness of 10 voxels,

of these nodes were plotted (plot3.m) for each lung using Matlab, and superimposed onto original source TLC CT images viewed in 3D slicer.

Heat-maps of SO_1, SO_5 and SO_10 presented negligible differences; SO_1 was selected to represent small airway proximity for simplicity. Values around and below the median (node mean) SO_1 were selected using distribution histogram, applying a distinct uniform colour to these nodes (selection colour in program). Images of networks with SO_1 low selected nodes, and heat-maps of features of interest (X, Y, Z, delHU and PRMfsad) were exported to MS PowerPoint environment using screen shots and MS Paint. SO_1 low selected images had transparency applied to the colour of selected nodes, essentially making a 'window' through the nodes deemed proximal to the small airways, and grey-scaled. Images were then combined, with grey-scale in front, and contrast enhanced (+40%) colour images behind, to present a small airway focused visualisation of the feature of interest across the given lung.

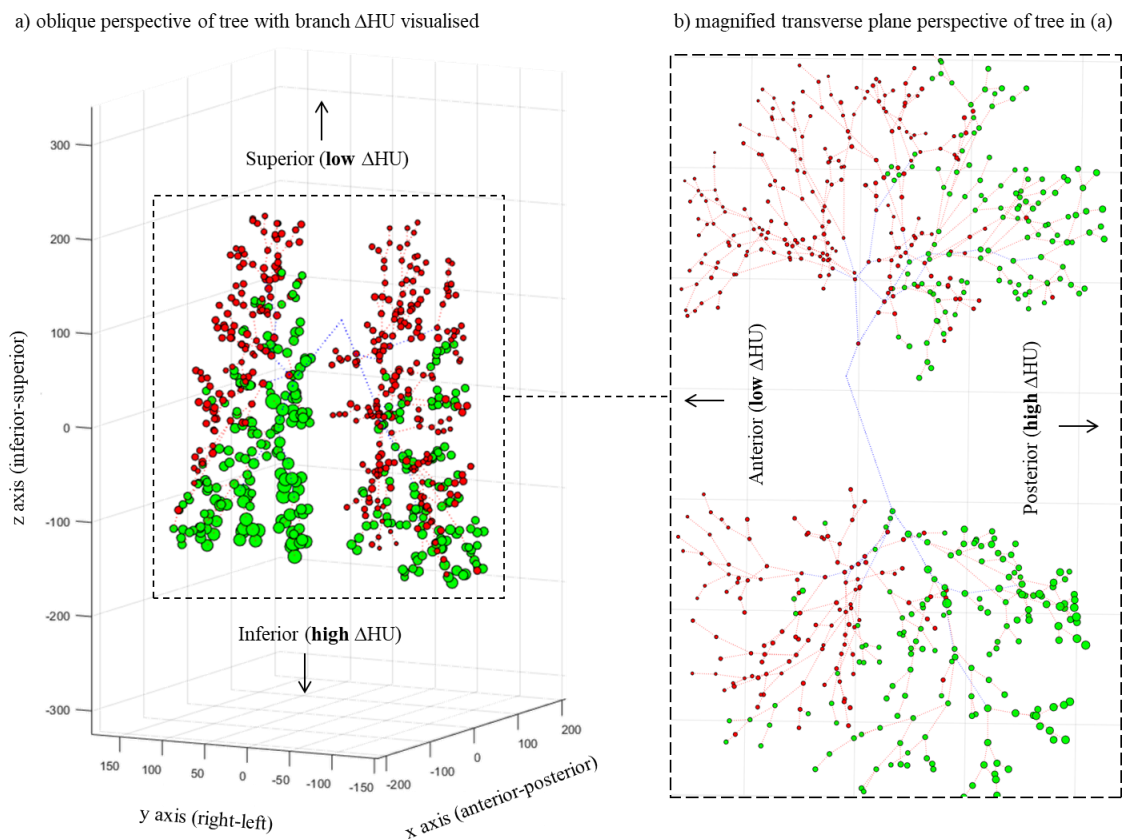
To appreciate the results of this process for 5 features X, Y, Z, delHU and PRM^{fSAD} and present this information concisely for a given patient, in each subject 10 networks (5 for each lung) were selected to be displayed coherently within a summary figure. Spatial coordinates were naturally grouped on the same line for immediate use in understanding distribution of nodes in space, with the other two features on the line below. Feature names were attached with node mean value histograms and range values illustrated in Ayasdi web interface. See, for example, **Figure 3.2-2** in the second study (section 3.2).

2.6.4 Tree Visualisation Algorithm

An algorithm to visualise the Δ HU distribution of branches in a simulated complete airway tree structure, as assigned by the method described in section 2.5.6, for a given range of SO, was designed to provide basic QC/QA (quality control and analysis) of the Δ HU assignment, i.e. judge effectiveness of using spheres of a given radius ($r = 5\sqrt{3}/2 + 0.1$) in capturing ventilation signal from HU changes about terminal branch points from manually aligned point sets (section 2.4.4); additionally visualisations would be useful for providing a compressed representation of Δ HU in a relatively small number of points in space ($10^2 \ll 10^7$), associated to the structure of airways in a subject (exact for the several generations from the trachea, approximate thereafter). Code developed to achieve this goal is provided in the appendices (**Appendix 10**).

An example of the output of this process is provided in **Figure 2-16**. The clear inferior-superior distribution pattern gives basic visual confirmation of adequate success in steps leading to this point, such as CT scanner acquisition at inspiration and expiration, image registration (E2I), Δ HU calculation, set alignment and Δ HU branch assignment based on spheres about the terminal branch points, and the graphical plotting algorithm just described. The method was implemented in the third study of this thesis.

Figure 2-16: SO1 Δ HU assignment visualisation of averages assigned to upper airways.



*Example of output from described Δ HU branch visualisation algorithm in three dimensions at an oblique angle (a) and in the transverse plane (b) of a healthy non-asthmatic subject. The biologically normal elevation of Δ HU in the inferior and posterior regions of the lungs, for a human subject lying supine, is evident, and shows successful conveyance of information from image registered CT voxels captured about the terminal branch points to the upper airways via simple averaging (**Figure 3.3-2**).*

2.6.5 Constriction / Correlation Analysis

To study how ΔHU assignment to branches of simulated tree structures may be used to improve accuracy of resistance estimation in FOT modelling, an algorithm to adjust, and in particular constrict, branch radii was designed and implemented. A function $f_w(\Delta HU(B))$ was defined mapping a branch B with initial radius $r(B)$, and associated ΔHU value $\Delta HU(B)$ to a scale factor for resizing, with a weight w parameterising the magnitude of rescaling. I.e. if r_2 is the radius of branch B after resizing, and $x = \Delta HU(B)$, then the equation determining r_2 is:

$$r_2 = f_w(x) \cdot r(B)$$

To map ΔHU to scale factors for radii adjustment, a linear transformation from the interval $[\min(\Delta HU), \max(\Delta HU)]$ to $[1 - w, 1]$ was sought, such that as w varied from 0 to 1, the most severe constriction would be $1-w$. E.g. if $w=0.6$, the airway with smallest ΔHU would have radii multiplied by 0.4 (reduced to 40% original size). The function effecting this transformation was defined as follows:

$$f_w(x) = 1 + w \frac{x - \max(x)}{\text{range}(x)}$$

Where $\text{range}(x) = \max(x) - \min(x)$ and statistical functions are acting on all branches with SO equal to the branch being resized.

An algorithm was implemented in Matlab to rescale all branches for SO from 1 to a given limit, with rescaling being performed on one SO set of branches at a time. The ΔHU distribution was pre-processed to map all points more than three standard deviations away from the mean to the point $\overline{\Delta HU} + 3 \cdot \text{std}(\Delta HU)$, in order to control for extreme outliers (common in high volume voxel data), and compute scale factors fitting closely to the described $[1 - w, 1]$ target interval. See **Figure 2-17** for histograms that illustrate mapping from a ΔHU distribution to a set of scale factors. Code implementing the described rescaling and pre-processing is illustrated in the appendices (**Appendix 11**).

Figure 2-17: Histograms of mapping raw ΔHU values to airway adjustment factors.

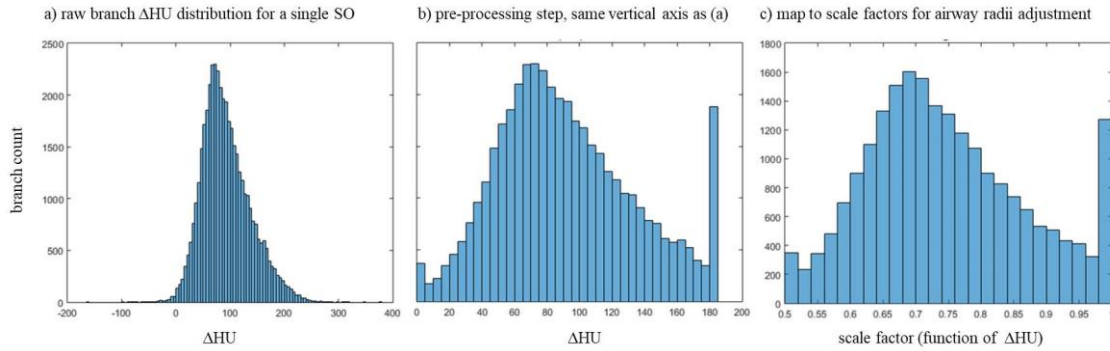


Illustration of mapping branch ΔHU values to scale factors for altering airway radii in direct proportion to ΔHU , in one subject. A, the raw ΔHU distribution, relatively normal in this instance, note the long tails. B, pre-processing step, points outside three standard deviations from the mean are mapped to the mean plus three standard deviations. C, range is rescaled to $[1-w, 1]$ where $w=0.5$ in this case.

To study the effect of ΔHU based radii constriction on FOT modelling of resistance, particularly R5-R20, an algorithm was designed and implemented to check a range of weightings $w = 0, 0.2, 0.225, \dots, 0.575, 0.6$ and SO ranges 1-to- n , where $n = 1, 2, 4, 6$ and 8, i.e. covering varying extents of the airway tree from the most distal airways upwards. FOT simulation code was then applied to a set of trees pre and post radii constriction, and both Pearson correlation and linear regression modelling was applied to study the outcome for every set of parameters. Sample code of the described parameter space search is provided in the appendices (**Appendix 12**).

Results of experiments were observed across the specified SO and weight ranges using a separate analysis script. Built-in function `imagesc.m` was used for heat-mapping matrices of correlation coefficients, from which trends were easy to observe. In general R5-R20 correlation was observed to increase with increasing radii constriction and extent of branch coverage from the terminal branches. R5 and R20 individually appeared to lose correlational strength with increasing constriction and branch coverage. This information is presented alongside R5-R20 correlation increase in **Figure 3.3-6** of section 3.3.

3 Studies

3.1 Functional CT imaging for identification of the spatial determinants of small airways disease in adult asthma

Abstract

Background: Asthma is a disease characterised by ventilation heterogeneity (VH). A number of studies have demonstrated that VH markers derived using impulse oscillometry (IOS) or multiple breath washout (MBW) are associated with key asthma patient related outcome measures and airways hyper responsiveness. However the topographical mechanisms of VH in the lung remain poorly understood.

Objectives: We hypothesised that specific regionalisation of topographical small airway disease would best account for IOS and MBW measured indices in patients.

Methods: We evaluated paired expiratory/inspiratory computed tomography in a cohort of asthmatic (n=41) and healthy volunteers (n=11) to understand the determinants of clinical VH indices commonly reported using IOS and MBW. Parametric response mapping (PRM) was utilised to calculate functional small airways disease marker PRM^{fSAD} and Hounsfield unit (HU) based density change from total lung capacity to functional residual capacity (ΔHU); gradients of ΔHU , in gravitationally perpendicular (parallel), inferior-superior (anterior-posterior) axes, were quantified.

Results: ΔHU gradient in the inferior-superior axis provided the highest level of discrimination of both S_{acin} and R5-20. Patients with a high inferior-superior ΔHU gradient demonstrated evidence of reduced specific ventilation in the lower lobes of the lungs and high levels of PRM^{fSAD} . A computational small airway tree model confirmed that constriction of gravitationally dependant lower zone small airway branches would promote the largest increases in R5-R20. Ventilation gradients correlated with asthma control and quality of life but not with exacerbation frequency.

Conclusions: Lower lobe predominant small airways disease is a major driver of clinically measured VH in adult asthma.

3.1.1 Introduction

Asthma is characterised by spatial heterogeneity in disease and consequent heterogeneity in airways function and lung ventilation (16,275). Ventilation heterogeneity (VH) may be captured using imaging approaches that can quantify and regionalise lung ventilation, such as hyper polarised 3-Helium/129-Xenon magnetic resonance imaging (MRI), Oxygen enhanced MR and single photon emission computed tomography (SPECT-CT) (27,33,276–278). Additionally VH can be measured clinically in patients using physiological tidal breathing techniques that measure heterogeneities in lung ventilation (captured using multiple breath washout (MBW) (279,280)) and mechanical behaviour (captured using impulse oscillometry (IOS) (23)). International guidelines for quality control and assurance of tidal breathing markers of VH derived from IOS and MBW have been proposed (20,133), supporting their potential role as tools to study early airways disease.

We have previously identified that two specific markers of VH, R5-R20 and S_{acin} , derived from IOS and MBW respectively, are associated with impaired asthma control, quality of life and exacerbations (135,280). These observations have been replicated by other groups in parallel studies of adult asthma (281,282). Additionally we have previously demonstrated, using computational small airway models and diffusion MRI, that IOS derived R5-R20 and MBW derived S_{acin} values, are anatomically grounded measures of small and acinar airway anatomical disease respectively, in adult asthmatics (280).

Heterogeneity of ventilation within the lungs is likely to be influenced by both gravitational effects and airway branching, as well as other factors that affect regional lung compliance (reviewed in (283)). However little is known about the spatial lung determinants of clinical measurements of VH derived using MBW and IOS. This is important as imaging tools are costly, difficult to implement in clinical trials and standardise across centres; physiological tools, if appropriately validated, could serve as simple surrogates of disease heterogeneity captured by sensitive imaging techniques.

Computed tomography (CT) of the lungs has been exploited widely to study lung structure and function relationships in asthma (284,285). More recently image registration applied to inspiratory and expiratory CT imaging has been utilised to derive indices of functional small airways disease (32,190,274,286). One specific and widely

deployed approach is parametric response mapping (PRM) (32,190,286). The PRM approach offers the potential to characterise spatial deformation of a voxel between different acquired CT lung volumes, e.g. functional residual capacity (FRC) and total lung capacity (TLC), over the entire lung, and hence the potential to identify spatial mechanisms of commonly measured MBW and IOS VH markers.

The purpose of this study was to use a range of global and regional airway density change (from functional residual capacity to total lung capacity) imaging biomarkers to understand how spatial variations in VH may contribute to widely reported clinical measurements of VH and small airways disease, captured by IOS and MBW in adult asthma.

Specifically we hypothesised that abnormal regional variations in ΔHU would be a major contributor to abnormal IOS and MBW physiological indices of VH in the small airways and sought to test this hypothesis using a functional CT imaging and computational simulation study.

3.1.2 Methods

3.1.2.1 Subjects

The total population for this study consisted of 52 subjects, 41 adult asthmatic and 11 healthy controls. Asthmatic subjects were recruited from Glenfield Hospital in Leicester, UK.

Asthma was defined by a clinician diagnosis with one or more of the following objective criterion (i) bronchodilator reversibility of FEV_1 to 400 mcg inhaled salbutamol of $\geq 12\%$ and 200 mls (17 of 41 asthmatics), (ii) Methacholine $PC_{20} \leq 16$ mg/ml (11 of 41 asthmatics) or (iii) peak flow variation of $\geq 20\%$ over a 2 week period (13/41).

Asthma severity was classified according to the current Global Initiative for Asthma (GINA) treatment steps (143). Severe asthmatics within the cohort had similar lung function (post bronchodilator FEV_1/FVC) to previously reported severe asthma cohorts in Leicester, UK (265), but higher average post bronchodilator $FEV_1\%$ predicted values.

Aged matched healthy volunteers were recruited via local advertising and staff with normal airway physiology and no features of respiratory disease. All subjects with asthma had been free from exacerbations for at least 6 weeks prior to study entry.

All subjects (asthmatic and healthy volunteers) were non-current smokers, however due to the known association of smoking and small airways disease, pack year smoking exposure was not an exclusion criterion. Only 3/41 patients with asthma had a smoking history of more than 15 pack years.

3.1.2.2 Ethical approval

The study protocol was approved by the National Research Ethics Committee – East Midlands Leicester (approval number 08/H0406/189), and all subjects gave their written informed consent.

3.1.2.3 Visits

Clinical and physiological assessment was performed in the following sequence and over 1-2 study visits, no more than 1 week apart. Asthma control was characterised using the modified 6 item Juniper Asthma Control Questionnaire (ACQ-6) (75) and Asthma quality of life using the standard 32 question Juniper Asthma Quality of Life Questionnaire (AQLQ) (74).

Exacerbations were defined according to ATS/ERS consensus criteria (287); a moderate-severe exacerbation is defined as one or more of the following: (i) worsening of asthma that requires use of systemic steroids or an increase in systemic steroids (for patients already receiving maintenance oral steroids) for 3 or more days, or (ii) an admission to hospital or an emergency department requiring systemic steroids.

3.1.2.4 Lung function measurements

All physiological tests were performed in the seated position by individuals with appropriate training and accreditation. Physiological tests were performed 15 mins after administration of a short-acting bronchodilator (salbutamol 400 µg). This was administered via a metered dose inhaler and spacer, with each 100 microgram actuation being inhaled in a separate inhalation to TLC, followed by a 5- to 10-s breath-hold. Spirometry was performed according to ATS/ERS standards (109).

Impulse oscillometry (IOS) was performed in triplicate as previously reported and in accordance with international guidelines (20,133). A volume calibration was performed

daily using a 3-L syringe, and the accuracy of resistance measurements was confirmed daily using a standard $0.2 \text{ kPa}\cdot\text{s}\cdot\text{L}^{-1}$ resistance mesh. Participants wore a nose clip and supported their cheeks, while an impulse waveform was delivered to their respiratory system via a loudspeaker connected to a mouthpiece, during 60 seconds of tidal breathing. Resistance at 5 Hz (R5), resistance at 20 Hz (R20), R5-R20, reactance at 5 Hz (X5) and AX were derived from pressure and flow measurements recorded throughout the 60-second period.

Multiple breath washout (MBW) was performed according to current guidelines (20) by using the sulphur hexafluoride (SF_6) wash-in method as previously described (135). Volume calibration of the pneumotachograph was performed daily using a 1-L syringe. Participants wore a nose clip and breathed an air mixture containing 0.2% SF_6 , while respiratory flows and exhaled breath SF_6 concentrations were monitored by an Innocor photoacoustic gas analyser (Innovision A/S, Odense, Denmark).

Participants maintained a steady respiratory rate of approximately 12 breaths per minute, and a constant tidal volume of 1L, using a real-time visual display of inspired volume as a guide. Once inhaled and exhaled SF_6 concentrations had equalized, participants were switched to breathing room air during expiration and asked to continue breathing at the same respiratory rate and tidal volume. The test was terminated once the end-tidal concentration of SF_6 in exhaled breath reached less than 1/40th of the original concentration (0.005%) for three consecutive breaths.

SF_6 was chosen as the inert tracer gas because of its heavy molar mass and based on previous simulation data from Dutrieue et al (288) suggesting that phase III slope sensitivity to SF_6 is maximal at the level of the alveolar duct. Lung clearance indices, S_{cond} , and S_{acin} , were calculated by using custom software written with TestPoint (Measurement Computing Corp, Norton, Mass) as previously described (135,280).

Body Plethysmography was performed with a constant volume plethysmograph, according to ATS/ERS recommendation (269). A minimum of three acceptable tests were performed and testing ended when repeatability criteria was achieved (FRC within 10% between highest and lowest value). Carbon monoxide uptake was determined using the single-breath method, according to standard guidelines (266). Alveolar volume (V_A) and the carbon monoxide transfer coefficient (K_{CO}) were calculated.

3.1.2.5 CT imaging and image analysis

Volumetric whole-lung scans were obtained with a Siemens Sensation 16 scanner with the following low-dose protocol: $16 \times 0.75\text{mm}$ collimation, 1.5-mm pitch, 120 kVp, 40 mA, 0.5-second rotation time, and scanning field of view of 500 mm, with dose modulation off. Scans were obtained following the administration of 400 mcg of inhaled Salbutamol at FRC and TLC in patients lying supine. Images were reconstructed with a slice thickness of 0.75mm at a 0.5mm interval by using B35f kernel. CT images were quantified using a panel imaging biomarkers (**Table 3.1-3**).

PRM was performed automatically using Imbio's Lung Density Analysis (LDA™) software application (Imbio, LLC, Minneapolis, MN) for all CT data, with registrations performed from TLC to FRC, on segmented voxel sets excluding the major airways (up to 3-4 generations from the trachea). Details on the PRM analysis have been previously reported (32,190,274,286). Relative lung volumes of normal parenchyma (PRM^{Norm}), fSAD (PRM^{fSAD}), emphysema (PRM^{Emph}) and unclassified PRM^{Uncl} were calculated by normalising the sum of all like-classed voxels by the total lung volume. Additionally features of the PRM joint density histogram (ellipse area, minor axis, major axis and angle to horizontal) were derived in MATLAB 2015a (MATLAB Release 2015a, The MathWorks, Inc., Natick, Massachusetts, United States), see **Figure 3.1-1 A**.

A novel algorithm for evaluating regional density change gradients in a given direction, termed stratified axial analysis (SAA), was developed from per voxel TLC to FRC density change (ΔHU , see **Figures 3.1-7** and **3.1-8 A**). This allowed us to investigate how ventilation, approximated by ΔHU , varied with respect to axes of interest; particularly the anterior-posterior (approximately parallel to gravity) and inferior-superior (approximately perpendicular to gravity). See **Figure 3.1-2 C D**. Straight line fitting by ordinary least squares criterion was applied to produce $\text{std}(\Delta\text{HU})^{\text{AP}}$, $\overline{\Delta\text{HU}}^{\text{AP}}$, $\text{std}(\Delta\text{HU})^{\text{IS}}$ and $\overline{\Delta\text{HU}}^{\text{IS}}$ as the gradients of fitted lines to SAA derived intervals (**Figure 3.1-7**), where superscript AP (IS) refers to axis used, anterior-to-posterior (inferior-to-superior); std refers to standard deviation, and \bar{x} refers to **arithmetic mean** of x . $\overline{\Delta\text{HU}}^{\text{IS}*}$ was calculated as the mean of $\overline{\Delta\text{HU}}$ values across every decile, equivalent to scaled (1/9) difference of extreme interval averages, in the inferior-superior direction. N.B. Additional markers classifying lung size asymmetry were also derived using custom scripts in MATLAB (**Figure 3.1-8 B**).

3.1.2.6 Computational simulations of regionalised bronchoconstriction

A computational model of airway impedance was designed, based on previous models in the literature to provide simulations of IOS derived R5-R20. In short, a 1D wave equation was used to estimate the impedance of each branch (203), with total impedance being calculated through summation of parallel and series contributions (205). Each terminal bronchiole was subtended by a constant-phase viscoelastic model parameterised using data from the literature (206). The models were adjusted for potential confounding upper airway shunting (272,289).

Simulations of total lung resistance over the frequency range (1-25Hz) were performed on the healthy conducting airway tree created through a combination of CT segmentation (to an average generation of 6), and algorithmic generation (to an average generation of 16) as previously reported (37). For each simulation, constrictions were applied to either the lowest or highest 25% of small airways ($\leq 2\text{mm}$ in diameter), relative to the supine or orthostatic position, to simulate the effects of gravitationally dependant airways. The constriction rates (the percentage an airway radius was reduced by, denoted c) were drawn uniformly from the range (0-70%), and applied homogeneously, using the same c for all airways, or heterogeneously, drawing each constriction from the normal distribution with mean c , and standard deviation $0.2c$. For each simulation the output R5-R20 was calculated.

3.1.2.7 Statistical analysis

Statistical analyses were performed in MATLAB 2015a. Kolmogorov-Smirnov tests were applied to check likelihood of a normal distribution. Binary group comparisons were performed using two-sample t-test (parametric data) and Mann-Whitney U-test (non-parametric data); for multiple group comparisons one-way ANOVA test was utilised (parametric data) and Kruskal-Wallis test (non-parametric data). Multiple-comparison procedures were performed with Tukey's honest significant difference criterion. Subgroups were determined by GINA treatment steps, and according to mean S_{acin} and R5-R20. 16 subjects (roughly one third of total population) at each end of R5-R20 and S_{acin} distributions were utilised in tertile polar analysis, at SAA inferior-superior deciles, where statistical significance was determined using two-sided Wilcoxon Rank-Sum test.

Average Pearson's correlation co-efficient (\bar{r}) is reported for correlational strength between CT biomarkers and selected non-CT features, in addition linear regression performed on raw CT data, selected CT features with minimal collinearity (see **Figure 3.2-12**), and principal components using Kaiser-rule (above mean variance explained). Linear discriminant analysis (LDA) was applied to determine class separation of VH markers S_{acin} and R5-R20 using combinations of CT imaging features, clinical features and spirometry. Negative binomial regression was used to evaluate the relationship between exacerbations and imaging biomarkers and Pearson correlations for association between asthma control/quality of life and imaging ventilation gradient biomarkers.

A p-value of $p < 0.05$ was utilised to define statistically significant results in all tests.

3.1.2.8 Transcriptomics Analysis

Nasal epithelial cell samples were obtained using a nasal scraping. Subjects underwent nasal scraping to collect airway epithelial cells. Scrapings were performed behind the inferior turbinate using a nasal cytology curette to obtain cellular sample, which was then stored in 500 μ l of RNA protect cell reagent, in an RNase/DNase-free cryovial, and kept at -80°C until shipment to central laboratory. RNA was extracted and processed using a Human Genome U133 Plus 2.0 Array from Affimetrix. Gene expression was derived and normalised.

Hierarchical single-linkage clustering was applied to the gene expression data for all subjects with successful sample acquisition / nasal scraping participation (45/52). A total of 13,400 sites were used to characterise genetic profiles for subjects. Clustering metric was Pearson correlation coefficient. Groups identified in clustering were evaluated for mean $\overline{\Delta HU}^{IS*}$; groups with highest and lowest means were further analysed with Wilcoxon ranksum test to isolate genes with lowest p-value in binary group discrimination, followed by selection of genes with highest absolute mean difference of gene expression between the two groups.

3.1.2.9 Topological Data Analysis

An integrated data set of CT with respiratory clinical/physiological features were phenotyped using TDA software Ayasdi, with the variance normalised Euclidean metric and L-infinity centrality lens computing a measure of how central points were within the data set:

$$L_{\infty}(y \in X) = \max_{x \in X}(d(y, X))$$

Where y is the point under evaluation, X is the total data set and d the selected metric.

Parameters were selected as follows: resolution = 30 and gain = 3. Bin equalisation was active. The column set for this analysis was: S_{acin}, R5-R20, LCI, FEV₁ and FEV₁/FVC, i.e. a five dimensional point cloud containing information on VH and lung function.

3.1.3 Results

3.1.3.1 Clinical characteristics

Clinical characteristics of the population are outlined in **Table 3.1-1**. Asthmatic patients were matched for age and sex to healthy volunteers. The asthmatics population had significantly greater eosinophilic airway inflammation and physiological evidence of airways dysfunction and VH when compared to healthy volunteers. There were no significant differences in VH markers R5-R20 and S_{acin} across GINA treatment intensity groups. Of the 3/41 asthmatic patients with a smoking history of more than 15 pack years, all had a PRM emphysema (PRM^{Emph}) score that was less than the mean + 1.96 SD (5% PRM^{Emph}) in a healthy aged matched population of 98 subjects (290) and preserved K_{CO} % predicted values (**Table 3.1-1**). The three patients all had evidence of asthma objectively (one had 78% FEV₁ reversibility, one had a PC₂₀ methacholine of 2 mg/ml, one had 49% FEV₁ reversibility). Furthermore of these three patients only 2 patients demonstrated a post BD FEV₁/FVC < lower limit of normal (LLN) (63% predicted in both patient respectively with a post BD FEV₁% of 72% and 57% respectively).

3.1.3.2 Imaging biomarkers of global lung VH are not associated with small airway VH markers R5-R20 and S_{acin}.

Table 3.1-3 outlines all of the CT derived imaging biomarkers. **Table 3.1-4** presents comparisons of the global and regional imaging biomarkers comparing asthmatic and healthy cases across the spectrum of GINA treatment intensity. Asthmatic cases demonstrated significantly smaller PRM ellipse major diameters and smaller ellipse

angles and had narrower distributions (standard deviations) of voxel HU change from FRC-TLC ($p < 0.05$) compared to controls – indicative of less overall VH.

Asthmatics did not differ from controls with respect to standard PRM markers (PRM^{Norm} , PRM^{fSAD} and PRM^{Emph}), see **Table 3.1-4**. In contrast patients who demonstrated FEV_1/FVC (%) less than the median population value (primarily asthmatics) had higher levels of PRM^{fSAD} and smaller PRM global ellipse areas (suggestive of less global heterogeneity) when compared to patients with FEV_1/FVC (%) \geq than the median value ($p < 0.05$), see **Figure 3.1-1 B C**. These observations were not replicated with the small airway indices of VH R5-R20 and S_{acin} (**Figure 3.1-9**) and indicate that global PRM indices in the lung track with spirometry defined airflow obstruction in contrast to small airway physiological indices.

3.1.3.3 Imaging biomarkers of regional VH are major determinants of small airway VH markers R5-R20 and S_{acin} .

To evaluate the relationship between regional imaging measures and small airways physiology, the population was split into Low/High sub groups (Low \leq mean, High $>$ mean) according to absolute S_{acin} and R5-R20 values. **Tables 3.1-2** and **3.1-5** (clinical features) and **Table 3.1-6** (imaging markers) summarise clinical and imaging features according to this stratification. Healthy cases predominated in the S_{acin} Low (9/11) and R5-R20 Low (8/11) groups, and asthmatic cases in the high groups.

Regional analysis identified that the gradient markers evaluating inferior-superior axis FRC-TLC deformation ($\overline{\Delta\text{HU}}^{\text{IS}}$ and $\overline{\Delta\text{HU}}^{\text{IS}*}$) slopes were the only markers that differed significantly in patients in the upper tertile of S_{acin} and R5-R20 when compared to the lower tertile ($p < 0.05$), see **Table 3.1-2**. Specifically for both S_{acin} and R5-R20 High cases, the ventilation gradient was reversed in the inferior-superior axis ($\overline{\Delta\text{HU}}^{\text{IS}}$ and $\overline{\Delta\text{HU}}^{\text{IS}*}$), such that ventilation was significantly reduced at the base of the lung. This is further exemplified in **Figure 3.1-3**, which presents ventilation gradient maps from the base to the apex of the lung comparing cases within the upper and lower tertiles of R5-R20 and S_{acin} respectively, and two exemplar subjects with and without ventilation gradient reversal. A similar but markedly less pronounced gradient change could be seen in the posterior regions of the lower lobes (anterior-posterior axis ($\overline{\Delta\text{HU}}^{\text{AP}}$)) when

comparing patients with high and low clinical levels of VH (R5-R20, S_{acin}), demonstrating reduced posterior ventilation in the lower lobes (**Figure 3.1-10**).

Further examination of the distribution of ΔHU and regional PRM^{fSAD} , in cases with a high and low $\overline{\Delta HU}^{IS*}$, identified that patients with high $\overline{\Delta HU}^{IS*}$ (ventilation gradient reversal) appeared to have focused regionalisation of lung disease (particularly *but not exclusively* in the lower lobes), see **Figure 3.1-4**. In contrast patients with a low $\overline{\Delta HU}^{IS*}$ had more homogeneous distributions of both ΔHU and PRM^{fSAD} .

15/16 $\overline{\Delta HU}^{IS*}$ **high** cases had abnormal regional ventilation in contrast to 5/16 $\overline{\Delta HU}^{IS*}$ **low** cases. Regionalisation of disease was in the lower lobes, i.e. generally focussed at the lung bases (see arrows). A chi squared analysis of the proportions of cases with abnormal regionalisation in each group demonstrated a p-value of $p < 0.0001$. ΔHU and PRM classifications correlated imperfectly (**Figure 3.1-4**), however lower ΔHU voxels were consistently associated with PRM^{fSAD} (**Figure 3.1-11**).

3.1.3.4 Linear relational analyses: associations between CT and non-CT features

Table 3.1-7 presents a high level summary of outcomes for linear analyses. ‘Corr.’ column presents average absolute Pearson’s correlation between all CT attributes and listed target attributes (rows). LR (subset) presents R^2 and F-statistic p-values from linear regression, using 14 feature subset selected to eliminate significant collinearity of predictor variables (**Figure 3.1-12 B**). LR (PCA) presents the same with principal components 1 through 6 as predictor variables. PCA-LR Coef. (abs. value) presents the loading scores of the principal components in cases where the F-statistic indicated statistically significant ($p < 0.05$) likelihood of a linear relationship, which should be used with **Table 3.1-9** for interpretation; that is linking principal component loading score magnitude with linear regression co-efficient magnitude to associate CT input features to targeted non-CT features.

Table 3.1-8 presents the coefficients derived in the application of linear regression to the 14 feature subset of CT attributes. Since the coefficient magnitude is dependent on ordering of predictor variables, the values for each input are assessed to determine the largest for a given input variable, and this may then be related to an target variable (row). Predictor variable with ID 9 (column), which is RLsizeRat (see **Table 3.1-9**

subscripts), has .44 as its highest magnitude, assigned to outcome variable R5-R20. In fact the outcome variable AX in this case is associated with magnitude .42, significantly higher than all other magnitudes, suggesting association between RLsizeRat and IOS (over other outcome variables). This result may be linked to the observed significance of feature RLsizeRat in **Table 3.1-6**; that is it appears as one of the CT variables which discriminates R5-R20 extremes.

Table 3.1-9 presents results using the PCA approach used in linear regression. The loading scores of all 22 CT attributes studied (rows) are presented for all 6 principal components used in linear regression. Combining this with the final column of **Table 3.1-7**, it is possible to study relational strength between outcome and predictor variables. R5-R20 has P2 as its highest loading principal component, and observing this column in **Table 3.1-9** it is clear that CT feature $\overline{\Delta HU}^{IS*}$ (practically equivalent to $\overline{\Delta HU}^{IS}$) has the greatest absolute loading value, suggesting association between $\overline{\Delta HU}^{IS*}$ and R5-R20. This association, in addition to vCntX (anterior-posterior segmented lung width) and ellMinL, may also be observed in **Table 3.1-6**.

3.1.3.5 Imaging gradient biomarkers and clinical disease

We examined the relationship between imaging gradient biomarkers and clinical disease expression. We found that the anterior posterior gradient biomarker $\text{std}(\Delta HU)^{AP}$ correlated significantly with both ACQ-6 ($r = 0.33$, $p=0.039$) and AQLQ ($r=-0.34$, $p=0.02$). We also found a significant association for the inferior-superior gradient imaging biomarker ($\overline{\Delta HU}^{IS*}$) and asthma quality of life ($r=-0.39$, $p<0.01$) but not asthma control. None of the gradient biomarkers were associated with exacerbation frequency.

3.1.3.6 Discrimination of S_{acin} and R5-R20 with imaging markers of density change (ΔHU) gradients and lung size asymmetry

Figure 3.1-5 and **Tables 3.1-10** and **3.1-11** present results of linear discriminant analysis (LDA) which sought to identify the relative contribution of spatial CT derived VH biomarkers, potential clinical contributors/confounders and spirometry to physiological VH indices S_{acin} and R5-R20. LDA demonstrated that the CT markers of ΔHU in the inferior-superior and anterior-posterior axes as well as right to left lung size asymmetry provided the greatest overall discriminatory value of small airway

physiological indices, confirming that these metrics contained most of the information content of the clinical small airway physiological indices.

3.1.3.7 Computational modelling validation of CT imaging PRM gradients

Computational modelling of regional bronchoconstriction in small airway patient specific conducting airway models (**Figure 3.1-6**) identified that increasing constriction of the small airways ($\leq 2\text{mm}$ diameter), that would be most influenced by gravity in the supine posture (lower lobe and posterior), promoted profound elevations in R5-R20 that were not seen with orthostatic simulations (i.e. constriction of small airways that would be most influenced by gravity in the orthostatic posture). Furthermore, similar regional constriction in the upper lobes did not promote the same difference on R5-R20 when considering orthostatic and supine postures. The computational models therefore provided further insight into associations between lower lobe regional focus of disease and R5-R20 response seen in the clinical imaging study (**Figures 3.1-3 and 3.1-4**).

3.1.3.8 Transcriptomics analysis: genes discriminating with respect to $\overline{\Delta HU}^{IS*}$

Results of hierarchical clustering and subsequent $\overline{\Delta HU}^{IS*}$ focussed discriminant analysis are presented in **Figure 3.1-13**. Correlation based clustering identified seven groups of subjects based on pattern of gene expression, annotated G1-G7. Columns were permuted by the clustering algorithm (Matlab's 'clustergram') to place similar subjects adjacent. Calculating mean $\overline{\Delta HU}^{IS*}$ for each group revealed that G3 (n=7) had the lowest mean gradient (-2.69) compared to G7 (n=3) which had the highest (2.73). To isolate a shortlist of 30 genes discriminating these groups, Wilcoxon ranksum testing followed by absolute mean difference maximisation was applied. Identified genes are presented in **Figure 3.1-13 C** and detailed in **Table 3.1-12**. Gene CXCR5 was elevated in subjects with abnormal $\overline{\Delta HU}^{IS*}$, and has been linked to inflammation (291).

3.1.3.9 Topological data analysis: phenotyping asthma

The results of TDA using Ayasdi to phenotype a data set including CT, clinical and physiological features, for asthma relevant phenotypes, are provided in **Figure 3.1-14**. TDA identified a set of subjects annotated C1 that appeared to occupy the central region of the data cloud (low L_∞) with higher acinar VH (S_{acin}), GINA score and FEV_1 reversibility due to bronchodilator (BD +/- %) with higher $\overline{\Delta HU}^{IS*}$ (abnormal inferior-superior ventilation profile). In contrast the other major group identified as C2, occupying an outer region of the data cloud (high L_∞) showed lower acinar VH, GINA

score and FEV₁ reversal, with a lower $\overline{\Delta HU}^{IS*}$. Statistical comparisons of these centrality based phenotypes are presented in **Table 3.1-13**.

3.1.4 Discussion

We have performed the first quantitative functional CT imaging study to understand the spatial determinant of small airway VH markers R5-R20 and S_{acin} in adult asthmatics and healthy volunteers. Furthermore we have coupled CT imaging with computational simulation of small airway physiology to understand the impact of disease regional pattern upon abnormal physiological indices of VH.

Using a panel of imaging biomarkers (**Table 3.1-3**), derived from inspiratory and expiratory CT scans, we have identified that gradients in ΔHU from the base to apex of the lung are a key determinant of both physiological measurements. Notably there is a reversal of the normal ventilation gradient in this axis, such that ΔHU is reduced at the base of the lung in patients with asthma and indeed occasionally in healthy volunteers with abnormal S_{acin} and R5-R20 values. In addition we have identified that other mechanisms including anterior-posterior ΔHU gradient decrease and other nonspecific regionalisation of ΔHU may underpin abnormal R5-R20 and S_{acin} indices in adult asthma. We found broadly similar but not identical results with the widely reported markers of small airways disease PRM^{fSAD} . Computational small airway tree models were then used to confirm the impact of gravitationally dependant lower lobe disease regional focus on IOS marker R5-R20, and matched our observation closely.

Previous studies have examined the difference in VH, between asthmatic and healthy subjects, using hyperpolarized 3He MRI (27), and another linked hyperpolarized 3He MRI with computational models to examine airway constriction in asthma (292). We are also aware of one study in bronchiectasis that attempted to correlate global burden of CT disease with physiological indices of VH (293). This study used correlations and regressions to identify associations between MBW lung clearance index LCI (a global marker of VH) and a CT scoring of the extent of bronchiectasis

Our study uses quantitative functional CT derived indices and specifically sheds insight into the topographical origins of abnormal R5-R20 (IOS derived) and S_{acin} (MBW derived) VH markers. Furthermore our observations, coupled with computer

simulations, see **Figures 3.1-3, 3.1-4 and 3.1-6**, suggest that regionalisation rather than global disease burden may be key determinants of S_{acin} and R5-R20 in asthma.

Our results are clinically important for a number of reasons. Reduced basal ventilation in asthma may be associated with reduced effective deposition of inhaled drugs, which may be a factor in poor asthma control reported in patients on ICS/LABA (inhaled corticosteroids / long-acting beta-agonist) combination therapies in European and other populations (294); this hypothesis would require testing with future studies.

Additionally our findings are important as they are the first to use spatial and functional information derived from quantitative PRM based CT imaging to shed insight into empirical lung physiological measurements R5-R20 and S_{acin} that are widely reported as small airway dysfunction detection tools.

Interestingly we found few differences in the PRM whole lung averages for functional small airways disease, emphysema and healthy (normally deforming lung voxels) in patients with and without high levels of clinical VH derived from MBW and IOS. In contrast average whole lung PRM values were associated with airflow obstruction measured using spirometry. The latter observations highlight both the importance of using the full information content of spatial imaging when trying to understand the topographical basis of VH indices, and the fact that expiratory flow limitation in asthma is a maker of total burden of lung unit damage rather than the heterogeneity of damage.

A likely factor of the observed ventilation gradients in the lung is the ‘slinky’ effect, which describes the compression of a slinky coil parallel to the gravitational field under normal gravitational conditions, isogravity and hypergravity (87). As the dependant regions of the lung are compressed by the weight of the lung above them, they have lower end expiratory volume and the surrounding pleural pressure is more positive (in comparison to the apex), consequently a given respiratory effort and change in pleural pressure will lead to a larger increase in volume.

Other factors responsible for ventilation and perfusion gradients are likely to include lung elastic recoil, nonlinear pressure-volume relationships, the influence of large vessels, and airways closure within dependant airways. These effects have been reported in imaging studies using both protocol MRI approaches (91) and more recently a CT imaging lung deformation study in severe asthma (274).

The finding of a reverse ΔHU gradient at the base of the lung in patients with abnormal S_{acin} and R5-R20 values, and asthma, may occur as consequence of a number of factors in asthmatic patients. Specifically basal airways may close at FRC in asthma, particularly when supine, and this may reduce the specific ventilation to the lung base; one study observed results to this effect in airway constriction due to methacholine challenge (24). Additionally the average BMI in our cohort was 30 kg/m², fat distribution in the abdomen and near the base of the lung may alter diaphragmatic and basal airway mechanics and promote airways closure. Additionally it is possible that there is preferential remodelling of the airways in the lung base in asthmatics. However this would need to be confirmed by pathological studies. Similar effects including the impact of gravity may promote the smaller anterior-posterior gradient decrease seen in patients with clinical VH.

TDA has previously been used to phenotype disease in type-2 diabetes (50) and asthma (52), using identified networks to support discovery of new disease subtypes. In this study we have presented evidence supporting the existence and importance of subtypes in the asthmatic population that relate to ventilation distribution across the lung in the inferior-to-superior axis. Additionally we have investigated nasal transcriptomics and presented genetic associations to extremes of ventilation gradient that have been linked to lung disease and asthma, such as CXCR5 (291) and VAV1 (295). These are exploratory findings limited by small group size (e.g. n=3) presented for interest, and should clearly be validated in larger studies for researching the micropathology underpinning abnormal lung ventilation gradients.

The current findings in this report add to our previous observations which have identified that both the degree and heterogeneity of small airway obstruction promote abnormal R5-R20 values (296), and that S_{acin} may be driven by asymmetries in the lung at length scales that equate to the level of the alveolar duct (280). Specifically here we show that ventilation gradients in the lung are a major discriminant factor associated with both abnormal IOS derived R5-R20 and MBW derived S_{acin} values.

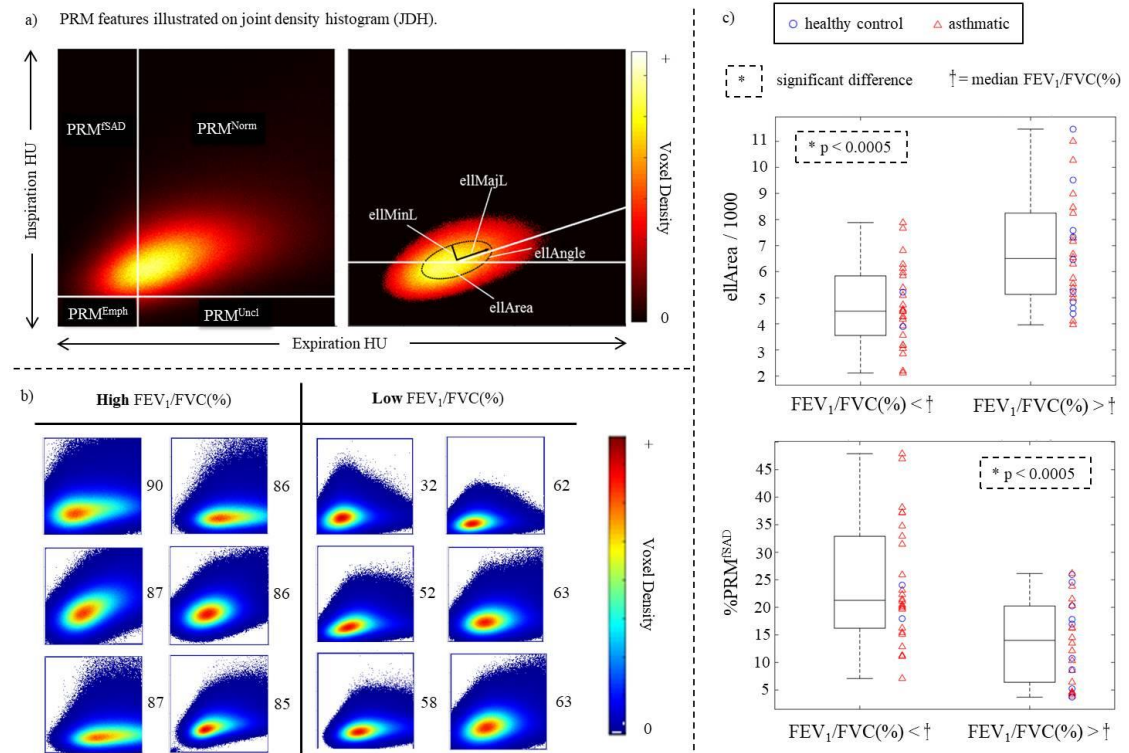
There are a number of limitations to our findings that warrant further evaluation. Firstly our study included asthmatics with a smoking pack year history of more than 15 pack years. Although these patients had no demonstrable imaging or physiological evidence of emphysema it is possible that smoking exposure rather than asthma per se was the

driver of disease gradients in these patients. As a consequence larger studies are required to evaluate the gradient biomarkers reported here, across the spectrum of asthma severity and treatment intensity, and in both smoking and non-smoking asthma populations. The same limitation of sample size warrants further evaluation of the imaging biomarkers in severe asthma populations, and considering the association of the biomarkers with patient related outcome measures in asthma. Such studies are underway and will report in due course (297). Our imaging gradient biomarkers (derived via image registration of inspiratory and expiratory CT scans) are likely to be sensitive to both reconstruction kernel and lung volumes as reported previously (191). However all of our CT scans were acquired at a single centre with the same reconstruction kernel and all patients were coached to expire to FRC for expiratory CT imaging prior to scanning. Nonetheless it is possible that expiratory imaging near residual volume would accentuate the imaging findings observed here and future studies are required to assess the impact of expiratory volume upon the imaging biomarkers reported here.

In conclusion, we have shown for the first time, using functional and computational approaches derived from CT imaging, that small airway VH, captured by IOS R5-R20 and MBW S_{acin} , is associated with CT density gradient reversal at the lung base, which is likely to be a direct consequence of reduced specific ventilation and small airways disease. The implications of these findings upon clinical disease expression, inhaled drug deposition and potential use in targeted inhaled drug delivery systems should now be considered in larger imaging cohorts and interventional studies.

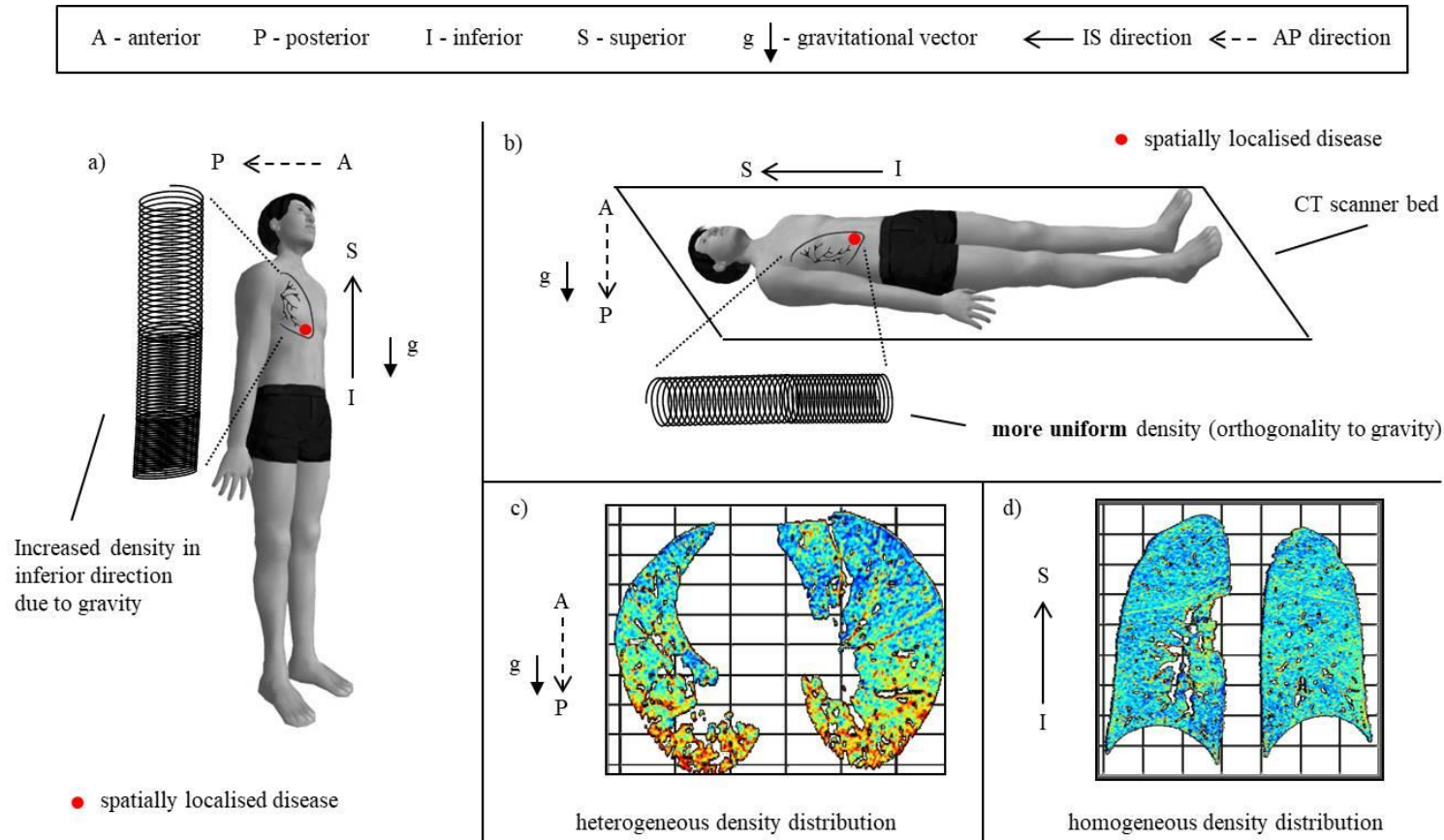
3.1.5 Figures and tables

Figure 3.1-1: JDH and association with obstruction (spirometry).



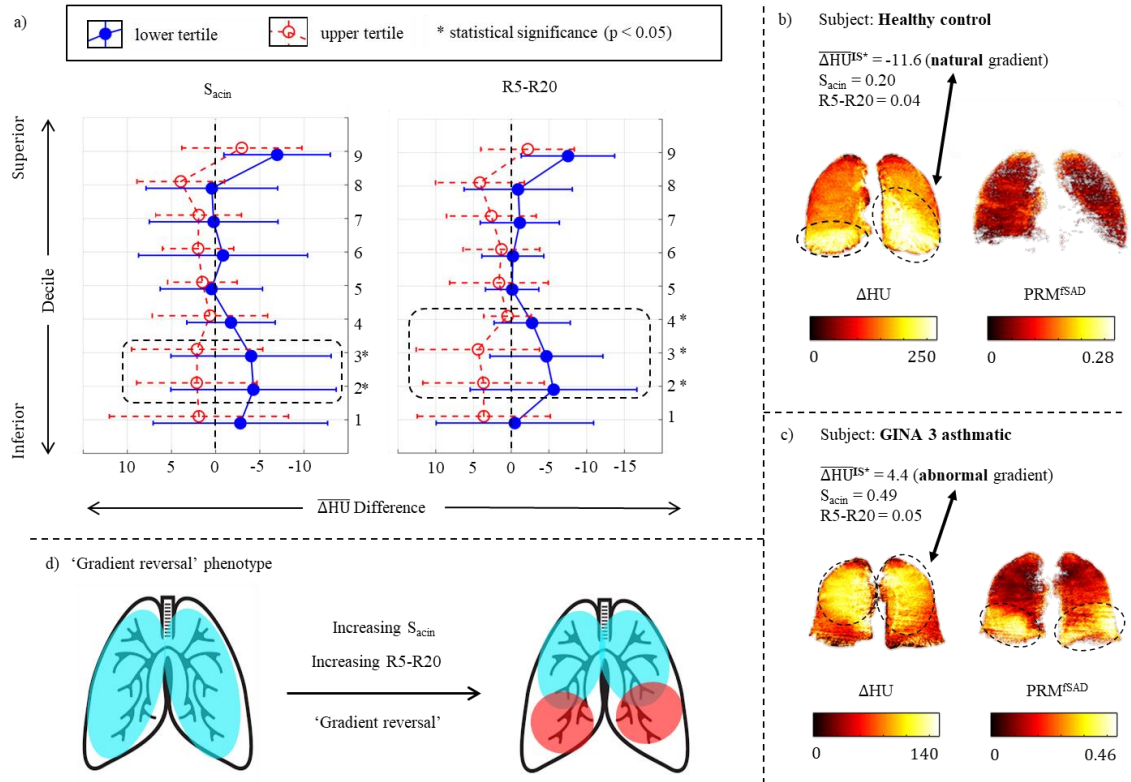
Global parametric response (PRM) mapping and spirometry. A: PRM features based on TLC and FRC HU joint density histogram (JDH). PRM voxel classification (left) defined by lines of expiration HU = -856 and inspiration HU = -950 utilised for defining PRM^{fSAD} and PRM ellipse geometry. B, JDH visualisation of FEV_1/FVC (%) extreme cases, demonstrating compact and left shifted ellipses in patients with airflow obstruction. C, box plot illustrating that patients with spirometric airflow obstruction have smaller PRM ellipse areas and significantly more functional small airways disease on CT imaging $\%PRM^{fSAD}$; groups formed about median FEV_1/FVC (%).

Figure 3.1-2: Subject postures with respect to gravity and lung density distribution.



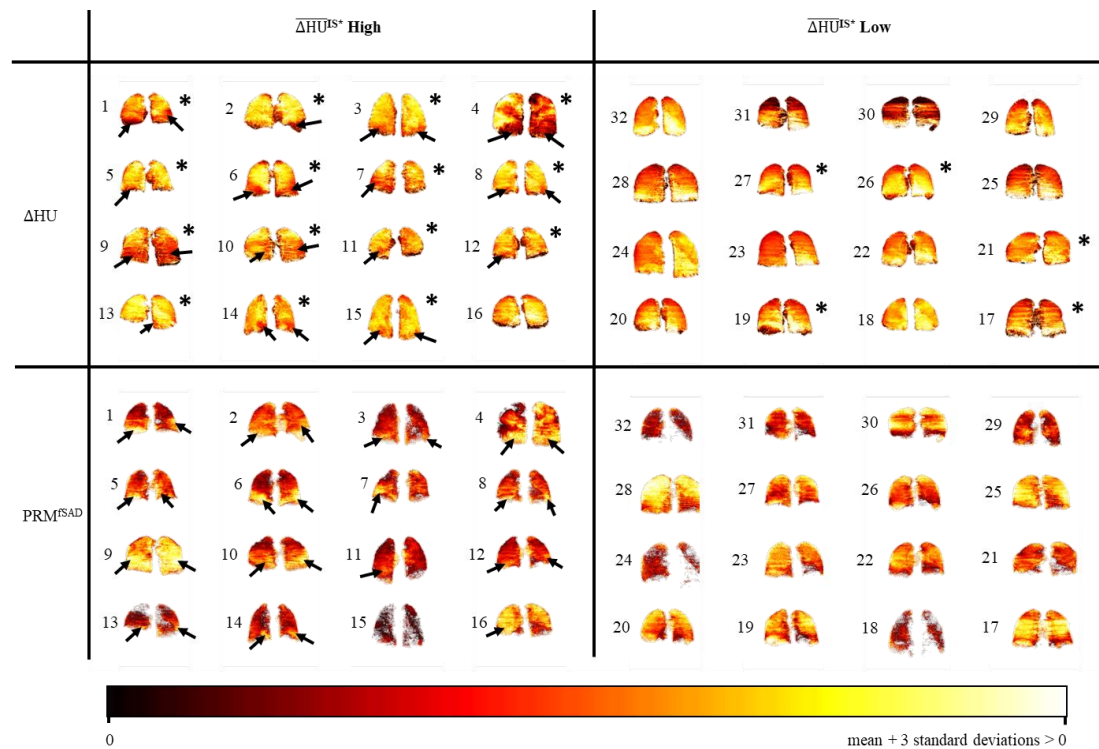
Overview of the slinky effect in the lungs in the standing (A) and supine (B) postures demonstrating the lung density distribution as a consequence of gravity. C, transverse cross section of expiratory HU voxels; the anterior-posterior lung density distribution is strongly influenced by gravity. D, coronal cross section of expiratory HU voxels; the inferior-superior lung density profile is relatively independent of gravity. The gradients of ΔHU in these axes were used to understand determinants of clinical VH measurements derived from IOS and MBW.

Figure 3.1-3: Decile-wise tertile analysis, min-max projections and phenotype observed.



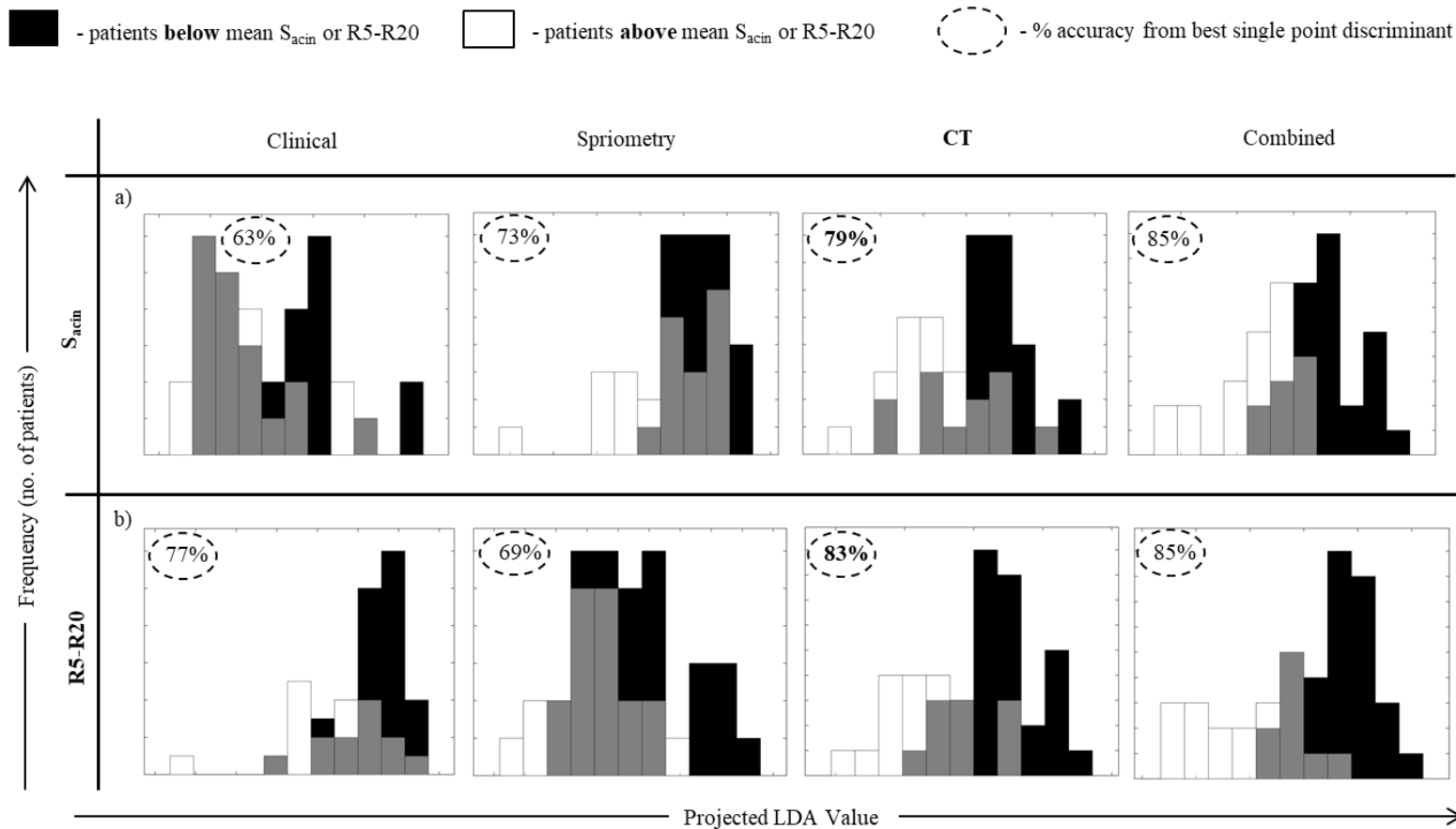
Inferior-superior ΔHU gradient analysis in patients with a high/low S_{acin} and $R5-R20$. A, decile-wise comparison of ΔHU mean differences, in the inferior-to-superior direction, of groups formed from the lower and upper tertiles of S_{acin} and $R5-R20$ distributions, specifically the lowest and highest 16 subjects with respect to these two markers. The inferior regions show significant differences when comparing lower and upper tertiles for both S_{acin} and $R5-R20$. B, joint density histogram of voxel mean ΔHU and PRM^{fSAD} percentage when projected onto coronal plane in subject showing typical (healthy) ventilation (surrogated by ΔHU) pattern and homogenous PRM^{fSAD} . C, as in B in a subject with abnormal ventilation pattern and basally focused PRM^{fSAD} . Colour bars labelled with min and max of occurring mean values. D, the concept of the inferior-superior gradient reversal phenotype is summarised in a simple visual schematic.

Figure 3.1-4: Coronal heat maps of ΔHU and PRM^{fSAD} in SAAz low and high tertiles.



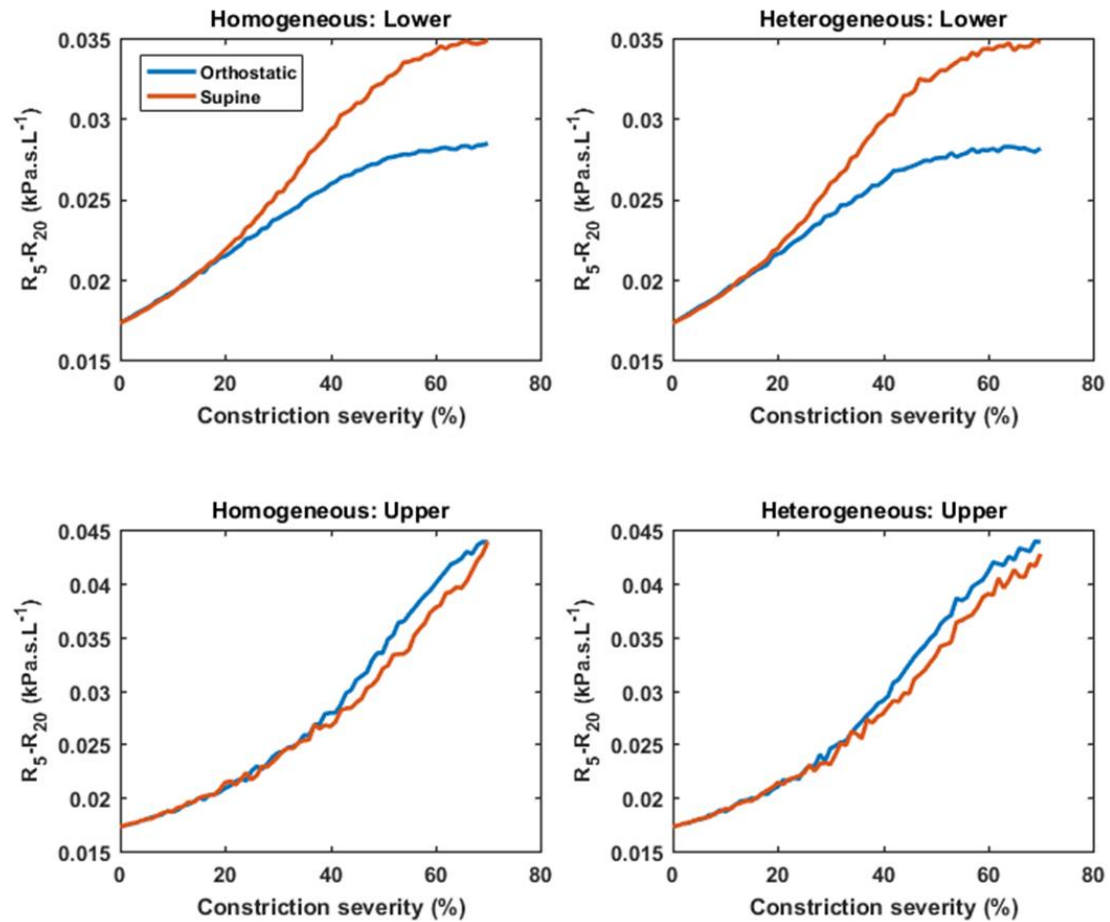
Coronal section heat maps of ΔHU and PRM^{fSAD} in $\overline{\Delta HU}^{IS*}$ low (no gradient reversal) and high (basal gradient reversal) tertiles of total population ($n=52$). The images labelled with an ID number assigned with respect to decreasing $\overline{\Delta HU}^{IS*}$ values. e.g. 1 = greatest $\overline{\Delta HU}^{IS*}$ (highest level of inferior/lower zone gradient reversal), 32 = smallest $\overline{\Delta HU}^{IS*}$ (lowest level of basal gradient reversal). H indicates non-asthmatic, G indicates asthmatic, with GINA level. It can be seen that patients with high $\overline{\Delta HU}^{IS*}$ values more often than not have inferior gradient reversal but also exhibit ΔHU and PRM^{fSAD} heterogeneity. In contrast patients with a low $\overline{\Delta HU}^{IS*}$ appear to have more homogeneous distributions of ΔHU and PRM^{fSAD} or upper lobe regionalisation of low ΔHU , as would be expected in the supine posture. Colour bar ranges determined per subject based on feature (ΔHU or PRM^{fSAD}) mean and variance as indicated. Arrows highlight specific disease regionalisation in $\overline{\Delta HU}^{IS*}$ abnormal subjects. Asterisks indicate subjects selected for chi squared test of proportions, having abnormal regionalisation of ventilation.

Figure 3.1-5: Visual overview of linear discriminant analysis on disparate data domains.



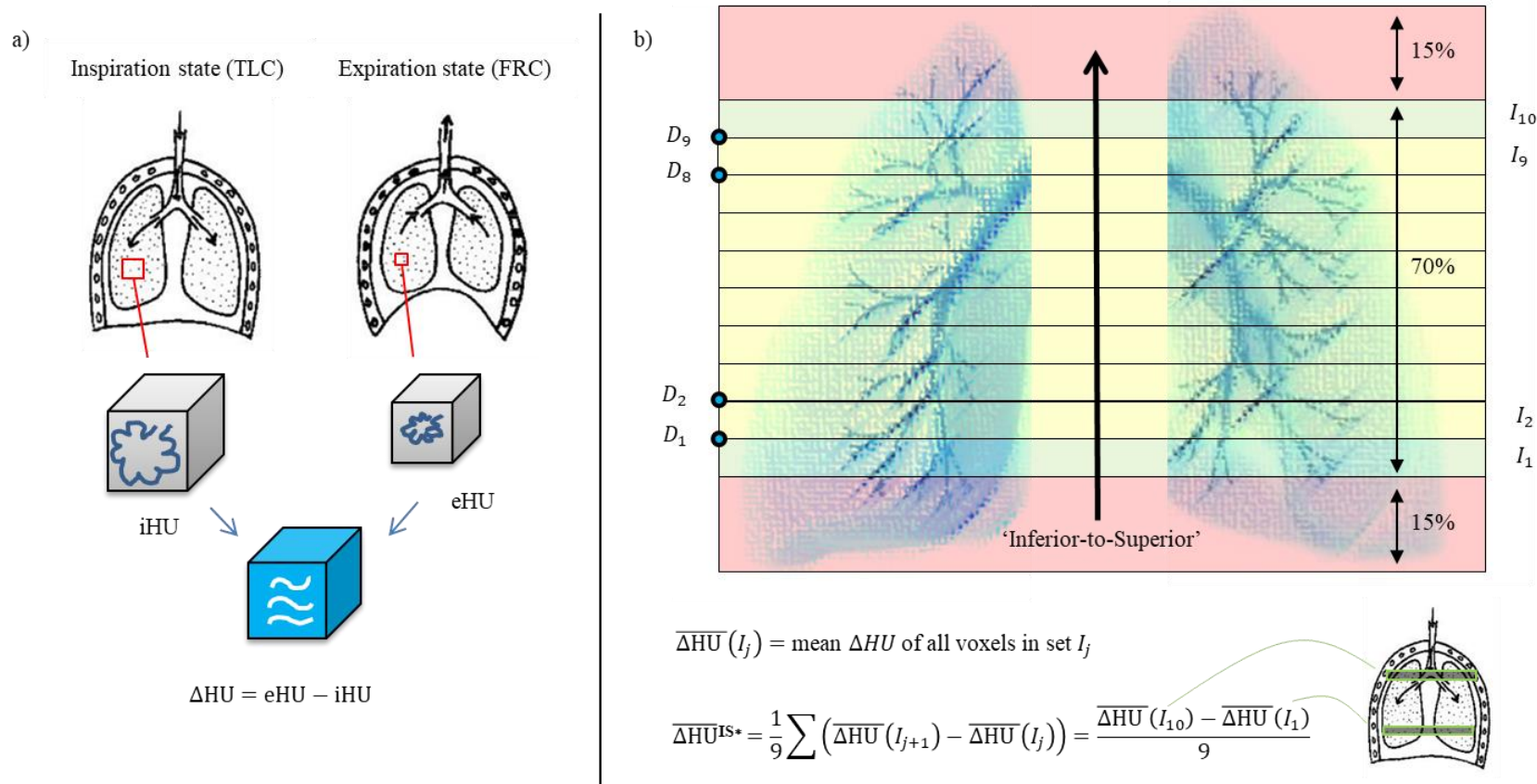
Histograms of linear discriminant analysis (LDA) applied to the total population ($n=52$), illustrating best linear separation of clinical ventilation heterogeneity (VH), R5-R20 and S_{acin} . Limited additional discrimination is added when considering potential clinical confounders of VH (e.g. age, height and weight) and spirometry appears to be less sensitive at discriminating patients with normal and abnormal clinical VH than CT imaging.

Figure 3.1-6: Effect of computation model based regionalisation on small airway physiology.



Comparison of R_5-R_{20} under varying regional small airway constrictions applied to healthy lung structure. The response of R_5-R_{20} can be seen for homogeneous (A, C) and heterogeneous (B, D) constriction of the small airways. In each case constrictions were applied to the lowest or highest 25% of airways, relative to the orthostatic or supine position. It can be seen that lower zone constriction and regionalisation produces far greater elevations in R_5-R_{20} than upper lobe constriction and regionalisation, in keeping with the observations in **Figures 3.1-3** and **3.1-4**.

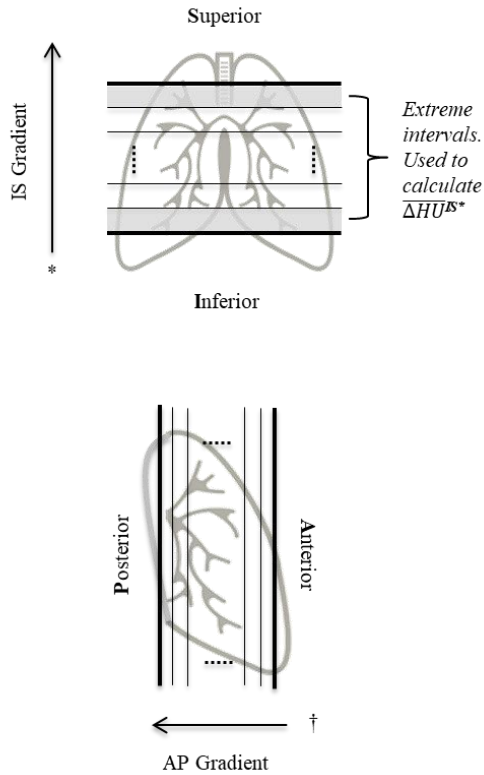
Figure 3.1-7: ΔHU derivation and inferior-superior SAA technical illustration.



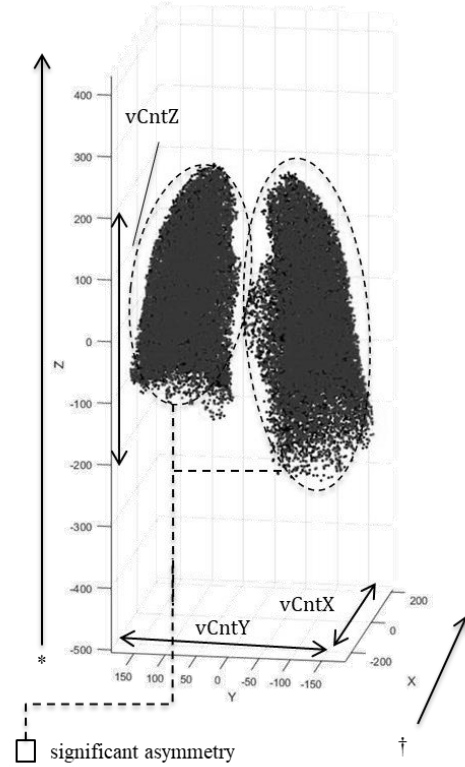
ΔHU derivation and inferior-superior SAA technical illustration. A, rationale behind assumption that simple change in HU (from inspiration to expiration), approximately change in local volume, is associated with ventilation / gas release. B, exemplar demonstration of SAA applied to the inferior-superior axis, and precise definition of ΔHU^{IS*} , notably being a scaled (1/9) difference in average HU between polar voxel intervals.

Figure 3.1-8: ΔHU gradient measure methodology and combinatorial voxel features.

a) Regional PRM, measuring gradients of ΔHU .

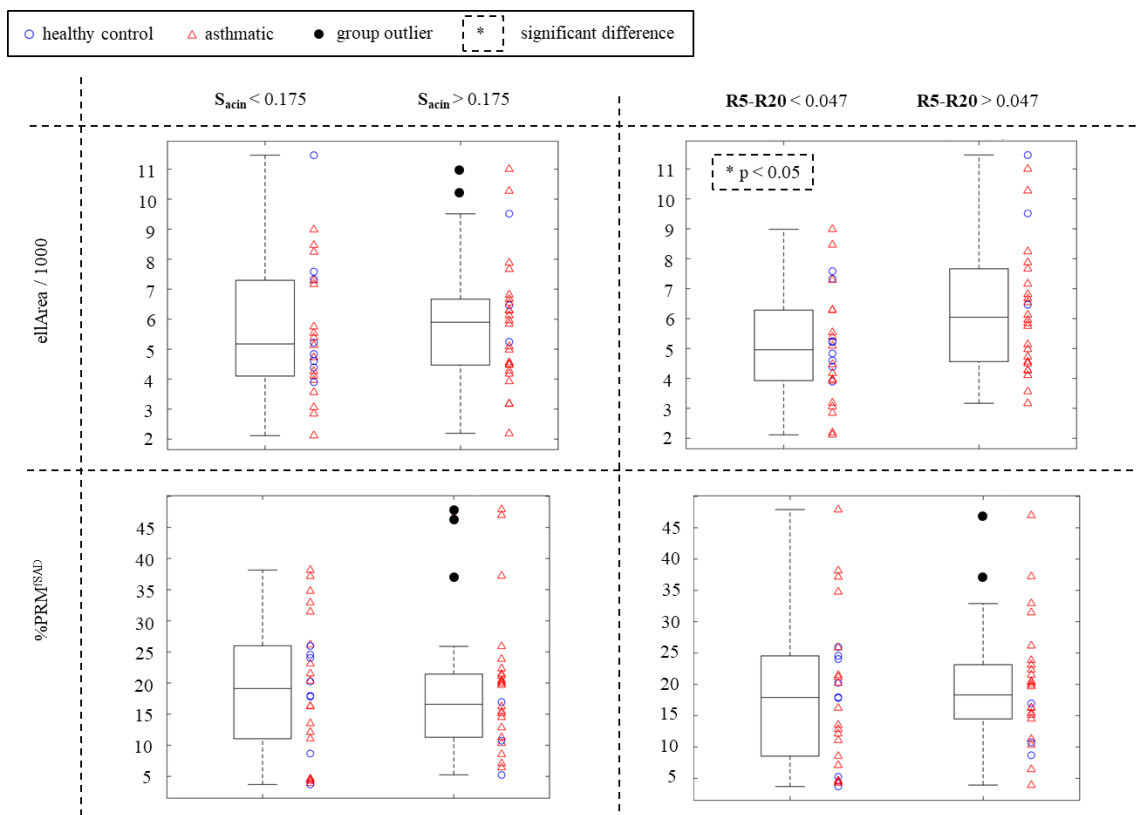


b) Combinatorial features, including inter-lung asymmetry.



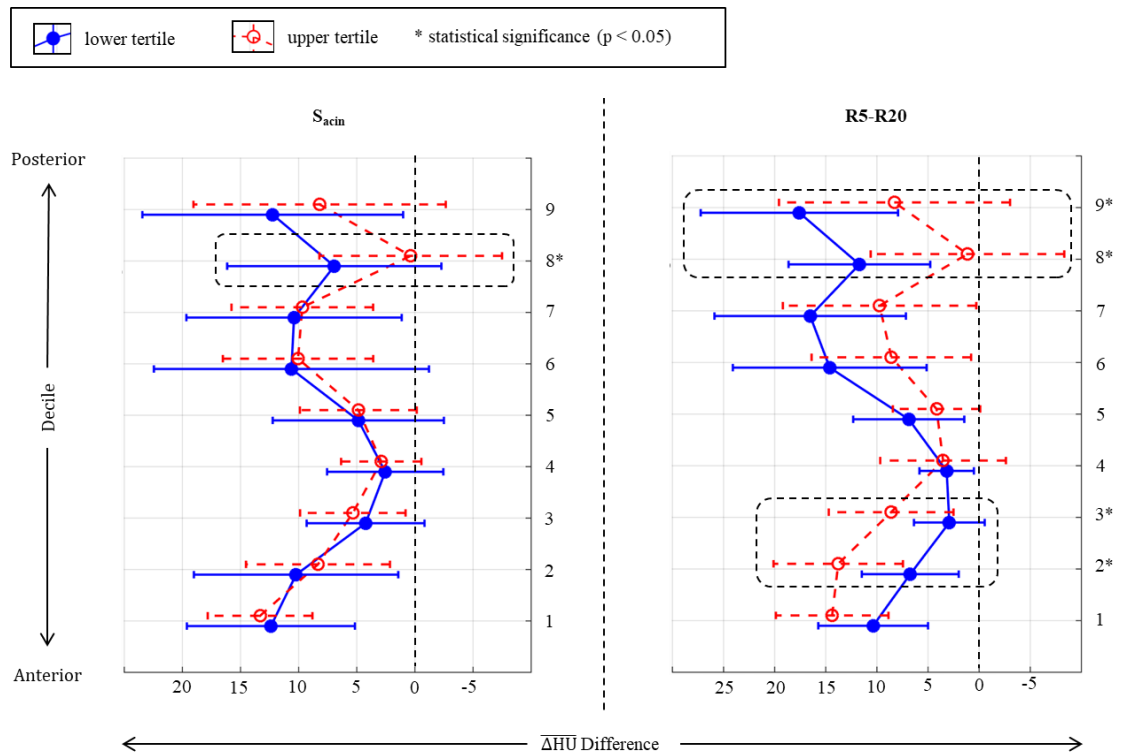
Decile based ΔHU gradient measures and combinatorial voxel features. A, inferior-to-superior (anterior-to-posterior) stratification of lungs, with 15% (10%) of range trimmed at ends; $\overline{\Delta HU}^{IS*}$ calculated as difference of extreme (shaded) strata. B, combinatorial features, based on coordinate axis ranges and voxel counts between segmented left and right lungs (lung asymmetry); subject illustrated chosen for clear case of visual asymmetry.

Figure 3.1-9: PRM ellipse area and PRM^{fSAD} association to VH markers S_{acin} and R5-R20.



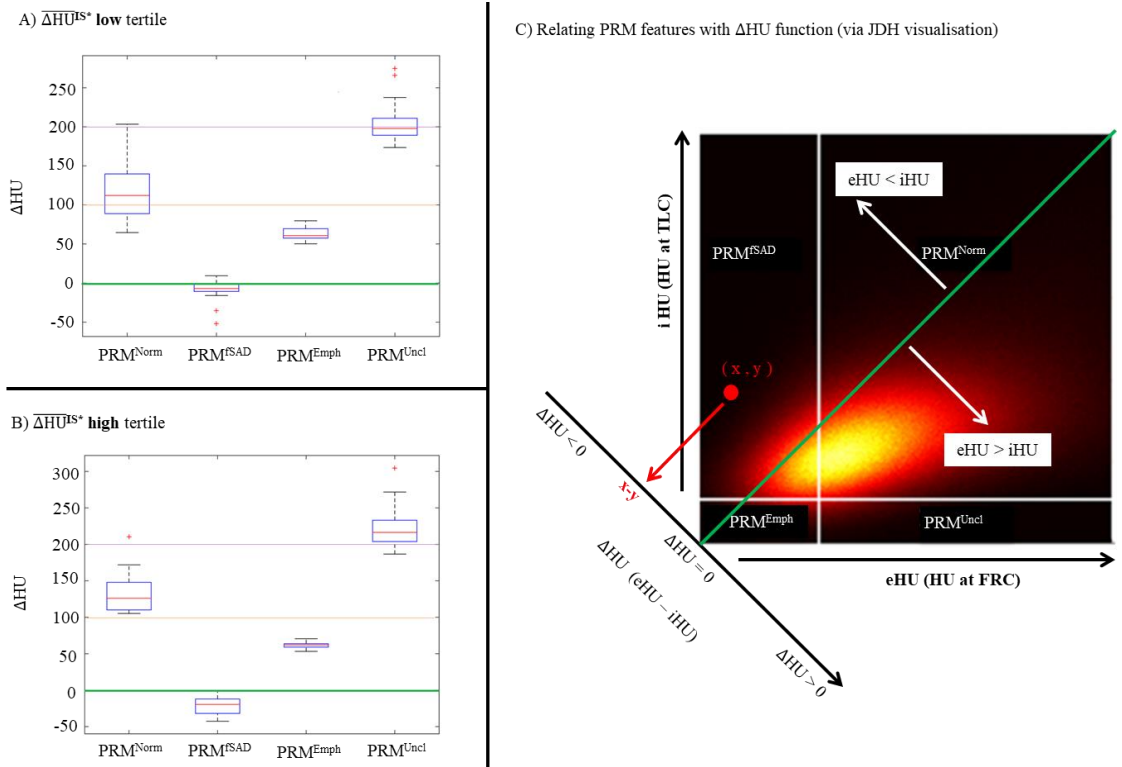
Box plot illustration of ellipse area and %PRM^{fSAD} association with VH markers S_{acin} and R5-R20; groups formed about median value of VH markers. Apparent lack of group separation relative to splitting on median FEV₁/FVC% (**Figure 3.1-1**).

Figure 3.1-10: Decile-wise tertile analysis of $\overline{\Delta HU}$ changes, anterior-to-posterior direction.



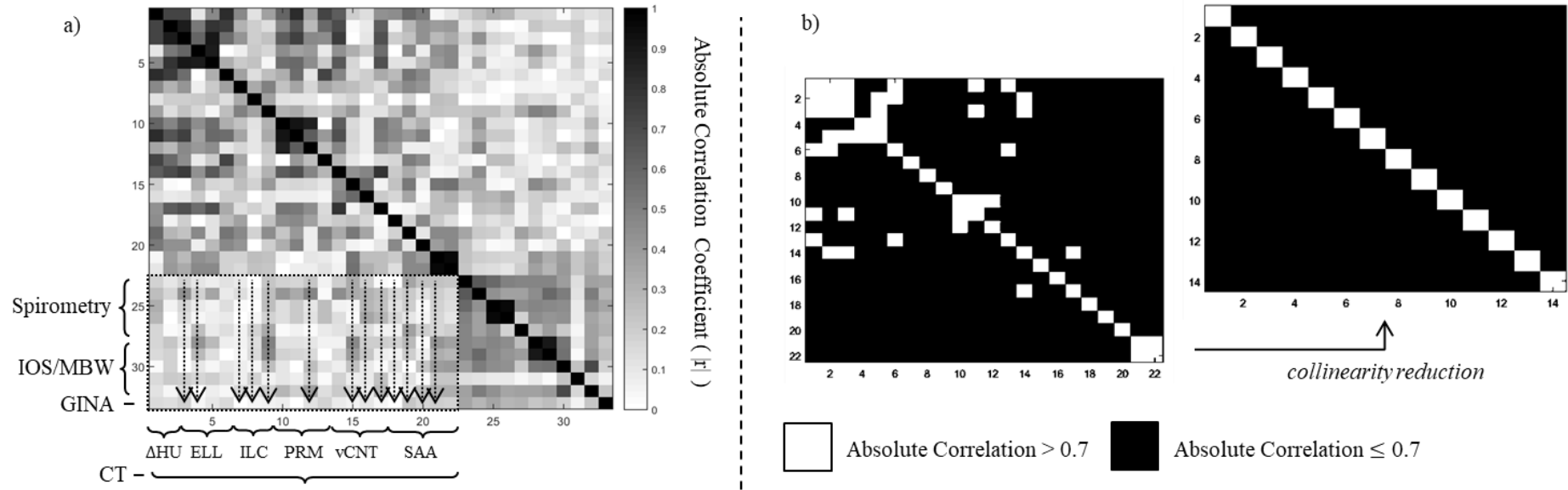
Anterior-to-posterior VH marker focused polar analysis. SAA deciles plotted as mean and standard deviation (bar lengths) of HU changes, highlighting significant regions related to both S_{acin} and R5-R20.

Figure 3.1-11: Comparing ΔHU across PRM^x voxel groups in $\overline{\Delta HU}^{IS*}$ low/high tertiles.



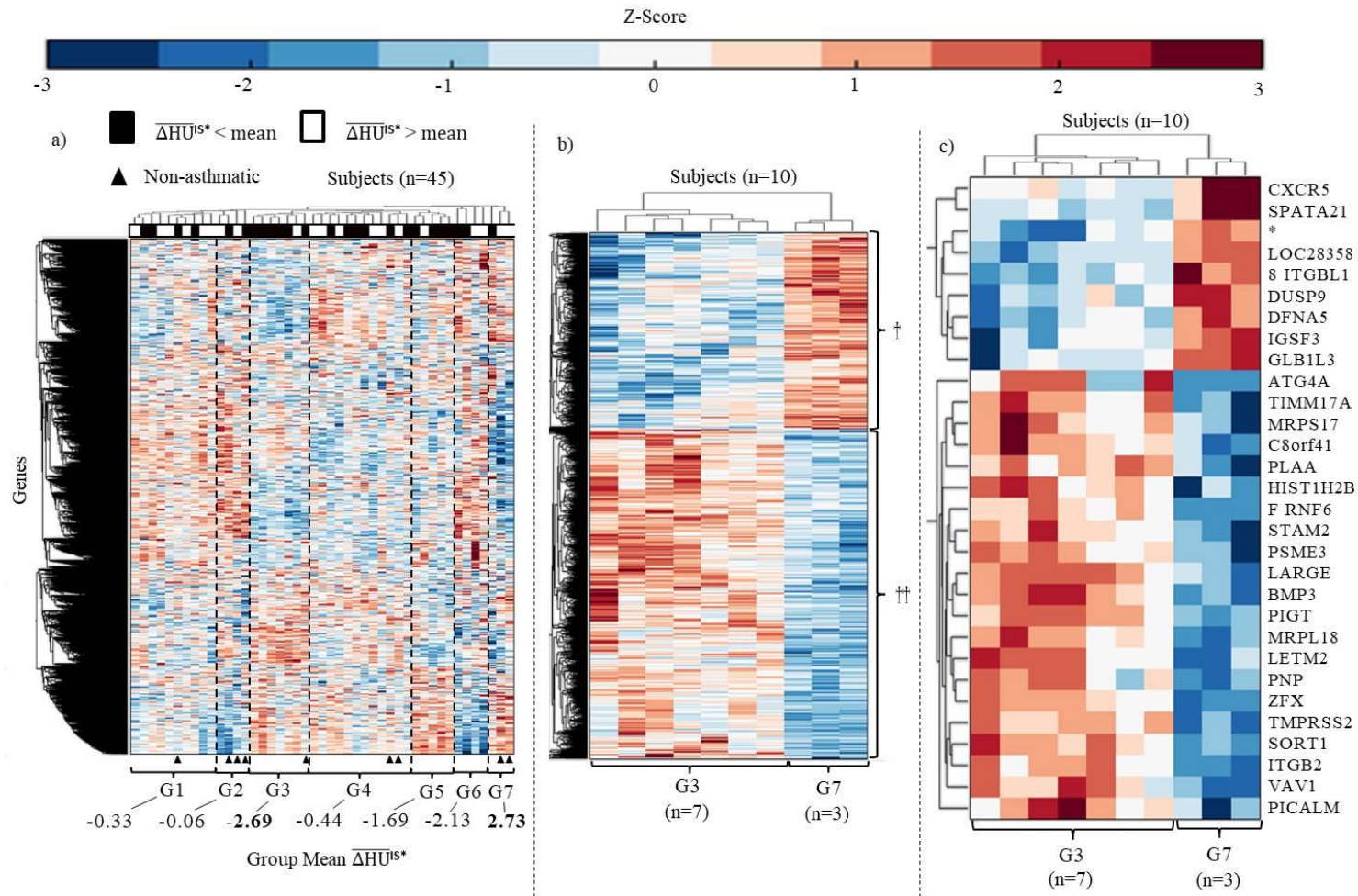
Comparison of ΔHU in PRM^x voxel populations, in $\overline{\Delta HU}^{IS*}$ low and high tertile, and illustration of relationship between ΔHU and PRM^x on exemplar JDH. A, $\overline{\Delta HU}^{IS*}$ low tertile boxplots of ΔHU by PRM^x class. B, $\overline{\Delta HU}^{IS*}$ high tertile boxplots of ΔHU by PRM^x class. Variance appears reduced in B relative to A (common observation: disease brings pressure to biological system, leading to reduced variance). C, illustration of ΔHU projection overlaid onto an exemplar JDH (same case as in **Figure 3.1-1 A**). Essentially it is a projection (x, y) in 2D onto $x - y$ in 1D. Line of no change ($x = y$) plotted in green, and relayed over boxplot figures in A and B; orange and magenta lines annotated for reference in comparing plots (slight difference in vertical scale).

Figure 3.1-12: Complete original data matrix correlation visualization and collinearity reduction from feature selection.



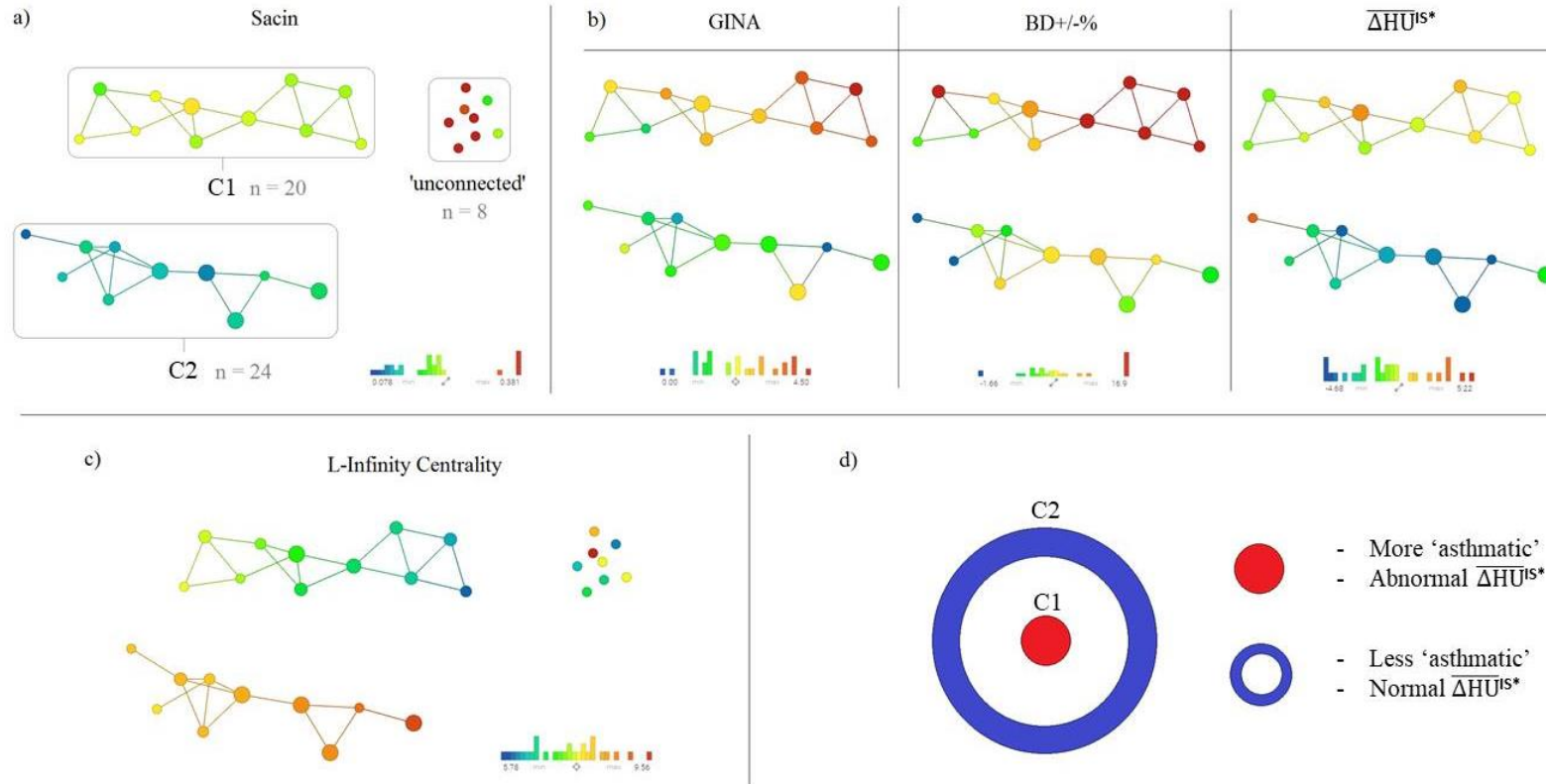
Δ HU, Hounsfield unit change; ELL, JDH ellipse based features; PRM, parametric response map features; ILC, inter-lung features; vCNT, voxel counting features; SAA, stratified axial analysis based features; IOS, impulse oscillometry; MBW, multiple breath washout; GINA, global initiative for asthma (treatment intensity score). A, illustration of absolute correlation matrix, annotated with feature set nomenclature, and indicating features selected for raw value linear regression. B, pairwise-correlation visualization and collinearity reduction on selected subset; binary colouring based on correlation coefficient value about 0.7.

Figure 3.1-13: Hierarchical cluster analysis of nasal gene expression discriminating $\overline{\Delta HU^{IS*}}$.



Hierarchical cluster analysis of nasal gene expression. across participating subjects (n=45), iterating on $\overline{\Delta HU^{IS*}}$ mean extreme groups. A, outcome for all 13,400 gene array sites submitted to clustering algorithm, with non-asthmatic (triangle) annotated above $\overline{\Delta HU^{IS*}}$ group means. B, outcome of cluster analysis iterated on matrix formed from G3 (low $\overline{\Delta HU^{IS*}}$) and G7 (high $\overline{\Delta HU^{IS*}}$), and genes with lowest p-value in Wilcoxon ranksum test. C, 30 genes from B with greatest absolute mean gene expression difference between G3 and G7.

Figure 3.1-14: Centrality based TDA network linking asthma with $\overline{\Delta HU}^{IS*}$



Centrality based TDA network. A, illustrates node groups of interest, C1 (n=20) and C2 (n=24), within network, covering over 80% of the study population. Colour overlay is S_{acin} , determined to be most significant discriminator with respect to C1 and C2. B, feature overlays relevant to asthma and $\overline{\Delta HU}^{IS*}$. All three are not included in the training column set. C, network coloured by L-Infinity centrality, single lens used in network generation. D, abstraction of results; C1 is central in 5D data cloud, C2 lies within some outer annulus of the data cloud.

Table 3.1-1: Clinical characteristics of asthmatic and healthy subjects.

	Healthy volunteers (n=11)	Asthma			
		All (n=41)	GINA 1 (n=8)	GINA 2/3 (n=20)	GINA 4/5 (n=13)
Clinical					
Age (years)	54.1 (± 14.4)	53.7 (± 12.6)	53.0 (± 9.2)	56.5 (± 12.8)	49.7 (± 13.7)
Sex [male/female]	[6/5]	[18/23]	[3/5]	[10/10]	[5/8]
BMI (Kg.m ²)	28.8 (± 4.5)	27.1 (± 4.9)	25.2 (± 4.4)	26.3 (± 4.6)	29.4 (± 5.3)
Atopic [Yes/No]	[3/8] _{1,2}	[30/11]	[6/2] _H	[16/4] _H	[8/5]
Smoking (pack years)	4.2 (± 7.8)	7.3 (± 17.0)	3.0 (± 4.9)	5.0 (± 8.1)	13.4 (± 27.9)
No. of exacerbations. (past 12 months)	-	1.4 (± 2.1)	0.5 (± 0.8)	1.1 (± 2.3)	2.2 (± 2.1)
ACQ-6	-	1.27 (± 1.04)	0.94 (± 0.85)	1.22 (± 0.85)	1.56 (± 1.35)
AQLQ	-	5.37 (± 1.11)	5.94 (± 0.90)	5.28 (± 1.21)	5.15 (± 1.00)
Asthma Duration (years)	-	17.6 (± 16.7)	13.3 (± 9.5)	18.9 (± 18.3)	18.2 (± 18.2)
Beclamethasone Dipropionate Equivalent. ICS Dose (micrograms/24 hours)	-	820 (± 698)	100 (± 282) _{2,3}	650 (± 371) _{1,3}	1523 (± 656) _{1,2}
Physiology					
Post-BD FEV ₁ (L)	3.7 (± 1.0) _{A,1,2}	2.7 (± 0.80) _H	2.5 (± 0.68) _H	2.8 (± 0.85) _H	2.7 (± 0.81)
Post-BD FEV ₁ %	116 (± 19.0) _A	97.2 (± 20.0) _H	99.5 (± 20.8)	97.7 (± 15.5)	95.1 (± 26.3)
Post-BD FEV ₁ /FVC (%)	80 (± 3.2)	74 (± 11)	76 (± 11)	76 (± 7.6)	72 (± 15)
Bronchodilator Response (% FEV ₁)	3.62 (± 3.38) _A	12.88 (± 16.33) _H	8.79 (± 8.19)	11.31 (± 13.52)	17.80 (± 22.77)
RV/TLC (%)	32.10 (± 7.71) _A	38.83 (± 8.68) _H	38.36 (± 8.41)	40.43 (± 9.65)	36.48 (± 7.13)
K _{CO} % pred	96.15 (± 12.69) _A	104.65 (± 15.78) _H	101.13 (± 14.68)	108.30 (± 15.53)	100.92 (± 16.77)
Multiple Breath Washout					
LCI	7.32 (± 1.01)	7.80 (± 1.28)	7.51 (± 1.93)	7.75 (± 1.01)	8.05 (± 1.24)
S _{acin}	0.131 (± 0.052) _A	0.207 (± 0.116) _H	0.193 (± 0.185)	0.203 (± 0.097)	0.220 (± 0.097)
S _{cond}	0.037 (± 0.034)	0.035 (± 0.024)	0.034 (± 0.026)	0.039 (± 0.026)	0.028 (± 0.018)

Impulse Oscillometry					
R5-R20 (kPa·s·L ⁻¹)	0.033 (± 0.029)	0.061 (± 0.058)	0.055 (± 0.035)	0.053 (± 0.042)	0.077 (± 0.085)
AX (kPa/L)	0.291 (± 0.219) _A	0.639 (± 0.752) _H	0.555 (± 0.423)	0.457 (± 0.298)	0.971 (± 1.209)
Induced Sputum					
Eosinophils (%)	0.46 (± 0.30) _{A,2}	10.14 (± 28.17) _H	#	12.94 (± 35.70) _H	6.17 (± 6.76)
Neutrophils (%)	46.3 (± 15.68)	59.73 (± 24.47)	37.1 (± 15.55)	61.72 (± 27.02)	63.96 (± 19.12)

M, male; F, female; BMI, body mass index; FEV, forced expiratory volume; FVC, forced vital capacity; BD, bronchodilator; LCI, lung clearance index; AX, area of reactance. Data expressed as mean (± standard deviation). Attribute normality was tested using one-sample Kolmogorov-Smirnov test over all subjects. Binary group comparisons ('healthy' vs 'all asthmatics') were performed using two sample t-tests for parametric variables, and Mann-Whitney U-tests for non-parametric variables. Non-intersecting group comparisons were performed using one-way ANOVA for parametric variables, and Kruskal-Wallis test for non-parametric variables. Multiple-comparison procedures were performed with Turkey's honest significant difference criterion. Groups with significant separation ($p < 0.05$) indicated by subscripts A (all asthma), H (healthy control), 1, 2, 3 (GINA 1, 2/3 and 4/5 respectively) and * (all other groups).

Table 3.1-2: Computed tomography imaging biomarkers and ventilation heterogeneity based stratification.

	S_{acin}		R5-R20	
	Low (n=29)	High (n=23)	Low (n=32)	High (n=20)
Clinical				
Asthma/ Healthy	[20/9]	[21/2]	[24/8]	[17/3]
Parametric Response Mapping				
%PRM ^{Norm}	0.74 (\pm 0.11)	0.71 (\pm 0.14)	0.72 (\pm 0.14)	0.73 (\pm 0.09)
%PRM ^{fSAD}	0.19 (\pm 0.10)	0.19 (\pm 0.12)	0.19 (\pm 0.12)	0.18 (\pm 0.08)
%PRM ^{Emph}	0.024 (\pm 0.018)	0.038 (\pm 0.037)	0.032 (\pm 0.033)	0.028 (\pm 0.021)
%PRM ^{Uncl}	0.046 (\pm 0.036)	0.063 (\pm 0.030)	0.053 (\pm 0.038)	0.055 (\pm 0.028)
Parametric Response Map Ellipse Properties				
ellMajL	126.2 (\pm 30.0)	130.2 (\pm 28.9)	122.0 (\pm 28.3)	137.5 (\pm 29.0)
ellMinL †	55.1 (\pm 9.4)	56.5 (\pm 11.8)	52.1 (\pm 9.1)	61.4 (\pm 10.0)
ellArea †	5590 (\pm 2067)	5922 (\pm 2251)	5092 (\pm 1763)	6769 (\pm 2312)
ellAngle	0.21 (\pm 0.13)	0.16 (\pm 0.09)	0.19 (\pm 0.11)	0.20 (\pm 0.11)
Ventilation gradient (ΔHU)				
std(Δ HU) ^{AP}	0.070 (\pm 0.078)	0.059 (\pm 0.057)	0.055 (\pm 0.068)	0.080 (\pm 0.070)
$\overline{\Delta$ HU) ^{AP} †	0.473 (\pm 0.207)	0.450 (\pm 0.234)	0.512 (\pm 0.226)	0.386 (\pm 0.183)
std(Δ HU) ^{IS} *	-0.077 (\pm 0.043)	-0.046 (\pm 0.035)	-0.064 (\pm 0.040)	-0.062 (\pm 0.047)
$\overline{\Delta$ HU) ^{IS} *, †	-0.043 (\pm 0.112)	0.021 (\pm 0.099)	-0.051 (\pm 0.100)	0.044 (\pm 0.102)
$\overline{\Delta$ HU) ^{IS} *, †	-2.033 (\pm 4.372)	0.489 (\pm 3.936)	-2.282 (\pm 4.018)	1.267 (\pm 3.987)

PRM, parametric response map; SAA, stratified axial analysis; Data expressed as mean (standard deviation). Attribute normality was tested using one-sample Kolmogorov-Smirnov test over all subjects. Binary group (i.e. S_{acin} low vs. S_{acin} high, and R5-R20 low vs. R5-R20 high) comparisons were performed using two sample t-tests for parametric variables, and Mann-Whitney U-tests for non-parametric variables. Groups with significant separation ($p < 0.05$) of S_{acin} (R5-R20) indicated by * (†).

Table 3.1-3: Reference list describing extracted CT attributes investigated in this study.

CT Feature	Description
$\overline{\Delta HU}$	Average ΔHU over all segmented voxels.
$std(\Delta HU)$	Standard deviation of ΔHU over all segmented voxels.
ellMajL	Length of major axis, of ellipse perceived in JDH.
ellMinL	Length of minor axis, of ellipse perceived in JDH.
ellArea	Area of ellipse perceived in JDH.
ellAngle	Angle (radians) between major axis and horizontal, of ellipse perceived in JDH.
RLmeanDiff	Absolute difference of average ΔHU between the two lungs.
RLstdDiff	Absolute difference of standard deviation in ΔHU between the two lungs.
RLsizeRat	Ratio of lung voxel counts (two lungs), with larger count in the denominator.
$\%PRM^{Norm}$	Percentage of voxels classified as PRM^{Norm} .
$\%PRM^{fSAD}$	Percentage of voxels classified as PRM^{fSAD} .
$\%PRM^{Emph}$	Percentage of voxels classified as PRM^{Emph} .
$\%PRM^{Uncl}$	Percentage of voxels classified as PRM^{Uncl} .
vCnt	Total voxel count (segmented voxel set).
vCntX	Maximum difference in voxel x-coordinates (anterior-posterior measure).
vCntY	Maximum difference in voxel y-coordinates (lateral measure).
vCntZ	Maximum difference in voxel z-coordinates (inferior-superior measure).
$std(\Delta HU)^{AP}$	Gradient of ΔHU standard deviation variability in anterior-posterior direction.
$\overline{\Delta HU}^{AP}$	Gradient of ΔHU mean variability in anterior-posterior direction.
$std(\Delta HU)^{IS}$	Gradient of ΔHU standard deviation variability in inferior-superior direction.
$\overline{\Delta HU}^{IS}$	Gradient of ΔHU mean variability in inferior-superior direction.
$\overline{\Delta HU}^{IS*}$	Difference of ΔHU mean in extreme deciles, in inferior-superior direction.

HU, Hounsfield unit; SAA, stratified axial analysis; fSAD, functional small airways disease; $\overline{\Delta HU}$ = average ΔHU , $std(\Delta HU)$ = ΔHU standard deviation, ellMajL = ellipse major length, ellMinL = ellipse minor length, ellArea = ellipse area, ellAngle = ellipse angle, RLmeanDiff = right-left mean difference, RLstdDiff = right-left standard deviation difference, RLsizeRat = right-left size ratio, $\%PRM^{Norm}$ = percent PRM normal, $\%PRM^{fSAD}$ = percent PRM fSAD, $\%PRM^{Emph}$ = percent PRM emphysema, $\%PRM^{Uncl}$ = percent PRM unclassified, vCnt = voxel count, vCntX = voxel count in x-direction (likewise for vCntY, vCntZ), $std(\Delta HU)^{AP}$ = 1st degree polynomial fitting to standard deviation in x direction (an SAA based measure, likewise for $std(\Delta HU)^{IS}$), $\overline{\Delta HU}^{AP}$ = as $std(\Delta HU)^{AP}$ with averages (likewise for $\overline{\Delta HU}^{IS}$), $\overline{\Delta HU}^{IS*}$ = extreme difference in inferior-superior direction.

Table 3.1-4: CT data across all groups, with treatment rationalized GINA stratification applied to asthma cohort.

	Asthma				
	Control (n=11)	All (n=41)	GINA 1 (n=8)	GINA 2/3 (n=20)	GINA 4/5 (n=13)
PRM					
%PRM ^{Norm}	0.74 (± 0.08)	0.72 (± 0.13)	0.69 (± 0.20)	0.72 (± 0.12)	0.75 (± 0.10)
%PRM ^{fSAD}	0.16 (± 0.08)	0.20 (± 0.11)	0.23 (± 0.17)	0.20 (± 0.09)	0.17 (± 0.10)
%PRM ^{Emph}	0.028 (± 0.018)	0.031 (± 0.031)	0.039 (± 0.040)	0.033 (± 0.034)	0.023 (± 0.019)
%PRM ^{Uncl}	0.072 (± 0.029) _A	0.049 (± 0.034) _H	0.041 (± 0.035)	0.045 (± 0.032)	0.058 (± 0.037)
ELL					
ellMajL	144.4 (± 30.0) _A	123.5 (± 27.8) _H	118.1 (± 33.2)	122.6 (± 28.8)	128.4 (± 24.0)
ellMinL	55.3 (± 10.0)	55.8 (± 10.6)	53.4 (± 10.9)	56.1 (± 12.2)	56.8 (± 8.1)
ellArea	6411 (± 2375)	5556 (± 2060)	5131 (± 2282)	5597 (± 2381)	5755 (± 1401)
ellAngle	0.13 (± 0.07) _A	0.21 (± 0.18) _H	0.25 (± 0.16)	0.21 (± 0.09)	0.18 (± 0.12)
ΔHU					
$\overline{\Delta HU}$	125.98 (± 39.18)	98.16 (± 42.85)	94.70 (± 63.85)	92.12 (± 27.06)	109.60 (± 48.71)
std(ΔHU)	112.58 (± 12.01) _A	102.83 (± 13.30) _H	97.77 (± 17.29)	102.11 (± 13.60)	107.04 (± 9.21)
ILC					
RLmeanDiff	11.33 (± 8.64)	10.42 (± 9.91)	10.05 (± 8.49)	9.46 (± 8.15)	12.13 (± 13.25)
RLstdDiff	4.74 (± 5.24)	4.70 (± 3.53)	4.89 (± 3.70)	4.98 (± 4.00)	4.15 (± 2.79)
RLsizeRat	0.198 (± 0.074)	0.170 (± 0.091)	0.177 (± 0.043)	0.159 (± 0.077)	0.183 (± 0.129)
vCNT					
vCnt†	8.23 (± 2.07)	8.76 (± 2.93)	8.81 (± 3.31)	8.66 (± 2.93)	8.87 (± 2.91)
vCntX	240.6 (± 23.3)	234.5 (± 31.5)	232.9 (± 13.9)	235.8 (± 35.8)	233.6 (± 33.9)
vCntY	343.5 (± 22.0)	335.3 (± 25.9)	327.3 (± 20.5)	327.7 (± 24.6) ₃	351.8 (± 24.4) ₂
vCntZ	436.0 (± 41.7)	439.9 (± 50.2)	449.3 (± 70.0)	442.3 (± 46.7)	430.5 (± 43.9)
SAA					
std(ΔHU) ^{AP}	0.078 (± 0.070)	0.061 (± 0.069)	0.038 (± 0.066)	0.070 (± 0.070)	0.063 (± 0.071)
$\overline{\Delta HU}^{AP}$	0.580 (± 0.182) _A	0.432 (± 0.217) _H	0.470 (± 0.223)	0.397 (± 0.211)	0.462 (± 0.232)
std(ΔHU) ^{IS}	-0.071 (± 0.046)	-0.061 (± 0.041)	-0.073 (± 0.043)	-0.062 (± 0.044)	-0.053 (± 0.037)
$\overline{\Delta HU}^{IS}$	-0.023 (± 0.159)	-0.013 (± 0.095)	-0.084 (± 0.083)	0.003 (± 0.098)	0.008 (± 0.083)
$\overline{\Delta HU}^{IS*}$	-1.460 (± 6.312)	-0.772 (± 3.725)	-2.835 (± 3.085)	-0.488 (± 4.199)	0.061 (± 3.008)

PRM, parametric response map; ELL, ellipse measurements; HU, Hounsfield unit; ILC, inter-lung comparison; vCNT, voxel count; SAA, stratified axial analysis. Data expressed as mean (± standard deviation). Attribute normality was tested using one-sample Kolmogorov-Smirnov test over all subjects. Binary group comparisons ('control' vs 'all') were performed using two sample t-test for parametric variables, and Mann-Whitney U-test for non-parametric variables. Non-intersecting multiple group comparisons were performed using one-way ANOVA for parametric variables, and Kruskal-Wallis test for non-parametric variables. Multiple-comparison procedures were performed with Turkey's honest significant difference criterion. Groups with significant separation ($p < 0.05$) indicated by subscripts A (all asthma), H (healthy control), 1, 2, 3 (GINA 1, 2/3 and 4/5 respectively) and * (all other groups).

Table 3.1-5: Non-CT features with ventilation heterogeneity based stratification.

	S_{acin}		R5-R20	
	Low (n=29)	High (n=23)	Low (n=32)	High (n=20)
Clinical				
GINA [H/1/2/3/4/5]	[9/5/0/8/5/2]	[2/3/3/9/5/1]	[8/6/1/10/6/1]	[3/2/2/7/4/2]
Age	50.7 (± 12.9)	57.6 (± 11.9)	52.0 (± 12.3)	56.6 (± 13.5)
Sex [M/F]	[12/17]	[12/11]	[13/19]	[11/9]
Weight (kg) †	79.0 (± 20.0)	76.8 (± 13.1)	72.9 (± 15.4)	86.3 (± 17.0)
Height (cm)	168.5 (± 11.2)	166.3 (± 9.7)	166.6 (± 11.2)	169.0 (± 9.6)
BMI†	27.6 (± 5.3)	27.3 (± 4.4)	26.2 (± 4.2)	29.5 (± 5.2)
Atopy [Yes/No]	[16/13]	[17/6]	[22/12]	[11/9]
Pack Years†	6.25 (± 18.45)	7.15 (± 10.98)	2.85 (± 5.76)	12.71 (± 22.91)
No. of Ex. (past yr)	1.1 (± 1.7)	1.0 (± 2.2)	1.0 (± 1.8)	1.2 (± 2.2)
Asthma Duration	15.4 (± 15.8)	19.7 (± 17.7)	18.9 (± 15.8)	15.8 (± 18.2)
Equiv. CS Dose	875.0 (± 779.3)	766.7 (± 625.6)	812.5 (± 726.1)	829.4 (± 678.0)
Spirometry				
FEV ₁ % *,†	108.2 (± 17.2)	92.6 (± 22.6)	107.5 (± 19.9)	91.4 (± 19.6)
FEV ₁ /FVC(%)*	79.3 (± 5.2)	71.1 (± 12.6)	77.1 (± 6.8)	73.3 (± 13.6)
Pre-BD FEV ₁ (L)*, †	3.03 (± 0.90)	2.23 (± 0.78)	2.92 (± 0.94)	2.28 (± 0.79)
Post-BD FEV ₁ (L)*, †	3.19 (± 0.91)	2.56 (± 0.79)	3.10 (± 0.93)	2.60 (± 0.80)
%BD+/-*	5.99 (± 9.2)	17.13 (± 18.48)	7.36 (± 9.93)	16.60 (± 19.75)
MBW				
LCI*,†	7.17 (± 0.91)	8.37 (± 1.28)	7.40 (± 1.04)	8.17 (± 1.39)
S _{acin} *	0.118 (± 0.039)	0.282 (± 0.102)	0.176 (± 0.097)	0.214 (± 0.126)
S _{cond}	0.032 (± 0.026)	0.040 (± 0.025)	0.038 (± 0.026)	0.030 (± 0.026)
IOS				
R5-R20†	0.043 (± 0.045)	0.070 (± 0.062)	0.025 (± 0.021)	0.103 (± 0.057)
AX†	0.414 (± 0.372)	0.757 (± 0.923)	0.278 (± 0.169)	1.025 (± 0.929)
Sputum				
Eos	0.294 (± 1.443)	0.788 (± 1.896)	0.385 (± 1.525)	0.714 (± 1.867)
Neut	56.10 (± 25.73)	59.22 (± 21.55)	58.50 (± 23.60)	56.23 (± 24.34)

M, male; F, female; BMI, body mass index; FEV, forced expiratory volume; FVC, forced vital capacity; BD, bronchodilator; LCI, lung clearance index; AX, area of reactance; Eos, eosinophil count; Neut, neutrophil count. Data expressed as mean (± standard deviation). Attribute normality was tested using one-sample Kolmogorov-Smirnov test over all subjects. Binary group (i.e. S_{acin} low vs. S_{acin} high, and R5-R20 low vs. R5-R20 high) comparisons were performed using two sample t-test for parametric variables, and Mann-Whitney U-test for non-parametric variables. Groups with significant separation (p < 0.05) of S_{acin} (R5-R20) indicated by * (†).

Table 3.1-6: Computed tomography imaging biomarkers and ventilation heterogeneity based stratification (full).

	S_{acin}		R5-R20	
	Low (n=29)	High (n=23)	Low (n=32)	High (n=20)
PRM				
%PRM ^{Norm}	0.74 (± 0.11)	0.71 (± 0.14)	0.72 (± 0.14)	0.73 (± 0.09)
%PRM ^{fSAD}	0.19 (± 0.10)	0.19 (± 0.12)	0.19 (± 0.12)	0.18 (± 0.08)
%PRM ^{Emph}	0.024 (± 0.018)	0.038 (± 0.037)	0.032 (± 0.033)	0.028 (± 0.021)
%PRM ^{Uncl}	0.046 (± 0.036)	0.063 (± 0.030)	0.053 (± 0.038)	0.055 (± 0.028)
ELL				
ellMajL	126.2 (± 30.0)	130.2 (± 28.9)	122.0 (± 28.3)	137.5 (± 29.0)
ellMinL †	55.1 (± 9.4)	56.5 (± 11.8)	52.1 (± 9.1)	61.4 (± 10.0)
ellArea †	5590 (± 2067)	5922 (± 2251)	5092 (± 1763)	6769 (± 2312)
ellAngle	0.21 (± 0.13)	0.16 (± 0.09)	0.19 (± 0.11)	0.20 (± 0.11)
ΔHU				
ΔHŪ	100.42 (± 48.02)	108.62 (± 36.97)	103.69 (± 45.02)	104.62 (± 41.46)
std(ΔHU) †	102.37 (± 14.32)	108.07 (± 12.01)	101.74 (± 12.55)	109.93 (± 13.81)
ILC				
RLmeanDiff	8.48 (± 9.37)	13.30 (± 9.36)	11.05 (± 10.38)	9.91 (± 8.36)
RLstdDiff	4.15 (± 3.94)	5.40 (± 3.81)	5.53 (± 4.33)	3.39 (± 2.68)
RLsizeRat †	0.184 (± 0.088)	0.166 (± 0.088)	0.201 (± 0.087)	0.135 (± 0.073)
vCnt				
vCnt ¥	8.67 (± 2.99)	8.62 (± 2.51)	8.97 (± 2.80)	8.12 (± 2.68)
vCntX †	231.9 (± 25.3)	240.7 (± 34.7)	225.9 (± 27.1)	251.7 (± 27.6)
vCntY	336.9 (± 27.5)	337.2 (± 22.4)	335.8 (± 25.3)	338.9 (± 25.4)
vCntZ †	438.4 (± 45.7)	439.9 (± 52.1)	451.3 (± 48.4)	419.5 (± 41.9)
SAA				
std(ΔHU) ^{AP}	0.070 (± 0.078)	0.059 (± 0.057)	0.055 (± 0.068)	0.080 (± 0.070)
ΔHŪ ^{AP} †	0.473 (± 0.207)	0.450 (± 0.234)	0.512 (± 0.226)	0.386 (± 0.183)
std(ΔHU) ^{IS}	-0.077 (± 0.043)	-0.046 (± 0.035)	-0.064 (± 0.040)	-0.062 (± 0.047)
ΔHŪ ^{IS} *,†	-0.043 (± 0.112)	0.021 (± 0.099)	-0.051 (± 0.100)	0.044 (± 0.102)
ΔHŪ ^{IS*} *,†	-2.033 (± 4.372)	0.489 (± 3.936)	-2.282 (± 4.018)	1.267 (± 3.987)

¥ values expressed have multiplier 10⁶.

PRM, parametric response map; ELL, ellipse measurements; HU, Hounsfield unit; ILC, inter-lung comparison; vCNT, voxel count; SAA, stratified axial analysis. Data expressed as mean (± standard deviation). Attribute normality was tested using one-sample Kolmogorov-Smirnov test over all subjects. Binary group (i.e. S_{acin} low vs. S_{acin} high, and R5-R20 low vs. R5-R20 high) comparisons were performed using two sample t-test for parametric variables, and Mann-Whitney U-test for non-parametric variables. Groups with significant separation (p < 0.05) of S_{acin} (R5-R20) indicated by * (†).

Table 3.1-7: Summary of linear statistical analyses: average correlation and linear regression (raw value & PCA).

Attribute ID	Corr.	LR (subset)		LR (PCA)		PCA-LR Coef. (abs. value)					
	\bar{r}	R^2	p	R^2	p	P1	P2	P3	P4	P5	P6
FEV ₁ % ₂₃	.158	.45	p < .05	.19	p > .5	-	-	-	-	-	-
FEV ₁ /FVC(%) ₂₄	.244	.63	p < .001	.34	p < .005	.37	.19	.40	.09	.08	.00
Pre-BD FEV ₁ ₂₅	.141	.61	p < .001	.29	p < .05	.12	.16	.10	.21	.08	.44
Post-BD FEV ₁ ₂₆	.133	.59	p < .001	.31	p < .05	.09	.10	.10	.30	.15	.42
%BD+/- ₂₇	.133	.41	p < .05	.16	p > .05	-	-	-	-	-	-
R5-R20 ₂₈	.203	.59	p < .001	.29	p < .05	.00	.52	.07	.09	.07	.01
AX ₂₉	.166	.51	p < .05	.19	p > .05	-	-	-	-	-	-
LCI ₃₀	.133	.53	p < .05	.19	p > .05	-	-	-	-	-	-
S _{cond} ₃₁	.132	.24	p > .05	.08	p > .5	-	-	-	-	-	-
S _{acin} ₃₂	.159	.52	p < .05	.28	p < .05	.12	.15	.42	.09	.09	.22
GINA ₃₃	0.103	0.46	p < .05	.12	p > .05	-	-	-	-	-	-

FEV, forced expiratory volume; FVC, forced vital capacity; BD, bronchodilator; AX, area of reactance; LCI, lung clearance index; GINA, global initiative for asthma; Single feature CT linear statistical analysis overview. \bar{r} = Pearson's correlation coefficient; R^2 = prediction strength as variance explained; p = p-value from F statistic (test likelihood of significant linear relationship); P1 – P6 = principal components selected by Kaiser rule (above mean variance explained in PCA), in order of variance explained. Corr. (correlation) column illustrates average correlation as crude measure of relation. LR columns (3, 4, 5 and 6) represent linear regression outcome using 14 feature subset and PCA principal components as predictor variables, and non-CT (first column) attributes as target variables. Final columns (7... 12) are absolute values of coefficients in linear regression model for PCA, used with **Table 3.1-9** to infer connection between target variables and CT features. Coefficients for 14 feature linear regression are listed in **Table 3.1-8**.

Table 3.1-8: Linear regression co-efficient table for linearity reduced CT subset regressions.

		LR Coefficients (absolute value) by Correlation Matrix ID (top row)													
		3	4	7	8	9	12	15	16	17	18	19	20	21	
LR Target Variable	FEV ₁ %	.97	.49	.14	.07	.04	.22	.28	.02	.48	.07	.06	.63	.23	
	FEV ₁ /FVC(%)	1.4	.64	.29	.28	.02	.31	.23	.13	.35	.13	.49	.55	.02	
	Pre-BD FEV ₁	1.5	.80	.38	.10	.17	.02	.14	.45	.60	.21	.41	.45	.11	
	Post-BD FEV ₁	1.4	.72	.35	.05	.11	.04	.15	.48	.61	.22	.46	.32	.18	
	%BD+/-	.80	.57	.29	.23	.29	.05	.13	.07	.21	.18	.06	.59	.22	
	R5-R20	.80	.83	.41	.04	.44	.09	.26	.02	.21	.03	.59	.03	.28	
	AX	1.0	.89	.43	.16	.42	.15	.16	.09	.19	.08	.71	.12	.27	
	LCI	1.1	1.1	.05	.05	.21	.15	.15	.14	.05	.06	.11	.36	.51	
	S _{acin}	.98	.95	.49	.07	.12	.23	.03	.13	.33	.01	.64	.20	.53	
	GINA	1.2	.41	.41	.21	.10	.37	.42	.19	.27	.28	.11	.49	.28	

LR, linear regression; FEV, forced expiratory volume; FVC, forced vital capacity; BD, bronchodilator; AX, area of reactance; LCI, lung clearance index; GINA, global initiative for asthma. Absolute value of linear regression coefficients over all 14 features selected for raw value linear regression. Refer to **Table 3.1-9** for linking ID numbers (heading row) to CT features. Each predictor feature (column) has highest occurring value in bold. Relational strength between individual predictor and target variables may be assessed through cross-referencing cell co-ordinates with relative co-efficient magnitude for a given feature.

Table 3.1-9: Principal component loading scores for components used in linear regression.

	Attribute ID	PC1	PC2	PC3	PC4	PC5	PC6
ΔHU	$\overline{\Delta HU}_1$	+.31	-.11	+.17	-.03	-.17	+.15
	std(ΔHU) ₂	+.31	+.10	+.13	+.19	+.08	+.11
ELL	ellMajL ₃	+.33	+.09	-.01	+.06	+.12	+.06
	ellMinL ₄	+.17	+.33	-.24	-.01	+.17	+.04
	ellArea ₅	+.27	+.23	-.12	+.05	+.16	+.03
	ellAngle ₆	-.25	+.09	-.30	-.12	+.15	-.07
ILC	RLmeanDiff ₇	+.17	-.21	+.12	-.21	+.34	-.15
	RLstdDiff ₈	-.05	-.19	+.04	-.24	+.63	-.04
	RLsizeRat ₉	+.12	-.25	-.00	-.19	+.34	+.00
PRM	%PRM ^{Norm} ₁₀	+.27	-.02	-.29	-.24	-.15	+.10
	%PRM ^{fSAD} ₁₁	-.32	+.05	+.08	+.17	+.15	-.16
	%PRM ^{Emph} ₁₂	-.19	+.01	+.42	+.23	+.11	-.07
	%PRM ^{Uncl} ₁₃	+.18	-.10	+.44	+.13	-.03	+.18
vCNT	vCnt ₁₄	-.29	+.05	+.17	-.28	-.08	+.17
	vCntX ₁₅	-.06	+.35	+.14	-.22	-.01	+.27
	vCntY ₁₆	-.07	+.22	+.11	-.40	+.08	+.56
	vCntZ ₁₇	-.25	-.12	+.19	-.25	-.15	+.06
SAA	std(ΔHU) ^{AP} ₁₈	-.14	+.16	+.04	+.41	+.34	+.32
	$\overline{\Delta HU}$ ^{AP} ₁₉	+.13	-.36	+.14	+.14	-.03	+.26
	std(ΔHU) ^{IS} ₂₀	+.19	+.03	+.30	-.35	-.09	-.37
	$\overline{\Delta HU}$ ^{IS} ₂₁	+.07	+.38	+.23	-.00	+.06	-.27
	$\overline{\Delta HU}$ ^{IS*} ₂₂	+.09	+.38	+.23	-.07	+.04	-.24

PC, principal component; PRM, parametric response map; ELL, ellipse measurements; HU, Hounsfield unit; ILC, inter-lung comparison; vCNT, voxel count; SAA, stratified axial analysis. Signed loading scores in PCA over all 22 CT features, as they load onto the first 6 principal components submitted to linear regression in linear statistical analyses. Magnitudes may be considered indicative of feature (row) relational strength to principal component (column) formation, which in turn may be associatively connected to predicted variables (see **Table 3.1-7** far right column).

Table 3.1-10: Linear discriminant analysis based classification percentage accuracy with selected feature sets.

	S_{acin} (% accuracy)	R5-R20 (% accuracy)
Clinical characteristics	63	77
Post BD spirometry	73	69
CT biomarkers	79	83
Clinical + Spirometry	75	77
Clinical + CT	81	87
Spirometry + CT	79	85
Clinical + Spirometry + CT	85	85

Data expressed as percentage of subjects correctly classified by best possible linear discriminant from linear discriminant analysis. Feature sets use attributes representing clinical (age, smoking history [pack years] and weight [kg]), spirometry (FEV₁% and FEV₁/FVC(%)) and CT (ellMinL, ellArea, std(Δ HU), RLsizeRat, vCntX, vCntZ, $\overline{\Delta$ HU^{AP}, std(Δ HU)^{IS} and $\overline{\Delta$ HU^{IS} [features differentiating S_{acin} or R5-R20]) data.

Table 3.1-11: Linear discriminant analysis coefficients on combined feature sets.

	S_{acin}	R5-R20
Clinical characteristics		
Age (completed years)	-0.014	+0.010
Weight (kg)	+0.028[▼]	+0.014
Smoking (pack years)	-0.0060	-0.016
Post BD spirometry		
FEV ₁ %	+0.017	+0.022
FEV ₁ /FVC(%)	+0.014	+0.013
CT biomarkers		
std(Δ HU)	-0.021	+0.0056
ellMinL	+0.012	+0.044[▼]
ellArea	-0.032[▲]	-0.080[▲]
RLsizeRat	-0.00040	+0.014
vCntX	-0.00073	-0.015
vCntZ	-0.022	+0.0047
$\overline{\Delta HU}^{AP}$	+0.0069	-0.00054
std(Δ HU) ^{IS}	-0.011	+0.028
$\overline{\Delta HU}^{IS}$	-0.0020	-0.022

Coefficients of linear discriminant analysis (LDA) applied across all selected features. S_{acin} or R5-R20 below mean groups are projected in positive direction. Above mean groups are projected in negative direction. Thus more positive coefficients may be associated with less ventilation heterogeneity (VH), and more negative coefficients with more VH. Extremes of greatest magnitude are emphasised in **bold**. Superscript **▲** indicates most associated with **high VH** (relatively most negative coefficient), and **▼** with **low VH** (relatively most positive coefficient).

Table 3.1-12: Genes highlighted from GE hierarchical cluster analysis (Figure 3.1-13).

Gene	Description	Chromosome	Connection to lung disease?
CXCR5	C-X-C Motif Chemokine Receptor 5	11q23	Inflammation.
SPATA21	Spermatogenesis Associated 21	1p36	-
*	Gene array site unlabelled	-	-
LOC283588	Uncharacterised	14q32	-
ITGBL1	Integrin Subunit Beta Like 1	13q33	Obesity. Smoking.
DUSP9	Dual Specificity Phosphatase 9	Xq28	-
DFNA5	Deafness Associated Tumor Suppressor	7p15	-
IGSF3	Immunoglobulin Superfamily Member 3	1p13	Emphysema.
GLB1L3	Galactosidase Beta 1 Like 3	11q25	-
ATG4A	Autophagy Related 4A Cysteine Peptidase	Xq22	Lung cancer.
TIMM17A	Translocase Of Inner Mitochondrial Membrane 17A	1q32	Lung cancer.
MRPS17	Mitochondrial Ribosomal Protein S17	7p11	-
C8orf41	TELO2 Interacting Protein 2 (TTI2)	8p12	-
PLAA	Phospholipase A2 Activating Protein	9p21	-
HIST1H2BF	Histone Cluster 1 H2B Family Member F	6p22	-
RNF6	Ring Finger Protein 6	13q12	Alveolar macrophage.
STAM2	Signal Transducing Adaptor Molecule 2	2q23	-
PSME3	Proteasome Activator Subunit 3	17q21	-
LARGE	LARGE Xylosyl- And Glucuronyltransferase 1 (LARGE1)	22q12	-
BMP3	Bone Morphogenetic Protein 3	4q21	Fetal lung development.
PIGT	Phosphatidylinositol Glycan Anchor Biosynthesis Class T	20q13	-
MRPL18	Mitochondrial Ribosomal Protein L18	6q25	Alveolar macrophage.
LETM2	Leucine Zipper And EF-Hand Containing Transmembrane Protein 2	8p11	-
PNP	Purine Nucleoside Phosphorylase	4q11	-
ZFX	Zinc Finger Protein, X-Linked	Xp22	Lung cancer.
TMPRSS2	Transmembrane Protease, Serine 2	21q22	Influenza.
SORT1	Sortilin 1	1p13	-
ITGB2	Integrin Subunit Beta 2	21q22	-
VAV1	Vav Guanine Nucleotide Exchange Factor 1	19p13	Protein interaction.
PICALM	Phosphatidylinositol Binding Clathrin Assembly Protein	11q14	-

* Gene label missing in data set (site unnamed).

Explanative listing of 30 gene expression attributes highlighted in hierarchical cluster analysis and cross examination against $\Delta\overline{H}U^{IS*}$ (**Figure 3.1-13**). Table columns detail (left to right) gene name, gene description and chromosomal locale (as detailed on ‘genecards.org’), and any known association to respiratory disease / asthma.

Table 3.1-13: TDA network group comparison (refer to Figure 3.1-14).

	C1 (Inner, n=20)	C2 (Outer, n=24)
Clinical		
GINA	2.900 (\pm 1.410)	1.667 (\pm 1.659)
Age	56.700 (\pm 13.742)	48.625 (\pm 11.447)
Sex	0.600 (\pm 0.503)	0.375 (\pm 0.495)
Weight (kg)	77.600 (\pm 14.424)	76.500 (\pm 19.377)
Height (cm)	167.300 (\pm 11.197)	168.125 (\pm 11.333)
BMI	27.613 (\pm 3.596)	26.803 (\pm 4.761)
FEV₁%	93.154 (\pm 14.696)	113.428 (\pm 17.551)
Atopy: Yes/No	0.750 (\pm 0.444)	0.542 (\pm 0.509)
(FEV₁/FVC)%	72.676 (\pm 7.176)	80.527 (\pm 4.314)
Pre-BD FEV₁(L)	2.308 (\pm 0.537)	3.260 (\pm 0.903)
Post-BD FEV₁(L)	2.635 (\pm 0.633)	3.380 (\pm 0.924)
%BD+/-	14.988 (\pm 14.428)	4.115 (\pm 7.127)
MBW		
LCI	7.930 (\pm 0.675)	6.932 (\pm 0.767)
S_{acin}	0.204 (\pm 0.041)	0.119 (\pm 0.055)
S _{cond}	0.041 (\pm 0.026)	0.029 (\pm 0.027)
IOS		
R5-R20	0.059 (\pm 0.026)	0.026 (\pm 0.027)
AX	0.509 (\pm 0.244)	0.281 (\pm 0.185)
Sputum		
Log(eos)	0.682 (\pm 1.649)	0.078 (\pm 0.994)
Neut	56.867 (\pm 24.210)	57.312 (\pm 22.325)
CT		
zPFavg	0.025 (\pm 0.090)	-0.071 (\pm 0.108)
ΔHU^{IS*}	0.305 (\pm 3.792)	-2.956 (\pm 4.240)

GINA, global initiative for asthma (treatment intensity); BMI, body mass index; FEV, forced expiratory volume; FVC, forced vital capacity; BD, bronchodilator; MBW, multiple breath washout; LCI, lung clearance index; IOS, impulse oscillometry; eos, eosinophil count; Neut, neutrophil count; HU, Hounsfield unit. Attribute normality was tested using one-sample Kolmogorov-Smirnov test over all subjects. Binary group comparisons performed using two sample t-test for parametric variables, and Mann-Whitney U-test for non-parametric variables. Groups with significant separation ($p < 0.05$) highlighted in bold type.

3.2 Topological data analysis to visualise the lungs and small airways in adult asthma

Abstract

Background: Asthma is characterised by ventilation heterogeneity (VH) in the lungs, across various anatomical depths of the airway tree. Tree depth can be measured using Strahler order (SO). Visualisation of the lungs and lung function with SO may be useful for evaluating the specific contribution of the small airways to VH in asthma.

Objectives: We sought to develop an approach to visualise VH in the lungs and small airways, by applying topological data analysis (TDA) to CT imaging data incorporating SO voxel proximity. Furthermore, we aimed to investigate patterns of Hounsfield unit change (ΔHU) and small airways disease marker PRM^{fSAD} .

Methods: 2 asthmatic and 2 non-asthmatic subjects underwent inspiratory (expiratory) CT scanning, at TLC (FRC). VH was quantified using multiple breath washouts. Subjects were chosen to represent a young healthy lung (no VH), ageing healthy lung (low VH), mild asthmatic lung (mild VH) and severe asthmatic lung (high VH). Parametric response map (PRM) registration was performed to calculate ΔHU and PRM^{fSAD} . A patient based conducting airway tree model was used to generate extended 16 generation airway centrelines for calculating SO proximity. TDA was then applied to produce compressed visualisations of the lungs.

Results: TDA networks provided a tool to observe patterns of small airway ΔHU and PRM^{fSAD} using SO proximity. Asthmatic and non-asthmatic subjects differentiated with respect to ΔHU inferior-superior distribution, PRM^{fSAD} anterior-posterior distribution and ΔHU heterogeneity.

Conclusion: We developed and applied a novel method using TDA to visualise the lungs and small airways, and provided insight into small airway VH in adult asthma.

3.2.1 Introduction

Asthma is a complex heterogeneous disease that involves combinations of factors from multiple length scales; systems medicine aims to integrate data from disparate sources to better understand such complex chronic diseases (41,42). Data driven phenotyping may then lead to better understanding of the disease (1), and provide a basis for the development of more effective targeted therapies (9). In this exploratory study we present novel methodology to visualise data from the small airways to the entire lung simultaneously, and demonstrate basic visual discrimination between asthmatic and non-asthmatic subjects.

Ventilation heterogeneity (VH) is a major characteristic of asthma (16,275), and may be captured using imaging approaches that quantify and regionalise lung ventilation, such as hyper polarised 3-Helium/129-Xenon magnetic resonance imaging (MRI), Oxygen enhanced MRI and single photon emission computed tomography (SPECT-CT) (27,33,276–278). VH can be measured clinically at the mouth using tidal breathing techniques such as multiple breath washout (MBW) (279,280). International guidelines for quality control and assurance of tidal breathing markers of VH derived from MBW have been proposed (20). In this study we used MBW markers LCI, S_{cond} and S_{acin} to characterise patient VH.

Computed tomography (CT) imaging of the lungs has been utilised to study lung structure and function relationships in asthma (284,285). Image registration has been applied to inspiratory and expiratory CT imaging to derive indices of functional small airways disease (32,190,274,286). Parametric response mapping (PRM) has been developed to characterise spatial deformation of regional volumes between different states, e.g. the lungs, between functional residual capacity (FRC) and total lung capacity (TLC) (32,190,286). Regional ventilation may then be approximated by, for example, Hounfield unit (HU) change (ΔHU) between FRC and TLC states (30).

Cancer heart and soft tissue environment (Chaste) has been developed by Oxford computer science for computer simulations in the domain of Biology (202). Algorithms for the modelling of a complete bronchial tree from an initial (major airways) image segmentation are included in this software package and have been applied to provide useful patient specific models of the airways (37). Complete airway tree approximation

may be used to measure voxel spatial proximity to the small airways, e.g. using Strahler order associated to simulated branches.

Combining Chaste airway tree modelling with PRM produces numerous voxel based features, leading to innately high dimensional data sets for analysis. Ayasdi is a software package providing tools for topological data analysis (TDA) that has proved useful in the visualisation and qualitative analysis of high dimensional data in medical research (50). TDA has recently been applied by our group to determine pathological heterogeneity in asthma (261).

The purpose of this study was to visualise patterns of ventilation and disease within the human lung through the application of TDA to high volume PRM registered voxel data sets; incorporating Strahler order (SO) approximation using Chaste extended centre line models, to highlight and visualise regions most representative of the smaller airways. See **Figure 3.2-1** for conceptual overview of data sources and TDA outcomes.

Specifically we aimed to:

- i. Develop an algorithm, incorporating Chaste extended centre line models of the complete airway tree, to measure voxel 'SO proximity'.
- ii. Apply TDA to sets of PRM registered lung voxels; incorporating said SO proximity measure, to identify and analyse small airways regions in visualisations.

We hypothesised that ΔHU and PRM derived functional small airways disease marker PRM^{fSAD} would discriminate between asthmatic and non-asthmatic subjects in the resultant visualisations.

3.2.2 Methods

3.2.2.1 Subjects

The study population consists of 4 male subjects, 2 asthmatic and 2 non-asthmatic. Asthmatic subjects were recruited from Glenfield Hospital in Leicester, UK. Asthma was defined by a clinician diagnosis with one or more of the following objective criterion (i) bronchodilator reversibility of FEV₁ to 400 mcg inhaled salbutamol of $\geq 12\%$ and 200 mls, (ii) Metahcholine PC₂₀ ≤ 8 mg /ml or (iii) peak flow variation of $\geq 20\%$ over a 2 week period. Asthma severity was classified according to the current Global Initiative for Asthma (GINA) treatment steps (143). Non-asthmatic subjects were recruited via local advertising, having normal airway physiology and no features of respiratory disease. Subjects with asthma had been free from exacerbations for at least 6 weeks prior to study entry. All subjects were non-current smokers, however due to the known association of smoking and small airways disease, pack year smoking exposure was not an exclusion criterion.

3.2.2.2 Ethical Approval

The study protocol was approved by the National Research Ethics Committee – East Midlands Leicester (approval number 08/H0406/189), and all subjects gave their written informed consent.

3.2.2.3 Visits

Clinical and physiological assessment was performed in the following sequence and over 1-2 study visits, no more than 1 week apart.

3.2.2.4 Lung Function Measurements

All lung function tests were performed 15 minutes after the administration of 400 mcg of inhaled salbutamol, administered via metered dose inhaler and spacer, with each 100 microgram actuation being inhaled in a separate inhalation to TLC, followed by a 5 to 10 second breath-hold. Spirometry was performed according to ATS/ERS standards (109).

Multiple breath washout (MBW) was performed according to current guidelines (20) by using the sulphur hexafluoride (SF₆) wash-in method as previously described (135). SF₆ was chosen as the inert tracer gas because of its heavy molar mass and based on previous simulation data from Dutrieue et al (288) suggesting that phase III slope

sensitivity to SF₆ is maximal at the level of the alveolar duct. Lung clearance indices, S_{cond} and S_{acin}, were calculated by using custom software written with TestPoint (Measurement Computing Corp, Norton, Mass) as previously described (109,269).

3.2.2.5 CT Imaging and Image Analysis

Figure 3.2-8 provides a detailed birds-eye view of the workflow taking CT input through to TDA output.

Volumetric whole-lung scans were obtained following administration of 400 mcg of inhaled Salbutamol at FRC and TLC in subjects lying supine.

PRM was performed using Imbio's Lung Density Analysis (LDA™) software application (Imbio, LLC, Minneapolis, MN) for all CT data, with registrations performed from TLC to FRC. Details on the PRM analysis have been previously reported (32,190,274,286).

Voxel classifications PRM^{Norm}, PRM^{fSAD} and PRM^{Emph} were calculated according to definitions published by Galbán et al (32). PRM^{Uncl} was defined as the associated compliment. Voxel feature ΔHU was defined as: (expiration HU) – (inspiration HU), representing local density change about the voxel, associated to volume change.

Trachea and upper airway centreline extraction was performed using Materialise Mimics software. Virtual modelling of the complete airway tree was achieved using Lung Chaste software and previously described methods (37). Resulting sets of distal branch points were assigned Strahler order (SO) values and overlapped with PRM registered voxel sets through manual visual alignment, following mean centralisation and matching ranges and rotational orientation of point sets.

Voxel SO proximity features SO_1, SO_5 and SO_10 were calculated using a k-nearest neighbour (k-NN) approach (see **Figure 3.2-9**), with k = 1, 5 and 10 respectively, providing measures of voxel proximity to the smaller airways. Features were calculated using bespoke scripts written with Matlab (Matlab Release 2015a, The MathWorks, Inc., Natick, Massachusetts, United States).

3.2.2.6 Topological Data Analysis

A set of 12 voxel based features were uploaded in .csv files, for each subject, to Ayasdi servers for analysis. The 12 features were Cartesian coordinates (X, Y and Z),

inspiration HU (iHU), change in HU (Δ HU), SO proximity (SO_1, SO_5 and SO_10) and PRM classifications (PRM^{Norm} , PRM^{fSAD} , PRM^{Emph} and PRM^{Uncl}). See **Table 3.2-2**.

Initial networks were generated using all voxels and column set X, Y, Z and iHU, with the neighbourhood lens functions, with parameters resolution = 80 and gain = 2.5. The histogram for Y was then utilized to split the population into predominantly left and right lung voxel sets, using the software's histogram interface (splitting approximately on the median Y value). The same process (neighbourhood lens functions with resolution = 80 and gain = 2.5) was applied to these sets, creating lung specific networks.

High density (iHU) voxel regions in lung specific networks were selected manually using the Ayasdi software lasso tool, and labelled. These high iHU node groups were then extracted using bespoke Python (v 2.7.1) script that wrote identified voxel coordinates (X,Y and Z) to a .txt file, that was read and plotted in 10 voxel thick slices in the transverse (see **Figure 3.2-1** top right section) and coronal (see **Figure 3.2-7 D**) planes using Matlab.

SO_1, SO_5 and SO_10 showed negligible differences in heat mapping, so SO_1 was used for simplicity, to determine proximity of network nodes to the smaller airways. Nodes with mean SO_1 below the histogram median were selected and made transparent in an exported image of the network. Images with transparent nodes were grey scaled and overlaid onto +40% contrast enhanced images of networks with heat maps of X, Y, Z, Δ HU or PRM^{fSAD} to produce multi-panel figures conveying this information concisely from the underlying high dimensional data.

3.2.3 Results

3.2.3.1 Clinical characteristics/Network Visualisations

Clinical characteristics of the subjects (n=4) are provided in **Table 3.2-1**. TDA networks for each subject are presented in **Figures 3.2-2** to **3.2-5**. Subject age, asthmatic/non-asthmatic class, and MBW based VH markers LCI, S_{cond} and S_{acin} , are provided at the top of each figure. In each case there are two panels of networks (lung specific as annotated), the top rows presenting spatial coordinate distribution, X, Y and Z for anterior-posterior, left-right and inferior-superior localisation respectively, and the bottom rows presenting Δ HU, as a signal of ventilation, on the left, and PRM^{fSAD} , as a

signal of disease, on the right. Colour map histograms with minimum and maximum value are included below each network. Greyed out areas represent voxels with high SO score. **Figure 3.2-6** illustrates reading networks with respect to spatial axes used.

3.2.3.2 Δ HU inferior-superior gradient

In non-asthmatic cases (**Figures 3.2-2** and **3.2-3**), Δ HU high areas appear spatially correlated with mid-to-low Z, in agreement with expected physiology (91); higher ventilation focused in the inferior regions of the lung, in the supine subject. In contrast, in the asthmatic cases (**Figures 3.2-4** and **3.2-5**), there appear high Δ HU areas overlapping mid-to-high Z, indicating biologically abnormal ventilation gradient in these lungs (ventilation distributed more toward apices). Solid ellipses highlight this observation.

3.2.3.3 PRM^{fSAD} anterior-posterior correlation

There is a clear and negative spatial correlation of PRM^{fSAD} and X in non-asthmatic subjects, with the associated networks being outlined by a dashed ellipse. In contrast, in the asthmatic cases, the correlation appears weaker, with contradictions illustrated by dashed sub-ellipses. The functional small airways disease marker PRM^{fSAD} thus appears restricted to the anterior region of the lungs in non-asthmatic subjects, whilst in the asthmatic subjects it spreads into the posterior regions.

3.2.3.4 Δ HU clustering pattern

When looking just at the Δ HU sub-networks, we observe a higher level of heterogeneity in the Δ HU distribution for asthmatic subjects, especially in the right lung of **Figure 3.2-4**; in comparison the Δ HU distribution of the non-asthmatic lungs appears less fragmented. This observation appears in agreement with the relationship of positive correlation between S_{acin} and VH, that is the subject with considerably highest S_{acin} has a markedly more heterogeneous distribution of ventilation (approximated by Δ HU).

3.2.3.5 Tissue Separation / Segmentation

In addition to small airway visualisation, we observed that anatomically driven network generation, using inspiration HU as a characteristic of anatomical matter, presented networks with a smooth gradient of tissue density (**Figure 3.2-7**). Investigating the high density low voxel count regions of the networks through plotting back in the physical space (**Figure 3.2-7 D**), it was clear this process was innately separating out airways and vascular structures.

3.2.4 Discussion

In this study we have developed a novel approach to visualise the small airways in the adult lung in order to evaluate patterns of ventilation and disease. This approach was then utilised to investigate differences between selected subjects from asthmatic and healthy control populations, and identified discriminatory patterns present. This is the first demonstration of applying TDA directly to high volume PRM registered lung voxel sets, and utilising extended airway centreline modelling to assign an SO proximity measure to lung voxels.

Previous studies have used TDA for disease phenotyping using clinical features and comorbidities to identify novel subgroups in type-2 diabetes (50), airway pathological and gene expression features identifying replicated micro clusters in asthma (261), and particles in exhaled air identifying phenotypes in small airways disease (52). Our study applies TDA more directly through medical imaging, to the organ characterising a disease, the human lung and asthma, to compress a large number of data points (order 10^7) and many features with information from disparate length scales, creating powerful visual summaries for qualitative analysis.

The subjects selected in this study were chosen to represent i) age-matched (with respect to asthmatic subjects) non-asthmatic lungs (**Figure 3.2-2**), ii) young and relatively healthy lungs (**Figure 3.2-3**), iii) more severe asthmatic lungs with higher (GINA 4) and MBW/IOS results (**Figure 3.2-4**), and iv) less severe (GINA 3) asthmatic lungs (**Figure 3.2-5**). Subjects were sex matched (male). The older non-asthmatic subject had a notably higher BMI relative to the other subjects, and was the only non-atopic present; this subject also had an LCI higher than the GINA 3 asthmatic, though their S_{acin} was lower. S_{cond} was only raised in the GINA 4 subject.

Our observations present clear hypothesis for testing in larger populations. The ΔHU inferior-superior gradient reversal has already been observed by our group in a larger ($n=52$) population of subjects, using quantitative measurements of ΔHU inferior-superior gradient (198). The loss of negative PRM^{fSAD} correlation in the anterior-posterior axis may also be tested with similar quantitative measures. The nature of ΔHU clustering in space, when compared with VH measures, from MBW/IOS for example, may also be tested, provided a suitable measure of ‘cluster count’ in an image such as these networks, or with clustering methods applied to ΔHU and spatial coordinate data.

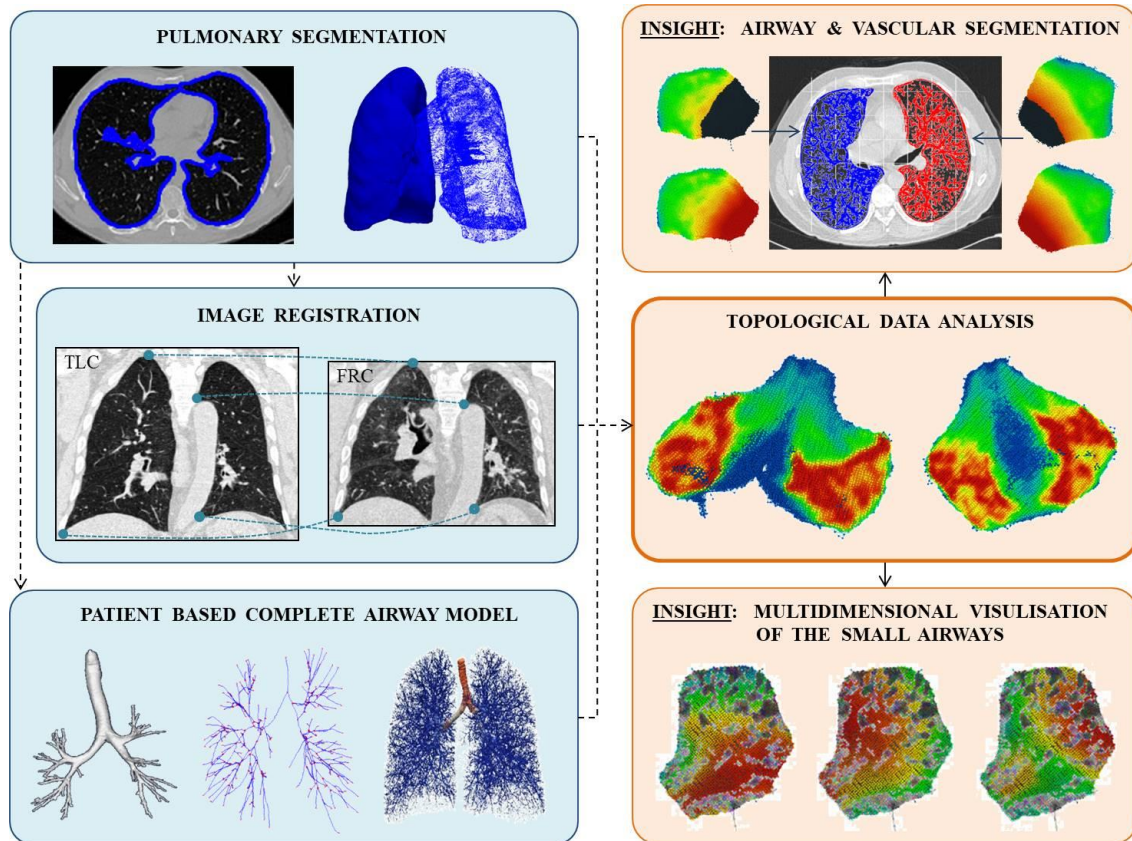
The anatomically driven approach to network generation, inputting tissue density alongside spatial coordinates as training features, naturally segmented high from low density anatomy in space, as illustrated in **Figure 3.2-7**. The bulk of the imaging data is relatively low density, due to gas abundance in inspiration or TLC state, and appears separated from voxels containing more of the higher density structural and viscous tissues, like airway walls, vascular walls, blood and pleura. It suggests potential for application as a segmentation tool with imaging data.

There are a number of important limitations to the work presented. The subjects selected were the same sex (male), and so did not account for significant biological differences observed between the sexes (298), and technically limit presented results to male lung anatomy; future studies should include female subjects to account for this. The number of subjects were limited, to focus on an in depth qualitative analysis of a novel development, rendering our observations as hypothesis that really require future study in larger populations to establish with statistical strength.

In conclusion we have developed and applied a powerful exploratory tool for high dimensional respiratory organ visualisation and analysis, given a set of voxels segmented from CT with associated measures such as HU. Summary figures, such as those presented in **Figures 3.2-2 to 3.2-5**, may be read by both researchers and clinicians, to gain insight into the spatial distributions of features, such as ventilation and disease markers, within the lungs. The methodology presented here is built upon a number of underlying processes, such as extended centre line modelling (Chaste), and has great potential for further research and development.

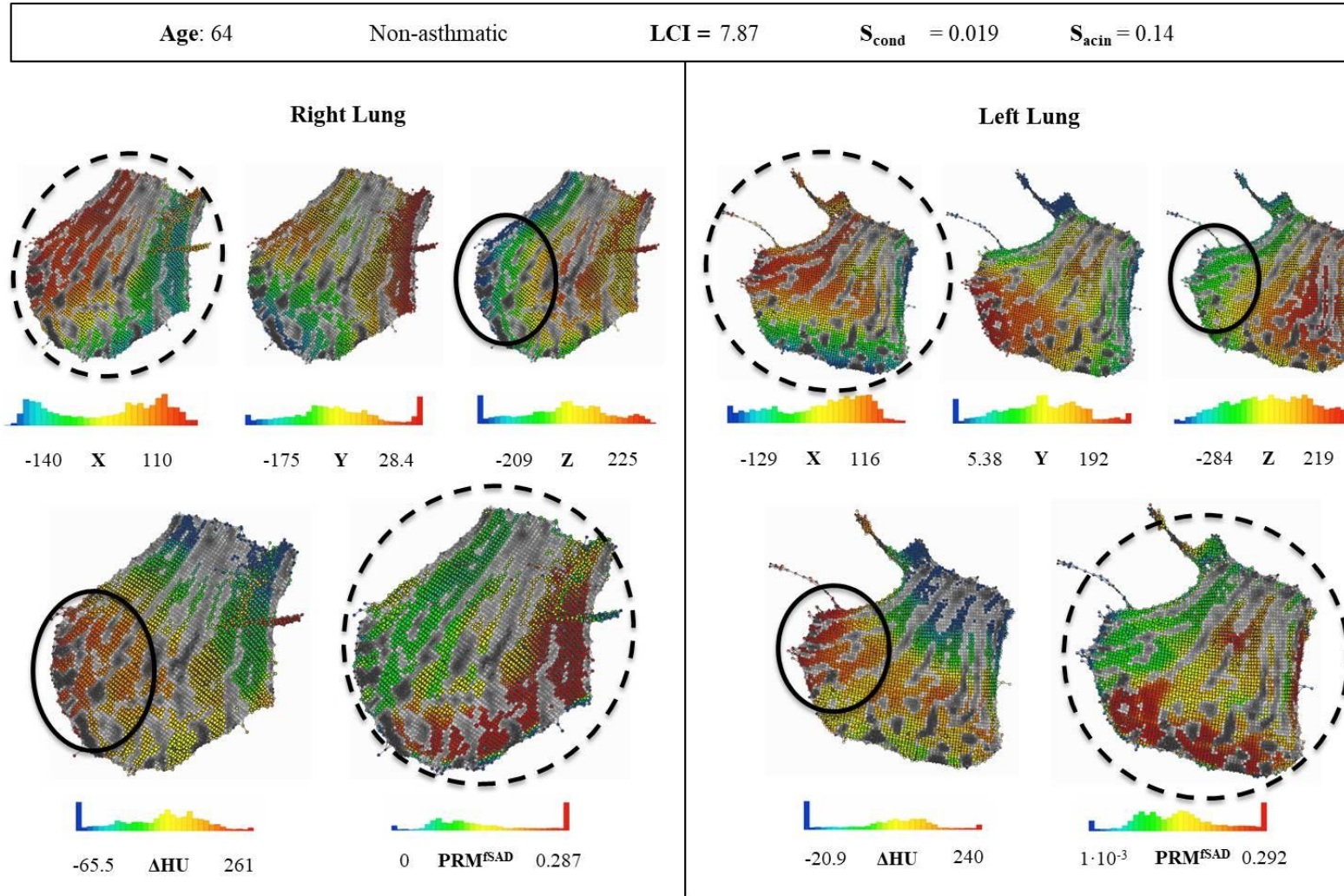
3.2.5 Figures and tables

Figure 3.2-1: Overview of TDA image analysis workflow developed in this study.



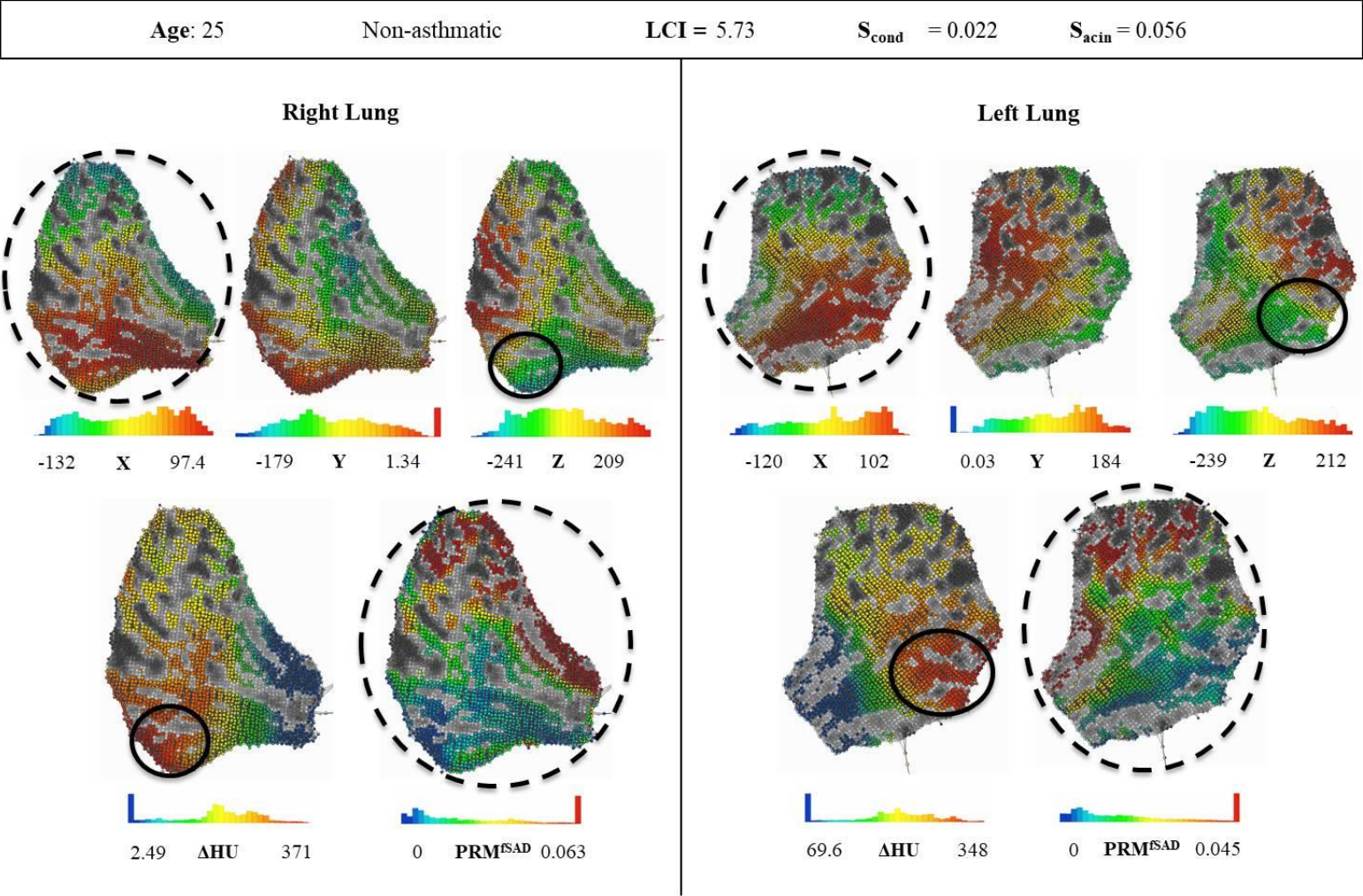
Overview of TDA image analysis workflow developed in this study, from data sources and pre-processing (left) to visualisations (right). Lung volumes are segmented from CT images at inspiration and expiration, then both are used for image registration, computing geometric deformation map between TLC and FRC states; TLC image is also used as the basis for major airway segmentation, from which a 16 generation airway tree model is simulated. Then, moving through to the right panel set, collated data sets are processed using TDA, resulting in compressed 2D representations of the lungs. Two outputs illustrated in this study are small airway focussed heat maps and innate separation of higher density tissues.

Figure 3.2-2: TDA network for older non-asthmatic subject (low VH).



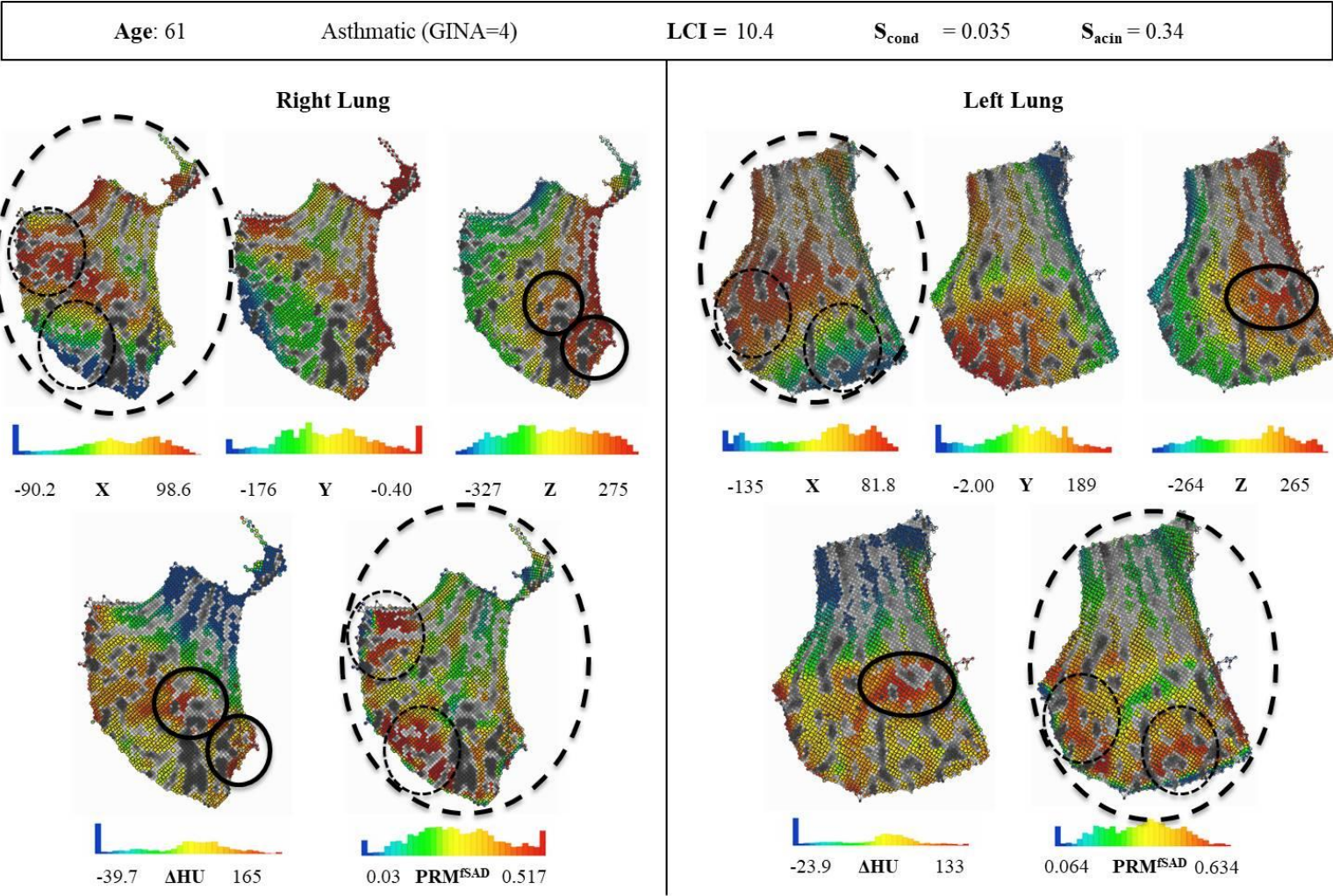
TDA network heat map panel for older non-asthmatic subject, age-matched to asthmatic subjects. Networks generated using parameters: gain = 2.5, resolution = 80, and using neighbourhood lens functions, 20-nearest neighbours based metrics. Biologically normal association of high ΔHU with inferior regions indicated with solid ellipses. PRM^{fSAD} negative spatial correlation to X indicated with dashed ellipses.

Figure 3.2-3: TDA network for younger non-asthmatic subject (minimal VH).



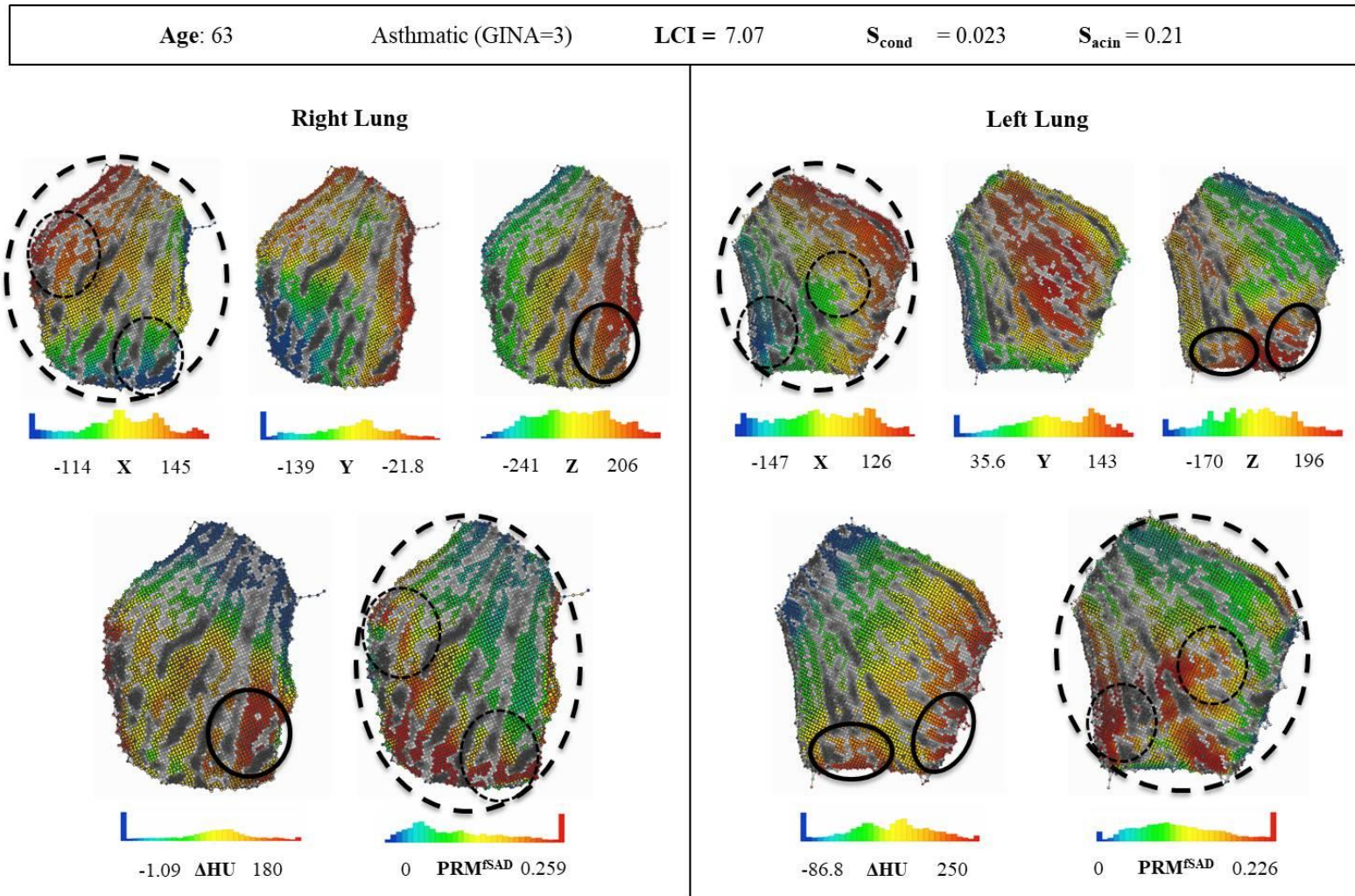
TDA network heat map panel for younger non-asthmatic subject, lowest LCI (5.73) and S_{acin} (0.022). Networks generated using parameters: gain = 2.5, resolution = 80, and using neighbourhood lens functions, 20-nearest neighbours based metrics. Biologically normal association of high ΔHU with inferior regions indicated with solid ellipses. PRM^{fSAD} negative spatial correlation to X indicated with dashed ellipses.

Figure 3.2-4: TDA network for severe asthmatic subject (GINA 4, high VH).



TDA network heat map panel for higher GINA step (4) asthmatic subject, highest LCI (10.4) and S_{acin} (0.34). Networks generated using parameters: gain = 2.5, resolution = 80, and using neighbourhood lens functions, 20-nearest neighbours based metrics. Biologically abnormal association of high ΔHU with superior regions indicated with solid ellipses. Contradictions to PRM^{fSAD} negative spatial correlation to X indicated with dashed ellipses.

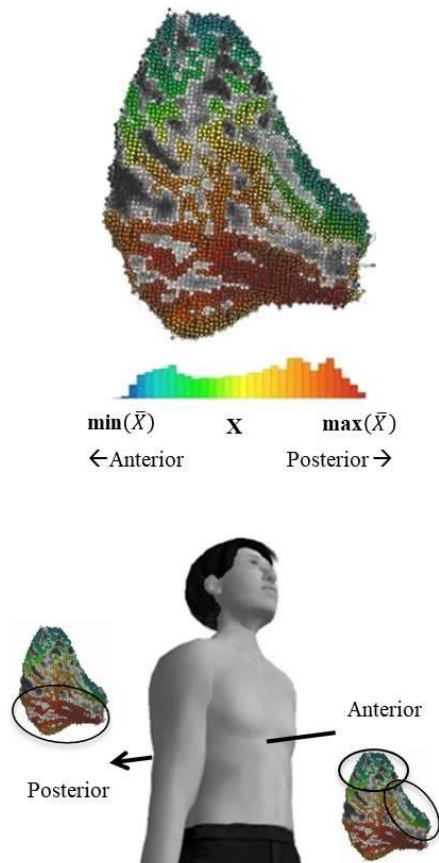
Figure 3.2-5: TDA network for mild asthmatic subject (GINA 3, mild VH).



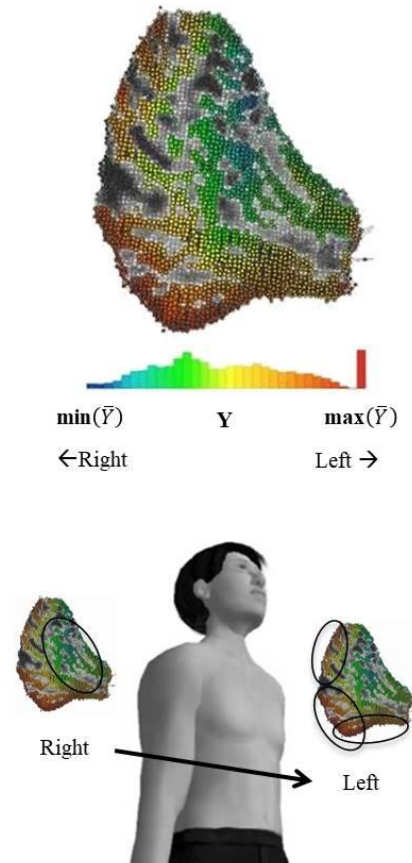
TDA network heat map panel for mild asthmatic subject (GINA = 3). Networks generated using parameters: gain = 2.5, resolution = 80, and using neighbourhood lens functions, 20-nearest neighbours based metrics. Biologically abnormal association of high ΔHU with superior regions indicated with solid ellipses. Contradictions to PRM^{fSAD} negative spatial correlation to X indicated with dashed ellipses.

Figure 3.2-6: Interpreting network heat maps for spatial features.

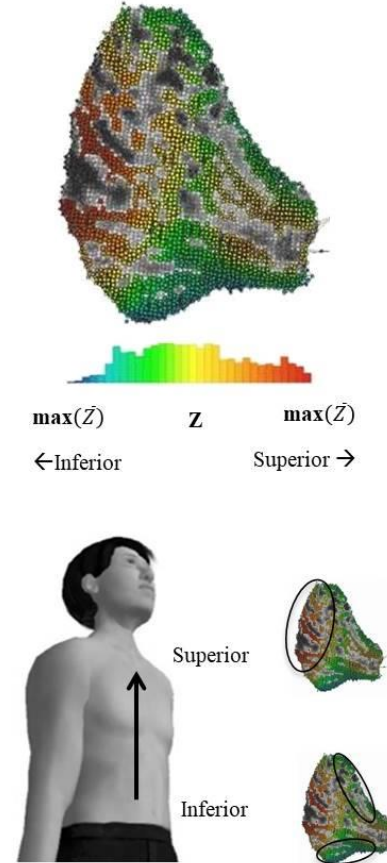
a) 'X' co-ordinate value.



b) 'Y' co-ordinate value.

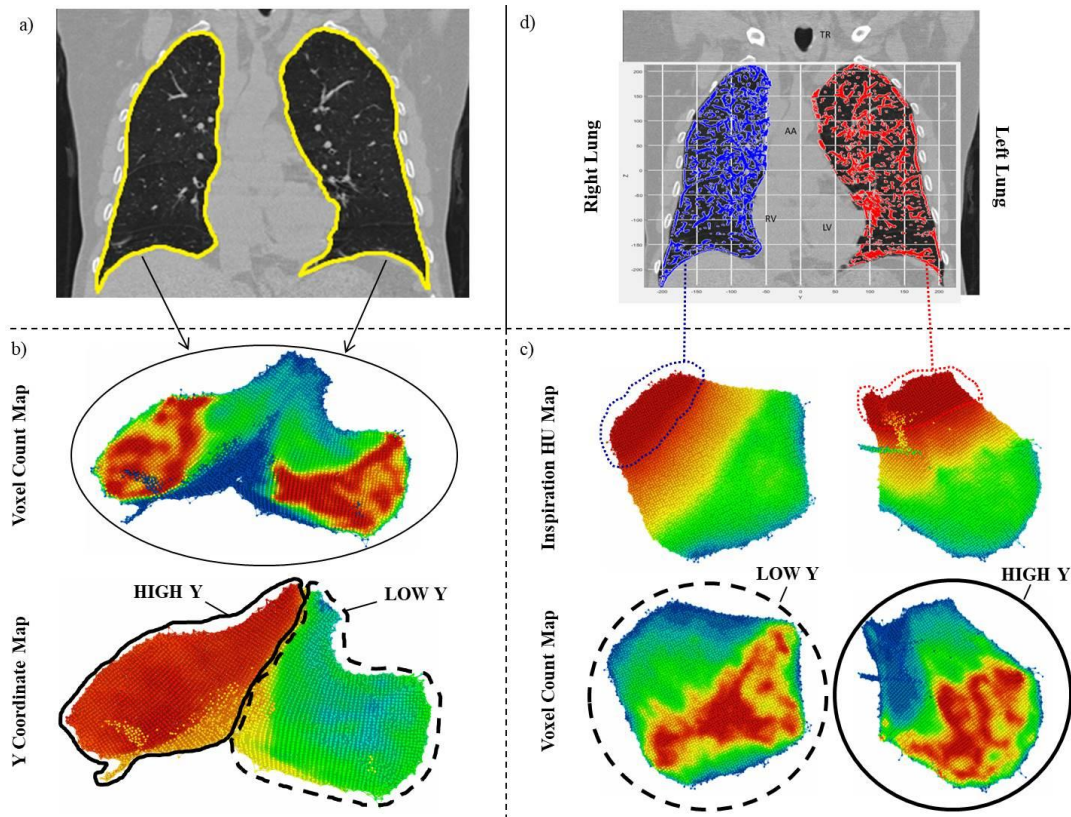


c) 'Z' co-ordinate value.



Interpreting network heat maps for spatial features. A, the 'X' feature, anterior-to-posterior voxel coordinate. Cool (blues and greens) nodes are more anterior, hot (reds and yellows) nodes are more posterior. B, the 'Y' feature, left-to-right voxel coordinate. Cool nodes are more to the left of the given lung, hot nodes are more to the right. A, the 'Z' feature, inferior-to-superior voxel coordinate. Cool nodes are inferior, hot nodes are more superior. In each case histograms illustrate distribution of node means, with minimum and maximum occurring averages listed for reference (indicates magnitude that the colour scale represents).

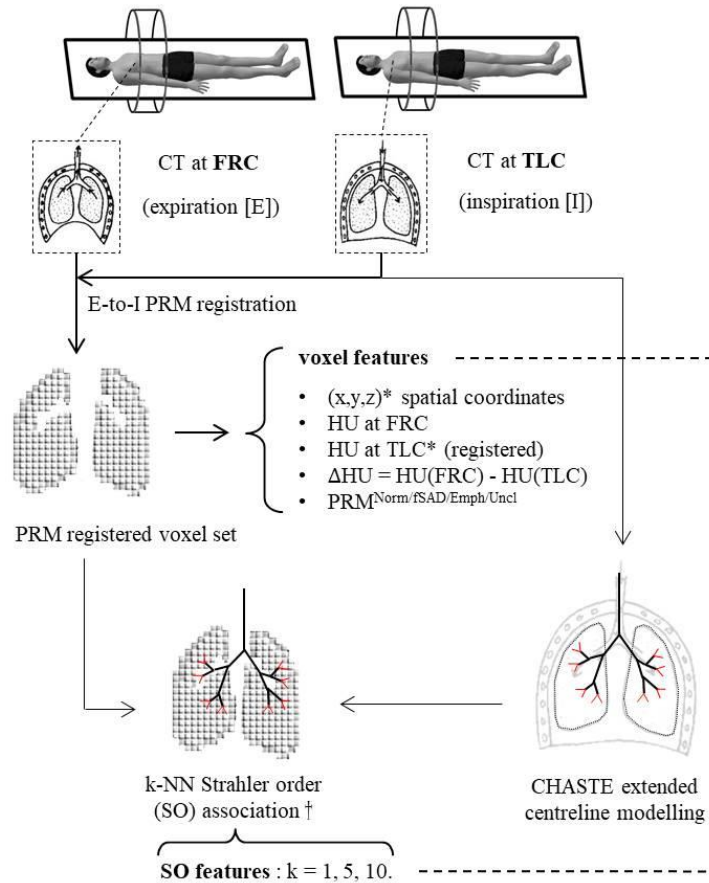
Figure 3.2-7: Anatomical segmentation using TDA.



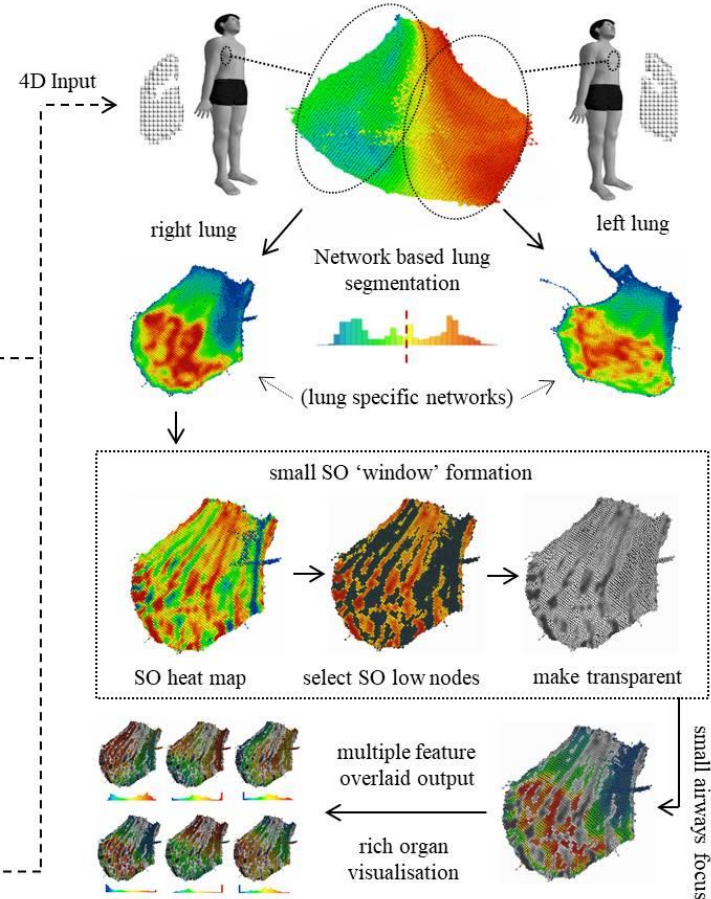
Anatomical segmentation using TDA. A) entire lungs are segmented, minus major airways, and stored as a set of voxels with spatial coordinates (X,Y,Z) and Hounsfield unit (HU) densities. B) TDA generates a network based on spatial and anatomical features, Y coordinate heat map confirms apparent splitting of left and right lungs. C) TDA iterated on left and right lung subsets generating lung specific networks. Inspiration HU heat map illustrates smooth transition across network with high density (red) relatively low voxel count nodes grouped to one side. D) Plotting spatial coordinates and overlaying original CT image appears to confirm left-right lung segmentation with vascular and pleural features (higher density tissues).

Figure 3.2-8: High level methodological overview of TDA based organ visualization workflow.

a) Parametric Response Mapping (PRM) based feature extraction

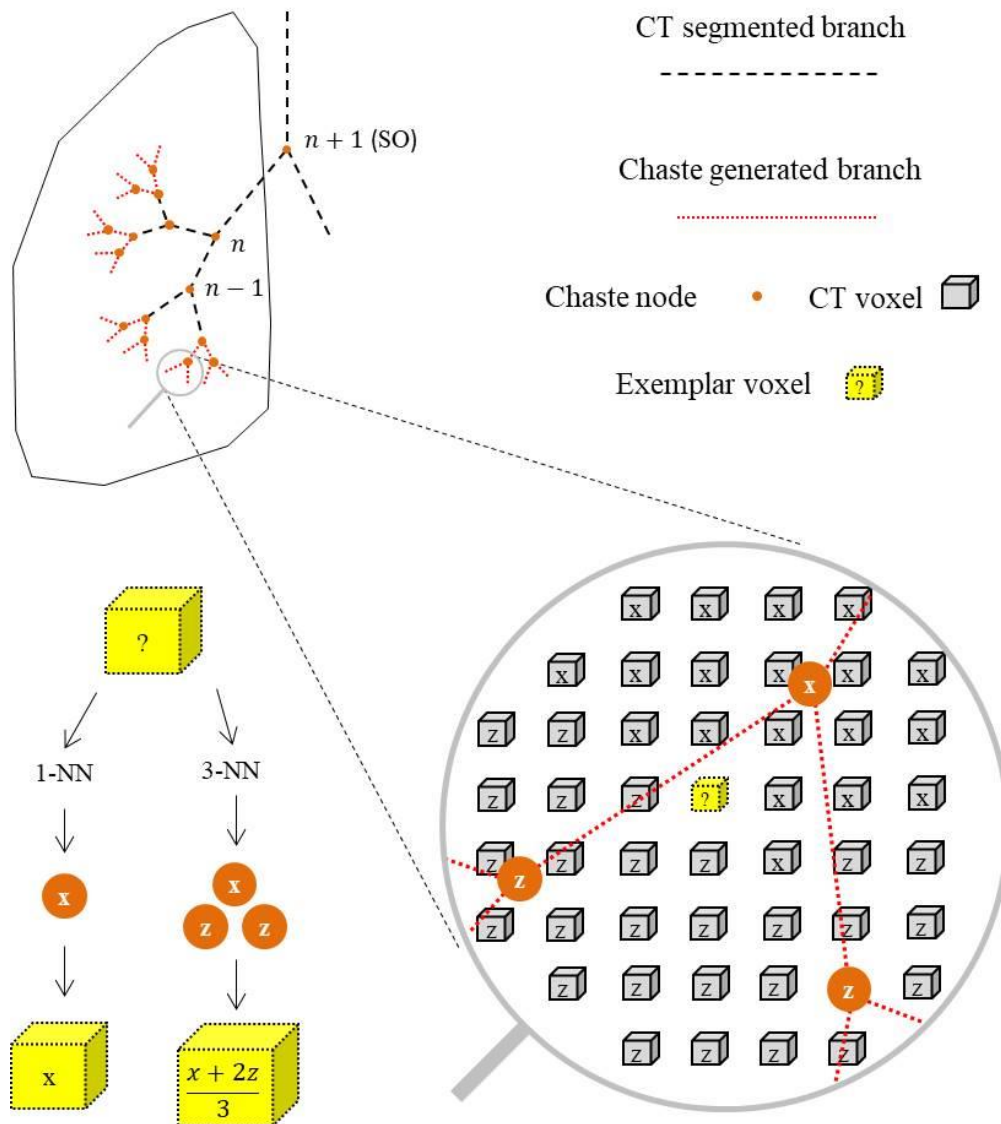


b) Topological Data Analysis (TDA) creates compressed representation



High level methodological overview of TDA workflow. A, obtaining data matrix for input to TDA. CT scanning undertaken at FRC and TLC, and registered to provide high volume data set of lung voxels. (*) spatial coordinates and TLC Hounsfield unit chosen for 4D input to TDA. Chaste modelling of extended centrelines were aligned to voxel sets to obtain Strahler order (SO) voxel features. B, 20-NN based TDA driven by (*), produces high volume network of nodes. Lungs are segmented using Y coordinate histogram, leading to lung specific networks. SO heat map identifies 'small airways' to be made transparent, and resultant greyscale image overlaid onto other feature heat maps results in small airways focussed compressed visualization of the lung.

Figure 3.2-9: Strahler order (SO) score assignment method with k-NN.



Auxiliary method to facilitate Strahler order (SO) incorporation and high quantity TDA network analysis summarisation. A simple k nearest neighbours (k-NN) algorithm is implemented to associate voxels with SO proximity within the lungs. The magnified region illustrates Chaste based nodes with associated SO values x and z (N.B. as depicted $x=z+1$). The cubes represent voxels from the PRM registered voxel set, grey cubes annotated with expected assigned value when $k=1$. Yellow cube demonstrates value assignment ideology and dependence on k .

Table 3.2-1: Clinical characteristics for selected asthmatic (n=2) and non-asthmatic (n=2) subjects.

	Non-Asthmatic		Asthmatic	
Figure No.	3.2-2	3.2-3	3.2-4	3.2-5
Age	64	25	61	63
GINA step	-	-	4	3
Weight (kg)	118	95	87	80
Height (cm)	180	185	177	178
BMI	36.42	27.76	27.77	25.25
Atopy	No	Yes	Yes	Yes
FEV₁%	146.0	120.0	47.9	67.0
FEV₁/FVC (%)	80.00	87.00	31.75	65.00
Pre-BD FEV₁(L)	4.80	5.52	1.25	1.75
Post-BD FEV₁(L)	4.95	5.67	1.60	2.00
%BD\pm	3.13	2.72	28.0	14.3
LCI	7.9	5.7	10.4	7.1
S_{cond}	0.02	0.02	0.04	0.02
S_{acin}	0.14	0.06	0.34	0.21

GINA, global initiative for asthma; BMI, body mass index; FEV, forced expiratory volume; FVC, forced vital capacity; BD, bronchodilator; LCI, lung clearance index; S_{cond}, conductive ventilation heterogeneity; S_{acin}, acinar ventilation heterogeneity. Data expressed as recorded values to apparent number of significant figures.

Table 3.2-2: Complete list of 12 features selected for upload to Ayasdi server.

Feature	Description
X	X-coordinate for voxel, anterior-to-posterior directed axis.
Y	Y-coordinate for voxel, left-to-right (lateral) directed axis.
Z	Z-coordinate for voxel, inferior-to-superior directed axis.
iHU	Inspiration (TLC) HU of voxel, chosen to incorporate ‘anatomical information’.
Δ HU	Δ HU = Expiration HU – Inspiration HU of voxel, to incorporate ‘ functional information’.
SO_1	k-NN SO association, with k=1.
SO_5	k-NN SO association, with k=5.
SO_10	k-NN SO association, with k=10.
PRM^{Norm}	1 if voxel is classified PRM^{Norm} , 0 otherwise. $PRM^{Norm} \leftrightarrow iHU \geq -950 \text{ AND } eHU \geq -856$
PRM^{fSAD}	1 if voxel is classified PRM^{fSAD} , 0 otherwise. $PRM^{fSAD} \leftrightarrow iHU \geq -950 \text{ AND } eHU < -856$
PRM^{Emph}	1 if voxel is classified PRM^{Emph} , 0 otherwise. $PRM^{Emph} \leftrightarrow iHU < -950 \text{ AND } eHU < -856$
PRM^{Uncl}	1 if voxel is classified PRM^{Uncl} , 0 otherwise. $PRM^{Uncl} \leftrightarrow iHU < -950 \text{ AND } eHU \geq -856$

HU, Hounsfield Unit; eHU, expiration (FRC) HU; iHU, inspiration (TLC) HU; SO, Strahler order; k-NN, k nearest-neighbours; PRM, parametric response map; fSAD, functional small airways disease. N.B. first four listed features (bold type) were used to train networks; spatial and anatomical information drove network creation.

3.3 Functional CT image analysis supports computational modelling of small airway resistance in asthma

Abstract

Background: Asthma is a disease characterised by ventilation heterogeneity (VH). The forced oscillation technique (FOT) provides a measure of VH in the small airways, R5-R20. Airway and lobar segmentation in CT imaging has been used to create models of the airway tree for computer simulation of FOT. Parametric response map (PRM) has been developed to extract functional information from paired TLC-FRC CT series.

Aims: In this study we aimed to improve the agreement of measured and simulated small airways resistance, by integrating PRM registered CT data with airway tree models, to parameterise small airway compliance.

Methods: Twenty three subjects (14 asthmatic and 9 healthy) were recruited as part of an observational study at Glenfield Hospital (Leicester, UK) and underwent CT scanning at total lung capacity (TLC) and functional residual capacity (FRC), and impulse oscillometry. Airways and lung surfaces at TLC were segmented using Mimics, and processed with cancer heart and soft tissue environment (Chaste) software to produce extended airway centrelines of the first 16 airway generations. Lung volumes at both TLC and FRC were segmented, and registered using parametric response map (PRM) software written in Matlab. Change in Hounsfield units (Δ HU) between FRC and TLC was calculated for each voxel. The airway tree and voxel set were spatially aligned to calculate a Δ HU score for each branch. Airway radii were constricted relative to assigned Δ HU in FOT computational modelling.

Results: Constriction of airways, based on Δ HU, increased measured and simulated R5-R20 correlation. The distribution of Δ HU about terminal airway branches was significantly right shifted above zero.

Conclusion: We have developed methods to combine PRM and airway tree modelling to visualise lung function in the airway tree from CT imaging data, and utilized this approach to improve FOT modelling of small airway VH.

3.3.1 Introduction

Asthma is a disease characterised by ventilation heterogeneity (VH) (16,275), which has been demonstrated to be strongly associated with the defining characteristic of airway hyper responsiveness (18), and predictive of inhaled corticosteroid treatment response (17). VH may be visualised and quantified using imaging approaches such as hyper polarised 3-Helium magnetic resonance imaging (MRI) (276,277), Oxygen enhanced MR (27) and single photon emission computed tomography (SPECT-CT) (33). VH can be measured clinically at the mouth using tidal breathing techniques such as multiple breath washout (MBW) (279,280) and impulse oscillometry (IOS) (23), a type of forced oscillation technique (FOT) (133). Guidelines for quality control of tidal breathing markers of VH derived from MBW and IOS have been proposed (20,299).

In IOS impulses generated by a speaker are superimposed onto tidal breathing and travel through a subject's airways to an extent inversely proportional to wave frequency (lower frequencies travel further); 5Hz has commonly been used to probe the whole airway tree, and 20Hz for the upper airways only (22,300). Respiratory resistance at x Hz may then be measured (R_x), and so given R_5 measures resistance due to the entire respiratory tract, and R_{20} only the upper sections including larynx, trachea and major airways, the difference R_5-R_{20} is used as a measure of peripheral airway resistance and heterogeneity (135), and has recently been validated as a small airways detection tool in adult asthma (24).

FOT modelling has been developed to computationally simulate IOS measurements (37). Major airways and lung volumes are segmented from CT imaging data and used to create a patient specific extended airway centreline model (201), a process developed by the AirPROM (Airway Disease Predicting Outcomes through Patient Specific Computational Modelling) consortium and implemented within Cancer heart and soft tissue environment (Chaste), developed by Oxford computer science (202). Thus small airways resistance and VH marker R_5-R_{20} may be simulated; however the concordance with clinically measured R_5-R_{20} is not well established.

The parametric response map (PRM) algorithm has been developed and applied to longitudinal CT imaging data to quantify changes in volumes as they deform geometrically in time (31). This has been applied to lung volumes at total lung capacity (TLC) and functional residual capacity (FRC) to extract functional information from

paired CT series (32). This has recently been utilised to link gradients of ventilation, surrogated by Hounsfield unit (HU) change ΔHU , across the lung to R5-R20 (198).

In this study we sought to combine PRM registered lung voxels with spatially aligned extended centreline airway trees, in order to improve the accuracy of FOT modelling of R5-R20, with regional compliance parametrisation provided by ΔHU measured about the terminal branch points. See **Figure 3.3-1** for overview.

Specifically we aimed to:

- Spatially align expiration-to-inspiration (E2I) PRM registered voxels with inspiration based airway tree models.
- Calculate and assign a ΔHU score to every branch of the airway tree model.
- Manipulate airway branch radii in direct proportion to ΔHU , to observe the effect on clinically measured and computationally simulated R5-R20 correlation.

We hypothesised that adjustment airway radii using ΔHU would improve the accuracy of R5-R20 simulation, as this provides greater inclusion of regional compliance (deformation extent as indicated by ΔHU) in the computational model.

3.3.2 Methods

3.3.2.1 Subjects

A total of 23 subjects are presented in this study. All subjects were non-current smokers and had a smoking exposure of equal to or less than 10 pack years; 9 past smokers with an average pack year of 4.4 (1 d. p.). Asthmatic subjects (n=14) were recruited from Glenfield Hospital in Leicester, UK. Asthma was defined by a clinician diagnosis given at least one of the following objective criterion (i) bronchodilator reversibility of FEV₁ to 400 mcg inhaled salbutamol of $\geq 12\%$ and 200 mls, (ii) Metahcholine PC₂₀ ≤ 8 mg/ml or (iii) peak flow variation of $\geq 20\%$ over a 2 week period. Asthma severity was classified according to the current Global Initiative for Asthma (GINA) treatment steps (143). All asthmatics experienced no exacerbations within 6 weeks prior to study entry. Non-asthmatic volunteers (n=9) were recruited via local advertising, presenting normal airway physiology and no features of respiratory disease.

3.3.2.2 Ethical Approval

The study protocol was approved by the National Research Ethics Committee – East Midlands Leicester (approval number 08/H0406/189), and all subjects gave their written informed consent.

3.3.2.3 Visits

Clinical and physiological assessment was performed in the following sequence and over 1-2 study visits, no more than 1 week apart.

3.3.2.4 Lung Function Measurements

All lung function tests were performed following the administration of 400 mcg of inhaled salbutamol. Spirometry was performed according to ATS/ERS standards (269). Impulse oscillometry (IOS) was performed in triplicate as previously reported and in accordance with international guidelines (133). Carbon monoxide uptake in the lung was determined using the single-breath method, according to standard guidelines (266). The carbon monoxide transfer coefficient (KCO) was calculated,

3.3.2.5 Statistical Analysis

Feature normality was tested using the one-sample Kolmogorov-Smirnov test at the 5% significance level. Mean difference significance for normal features between two

groups was performed using a two-sample t test at the 5% significance level, and for non-normal features using a Mann-Whitney U-test at the 5% significance level. This workflow was applied to asthmatic (n=14) and non-asthmatic (n=9) groups to test difference of means for all features reported in **Table 3.3-1** and **3.3-2**.

3.3.2.6 CT Image Acquisition and Segmentation

Volumetric whole-lung scans were obtained following administration of 400 mcg of inhaled Salbutamol at FRC and TLC in subjects lying supine. Lung volumes were segmented using ‘yet another CT analyser’ (Yacta) software (301) (v2.8.5.11). Trachea and upper airway centreline extraction was performed using Materialise Mimics software. Virtual modelling of complete airway tree was achieved using Lung Chaste software and previously described methodology (37) applied to TLC images.

3.3.2.7 Image Registration

Expiration (FRC) to Inspiration (TLC) image registration (E2I) was performed using cmi (Center for Molecular Imaging, University of Michigan, United States) Matlab (release 2015a, The MathWorks, Inc., Natick, Massachusetts, United States) scripts, incorporating elastix open source code. Voxel feature ΔHU was defined as: (expiration HU) – (inspiration HU), representing local density change about a voxel, associated closely with tissue volume difference due to TLC/FRC state change.

3.3.2.8 Set Alignment

Point sets of the segmented lung volumes (voxels with assigned ΔHU values) and simulated airway trees were aligned semi-manually: both sets were mean centred at 0 (subtraction of mean positional vector), then the airway tree set was rotated 90 degrees clockwise about the inferior-superior axis, and reflected in the sagittal plane. Finally the airway tree set was rescaled to match the range of the registered voxel set in all three Cartesian axes, with optional perturbation of scaling and translation. Visualisation of both sets as colour coded points in space was used to judge set overlap, and perturbation parameters edited until visual agreement appeared satisfactory (maximisation of apparent volume intersection).

3.3.2.9 ΔHU Assignment

To assign a ΔHU value to all branches of Strahler order 1 (SO1), terminal branches, of an airway tree, the distal nodes of said branches were defined as centre points for

spheres of radius $r = \frac{5\sqrt{3}}{2} + 0.1$; the mean of ΔHU for all voxels within a sphere was the value then assigned to that branch, see **Figure 3.3-2**. A cell array method, working on a lattice of partition length r , was applied to the E2I registered voxel set (302), for computational feasibility, and SO1 ΔHU values were assigned in sequence to every terminal branch of a tree. An iterative algorithm working up Strahler orders from 2 to the tree maximum was then employed, assigning to every other branch a ΔHU value equal to the average of the ΔHU of connected branches of one Strahler order less.

3.3.2.10 Branch ΔHU Visualisation

To visualise tree structures with assigned ΔHU , an algorithm was implemented to plot branches with Strahler order belonging to a given set, $\{5, 6 \dots 12\}$ presented here (**Figures 3.3-1 and 3.3-5**), with circle markers positioned at the distal branch nodes to display information on the assigned ΔHU value through circle size, with radii directly proportion to z score of branch ΔHU , and interior colour green (red) for above (below) average ΔHU , averages taken over all branches selected for plotting for the given subject.

3.3.2.11 FOT Modelling

A computational model of airway impedance was designed, based on previous models in the literature to provide simulations of IOS derived resistance values across a range of frequencies. A 1D wave equation was used to estimate the impedance of each branch (203), with total impedance being calculated through summation of parallel and series contributions (272). Each terminal bronchiole was subtended by a constant-phase viscoelastic model parameterised using data from the literature (207). The models were adjusted for potential confounding from upper airway shunting.

3.3.2.12 Radii Constriction

Airway radii of branches for a given Strahler order were rescaled as a function of ΔHU to observe the resulting change to FOT simulated results against those measured in clinic. Radii scaling factors were computed from the associated ΔHU distribution. ΔHU values outside of three standard deviations were mapped to the mean plus three standard deviations. Then for a given branch, the new radius (r_2) was calculated as:

$$r_2 = \left(1 + w \cdot \frac{\Delta\text{HU} - \max(\Delta\text{HU})}{\text{range}(\Delta\text{HU})}\right) \cdot r_1$$

Where r_1 is the initial radius, ΔHU statistics are calculated across all branches of the same Strahler order, and w is a weighting parameter between 0 and 1 inclusive. This produces scale factors in the interval $[1-w, 1]$, i.e. as w varies from 0 to 1, the extent to which radii are reduced, for greater ΔHU , varies from 0 to 100%. See **Figure 3.3-3** for visual demonstration.

3.3.2.13 FOT Parameterisation Testing

An algorithm was implemented to compute Pearson's correlation coefficient between measured and simulated resistance features (R5, R20 and R5-R20) for Strahler order ranges 1 to k , $k=1, 2, 4, 6, 8$ and 10, and weighting parameter $w=0$ and $w=0.2, 0.225, \dots, 0.6$. At each iteration, airway tree radii were rescaled according to the described scheme and FOT modelling run using the new radii distribution, from which a new set of simulated results were compared to the clinically measured resistance values.

3.3.3 Results

3.3.3.1 Clinical Characteristics

Asthma ($n=14$) / Non-asthma ($n=9$) group characteristics, across all subjects used in analysis ($n=23$), are presented in **Table 3.3-1**. Groups are well matched for age and smoking history. There is a notably higher ration of female subjects in the asthmatic group ($9/5 > 5/4$). Asthmatic subjects have a significantly higher proportion of atopy ($9/5 \gg 2/9$). Only 2 out of 14 asthmatic subjects are GINA stage 1 or 2, whilst stage 3 accounts for 50% of the subjects; the sample is most representative of moderate to severe asthma. In spirometry the only statistically significant ($p < 0.05$) difference occurs in post-bronchodilator FEV_1 , where asthmatic subjects have a lower average (2.65 L) relative to non-asthmatic subjects (3.74 L).

3.3.3.2 Airway Tree Modelling and ΔHU Assignment

Airway trees for all 23 subjects presented were generated with no major faults on visual inspection, and were aligned successfully to PRM E2I registered voxel sets. ΔHU about the terminal branch nodes in every subject was calculated. Asthma / non-asthma branch

radii and ΔHU statistics are presented in the first section of **Table 3.3-2**, and coronal (anterior perspective) heat maps across all 23 subjects for radii and ΔHU are presented in **Figures 3.3-8** and **3.3-9**. Radii average (\bar{r}) was indistinct between groups ($\Delta \bar{r} < 0.005\text{mm}$), however non-asthmatic subjects presented with notably less standard deviation ($0.007\text{mm} < 0.026\text{mm}$). ΔHU in asthmatics was higher on average, though without statistical significance (5%). ΔHU averages for both groups between global and SO1 voxel sets showed little difference, however a comparison across all subjects, illustrated in **Figure 3.3-4**, showed a marked focus toward $\Delta HU > 0$. Subject specific ΔHU distributions, global and SO1, are presented in **Figures 3.3-10** and **3.3-11**.

Visualisations of ΔHU across a range of Strahler orders were produced and provided a clear picture of functional information captured about the terminal branch points, providing basic QC/QA of the workflow through observation of expected ΔHU patterns replicated in the SO1 based visuals. **Figure 3.3-5** demonstrates this in 2 subjects showing distinct ΔHU patterns with respect to the inferior-superior direction. All branches with a Strahler order between 5 and 12 inclusive are plotted. Circles are plotted at terminal branch points with radii proportional to branch ΔHU and coloured green (red) is above (below) mean ΔHU for the given branches.

3.3.3.3 FOT Modelling and Radii Constriction

FOT modelling of resistance was conducted across all subjects with raw (unaltered) airway tree structures; Clinically measured and computationally simulated R5, R20 and R5-R20 is reported for asthma/non-asthma subjects in the second and third sections of **Table 3.3-2**. Asthmatic subjects had higher resistance values in clinical measurements, though without statistical significance (5%). In raw airway tree based simulations there was a significant ($p < 0.05$) difference in R5, asthmatics maintaining higher resistance in general, however simulated values were notably lower in both R5 and R20, and very poor replication in R5-R20.

Constriction was applied across all subjects, reducing branch radii across a specified range of Strahler orders, and to an extent determined by a variable weighting parameter w . Resulting correlation coefficients (Pearson) between measured and simulated resistance, for unaltered ($w=0$) and $w = 0.2, 0.225, \dots, 0.6$, and for Strahler order ranges from 1 to k , $k = 1, 2, 4, 8$ and 10, are presented in **Figure 3.3-6**. At $w=0$, R5-R20 has 0 correlation, whilst R20 and R5 have weak and mild positive correlation respectively.

Then with increasing constriction, R5-R20 correlation rises sharply, whilst R5 and R20 correlations decrease to a lesser extent. The point at which R5-R20 correlation supersedes R5 is indicated ($w=0.4$).

The case for R5-R20 with $w=0.4$, and Strahler order range = 1 to 10, is presented as scatter and Bland-Altman plots in **Figure 3.3-7**. The left plots illustrate the situation with no radii alteration, where there is close to 0 correlation between measured and simulated values, and a strong negative bias between the measurements. The right plots show the change due to radii constriction, where mild positive correlation is observed, with a lesser negative bias. The correlation coefficient is 0.00 (2 d. p.) prior to alteration and 0.43 (2 d. p.) post alteration. Simulated results for $w=0.4$ across all asthma / non-asthma subjects are presented in the final section of **Table 3.3-2**. R5-R20 average is significantly ($p < 0.05$) greater in asthmatic subjects, and constricted R5 and R5-R20 values are notably closer to the clinically measured values, relative to the unaltered simulation results.

3.3.4 Discussion

In this study we have developed a methodology to link functional information from registered inspiratory-expiratory CT scans, ΔHU , to airway tree models for effective visualisation and analysis of branch functionality; furthermore we utilised this technique to improve the ability of FOT modelling to simulate resistance in the small airways, measured by R5-R20, in adult asthma. Additionally we showed that voxel sets acquired about the terminal branch points present behaviour in agreement with gas release, $\Delta HU > 0$, relative to all voxels of the segmented lung volumes. We have shown that image registration provides access to functional information about the small airways, and that combining registered voxel sets with segmented airway trees allows for more effective modelling of small airway resistance.

This study is the first to combine and link CT imaging information between PRM image registrations and extended airway centreline models in patients, to simulate small airway resistance in asthma. Past studies have utilized extended airway centreline models to study airway hyper-responsiveness (303), multiple breath washout (MBW) (304) and ventilation distribution (305). A recent study employed FOT computational

modelling in extended airway models, with airway constriction (radii reduction) applied homogeneously and heterogeneously to varying degrees, to observe the effect on R5-R20 (38). This study concluded that the contribution of narrowing the small airways, relative to the conducting zone, was greatest in determination of R5-R20. The visualisation of gradient reversal, of ΔHU , in an asthmatic subject echoes observations made in another recent study (198), where computational modelling of R5-R20 in extended centreline structures was used to validate this effect, in that constriction applied in inferior lung regions demonstrated a marked dominance in R5-R20 determination.

The methodology presented in this study can be applied to measure small airway function and resistance given only paired inspiratory-expiratory CT scans for a subject. The visualisation algorithm can then provide clinicians and researchers with a simple representation of branch function summarised at and across selected ranges of Strahler order, translating the complex high resolution point cloud of voxels into a relatively small set of branches indicating spatial lung function which is easier to read and understand. The observation of inferior-superior gradient reversal and expected positive anterior-posterior ΔHU gradient confirms the accuracy of ΔHU assignment to simulated tree branches and reliable reflection of functional behaviour. The marked improvement to FOT modelling of R5-R20, from a correlation coefficient of 0 to 0.43, supports the successful capture of small airway functional information by this method.

This study has a number of limitations. The sample size of 23 is relatively small compared to other studies using paired inspiratory-expiratory CT analysis (32). The imbalance of sex in the asthmatic group limits the scope of observations given important differences in the thoracic geometry (298) and airway disease pathophysiology (306) between the sexes. The R5-R20 model applied with constriction weight $w=0.4$ lies away from the identity line ($y=x$), still exhibits notable negative bias in measurements and under predicts, on average, clinically measured R5-R20 to a significant degree. However, the extent of correlation increase is not negligible (43%), and shows clearly that the inclusion of ΔHU as a parameterisation for radii reduction is increasing R5-R20 in a non-uniform manner that mimics the clinically observed measurements. It is clear that more work is required in order to build, on this observation, a more accurate model that can better match the range of the clinical output.

The increase to small airway resistance (R5-R20) correlation as oppose to relatively total (R5) or conductive (R20) airway resistance, which indeed lost some correlation, may be due to the direct utilisation of the distal airway branches, Strahler order 1, in the acquisition of functional information. That is the accuracy of ΔHU in representing airway function is greatest at the level of the small airways, and drops off when averaging up the airway tree. It would be useful in further study to compare airway prediction of function using ΔHU in this manner, to a gold standard measurement of ventilation in the small and major airways, for example lobar ventilation using hyperpolarised 3He gas MRI, to test this hypothesis.

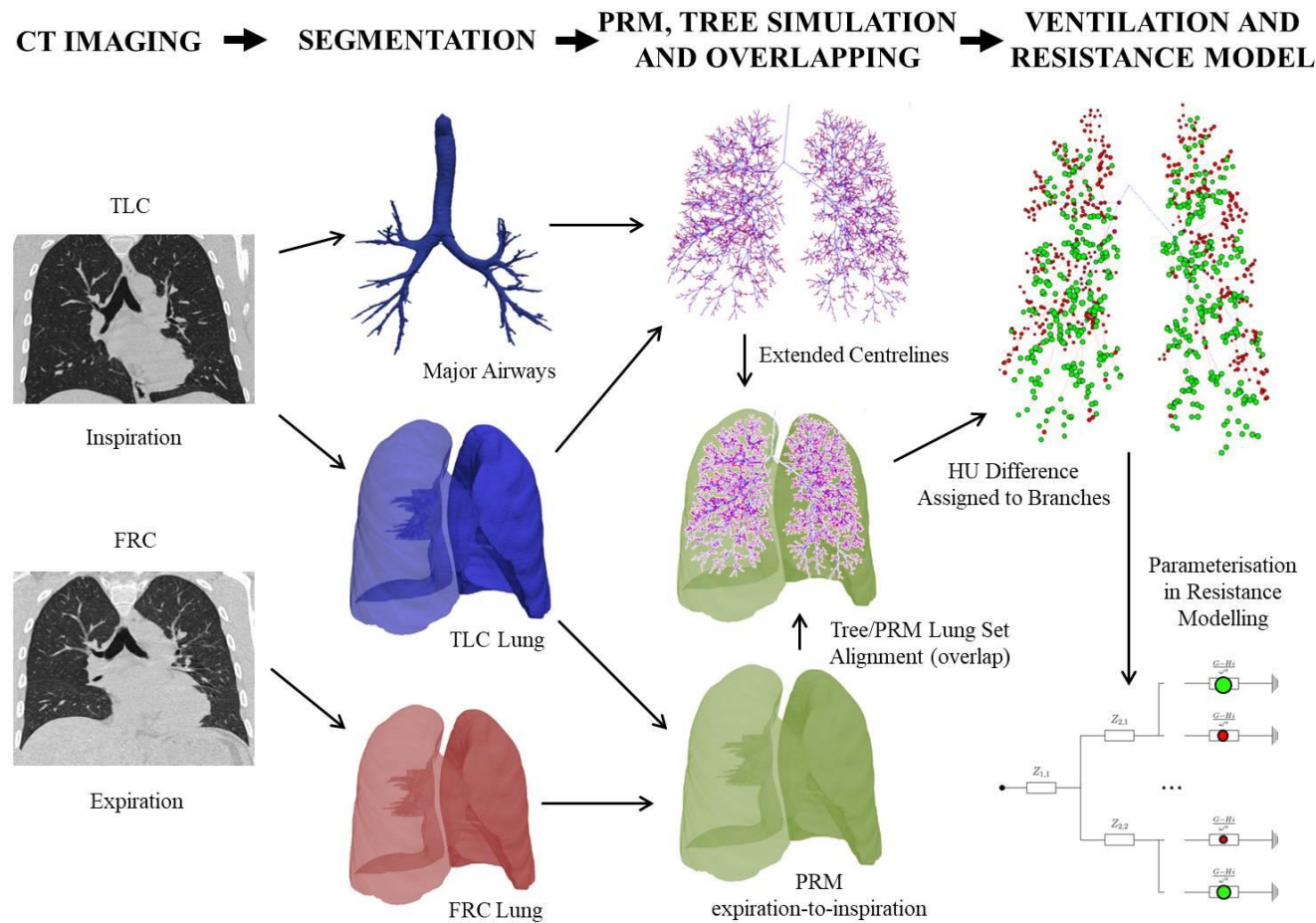
The assignment of parenchymal function to airway branches may have application for any medical operation concerning the airway tree, bronchial thermoplasty for example (167). The indication of precise branches and regions of the lung, likely to be least functioning as predicted, for example by a low assigned ΔHU , may allow for optimisation of surgical techniques through the targeting of airways which will benefit most from treatment. It would be useful to cross analyse ΔHU assignment with aerosol deposition studies (307,308), including computational fluid dynamics (CFD) predictions, to see how particle concentration correlates spatially with ΔHU , which may then predict spatial response to treatment from inhaled therapies under varying conditions imposed on the airway tree.

The visualisation algorithm may be developed into a tool to provide regional and airway branch specific evaluations of lung function; this would essentially be a software development project. Improved FOT modelling of small airway resistance will provide more accurate models of lung function, particularly in the smaller airways, recently recognised for their important role in lung disease (281), and provide better insight into the nature of modelled processes (IOS) in asthma.

In summary, we have developed a powerful new method to visualise lung function in asthma, capture regions of segmented lung volumes from CT most representative of functional tissue, and combined image registration with extended centreline models in FOT modelling to demonstrate significant increase in modelling accuracy with respect to the small airways.

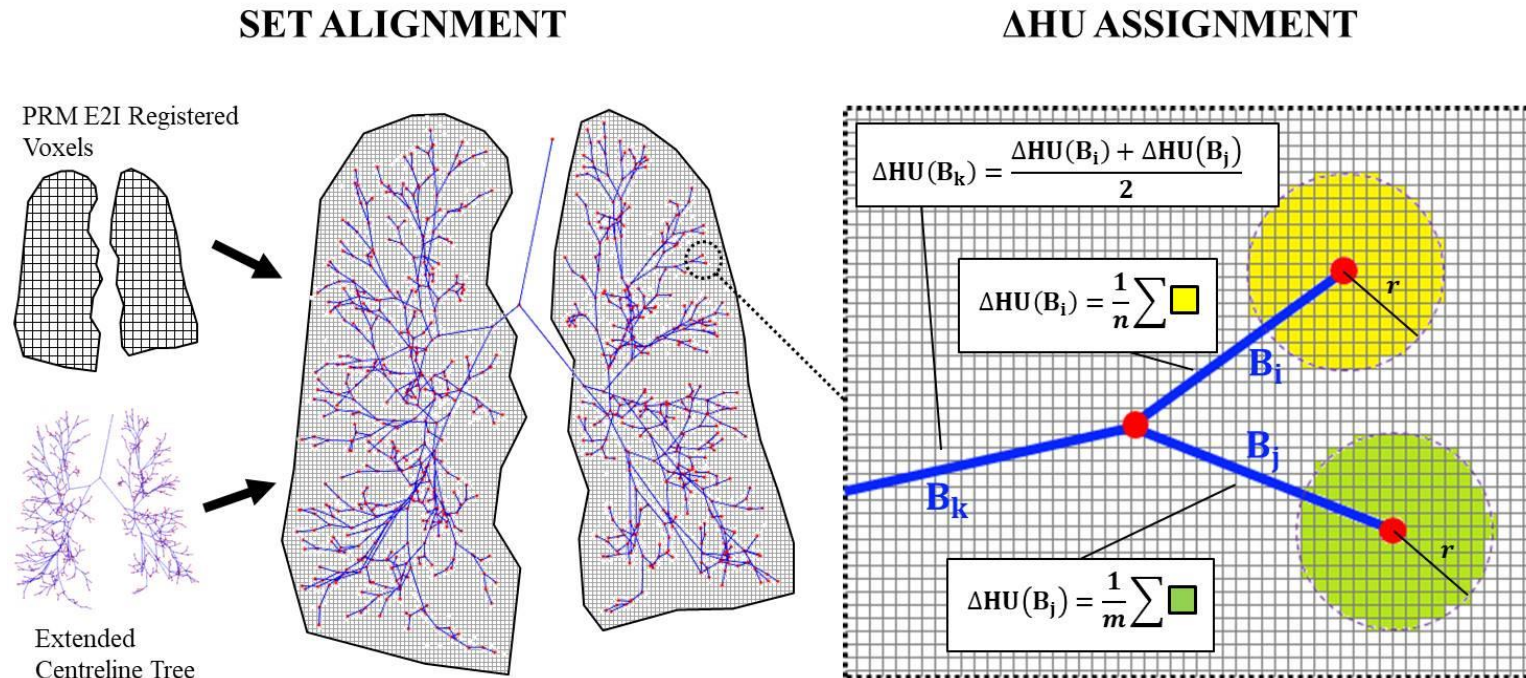
3.3.5 Figures and tables

Figure 3.3-1: Graphical abstract of study: raw CT data to augmented FOT modelling.



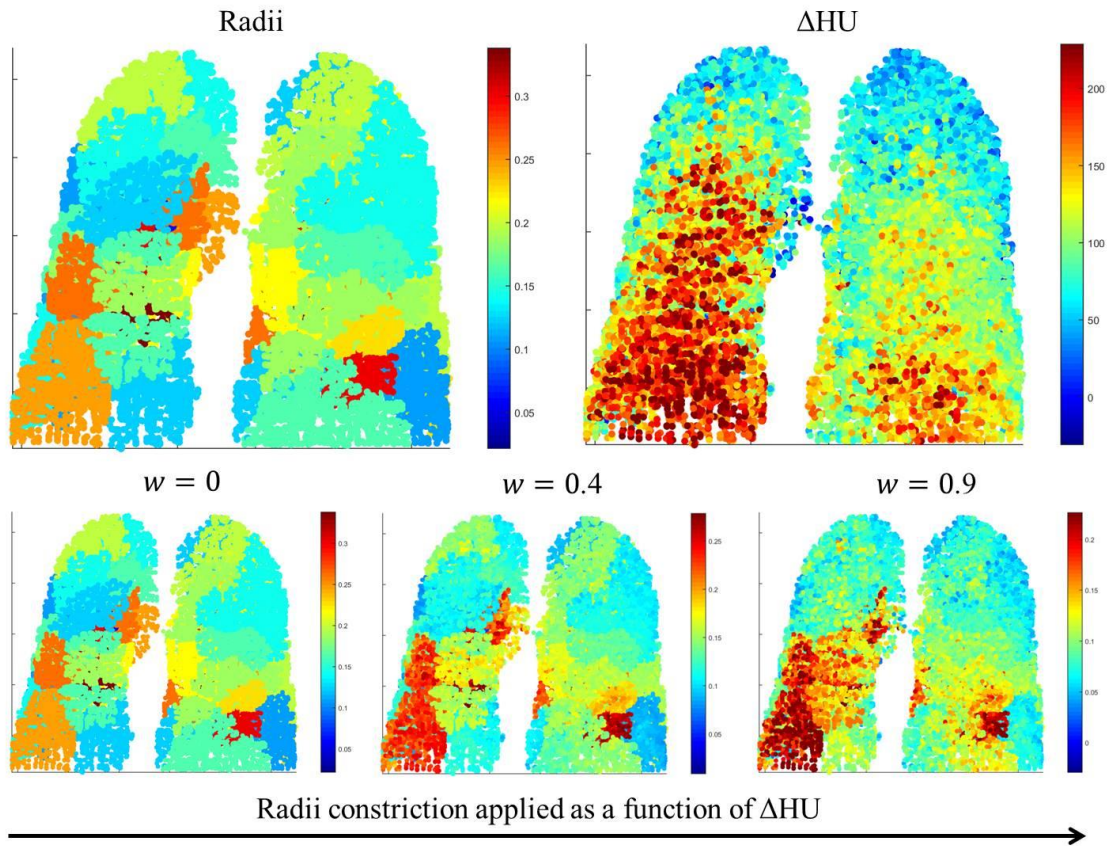
Graphical abstract: data flow from CT series to visualisation and modelling output, through four stages. Stage 1, CT imaging, series are captured at TLC (inspiration) and FRC (expiration). Stage 2, major airways are segmented from the TLC series, and lung volumes are segmented from both TLC and FRC series. Stage 3, the TLC segmentations are combined to create an extended centreline tree, both TLC and FRC series are used for image registration (PRM), then the results of these processes are spatial aligned. Stage 4, HU changes calculated from PRM are linked to airway tree branches and used for regional compliance parametrisation in resistance modelling.

Figure 3.3-2: Calculation of ΔHU scores assigned to terminal branch nodes.



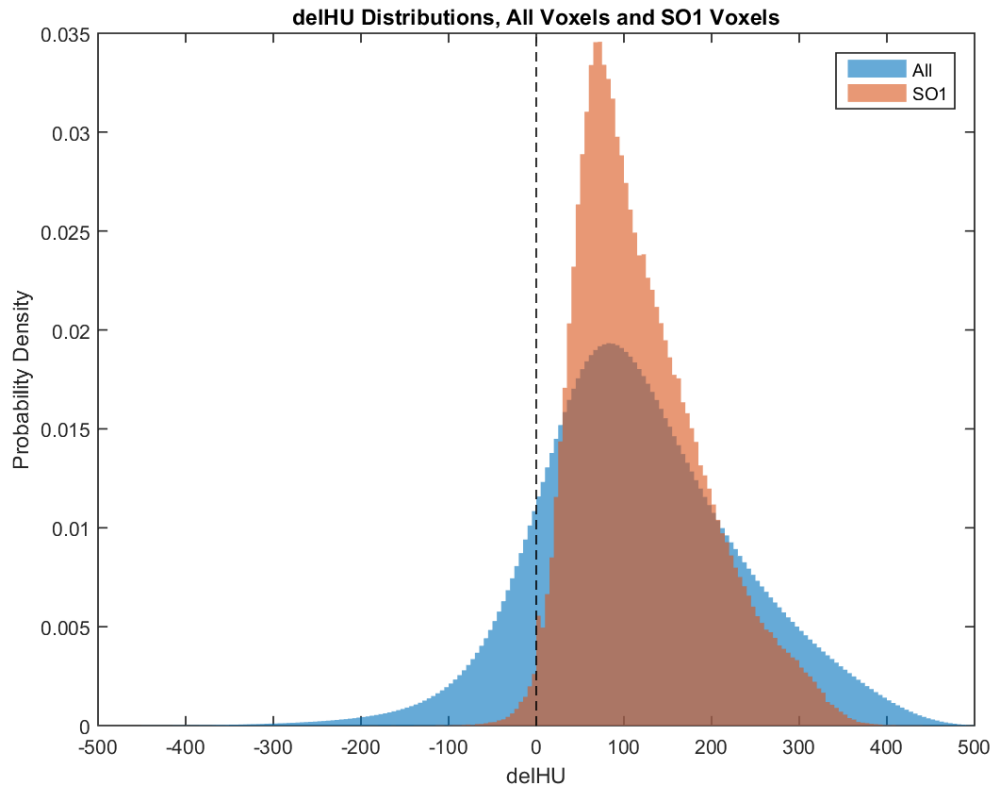
Calculation of ΔHU scores to branches of an extended airway centreline tree. The PRM registered voxel set is first spatially aligned with the tree. Then at every terminal branch node a sphere of fixed radius (r) is used to identify a proximal set of voxels (illustrated as squares here) and assign the arithmetic mean ΔHU of this set to the branch. Every branch thereafter is assigned a ΔHU value based on the average of the branches joined to it, until every branch in the tree has a ΔHU score.

Figure 3.3-3: Airway radii distribution and constriction as a function of ΔHU .



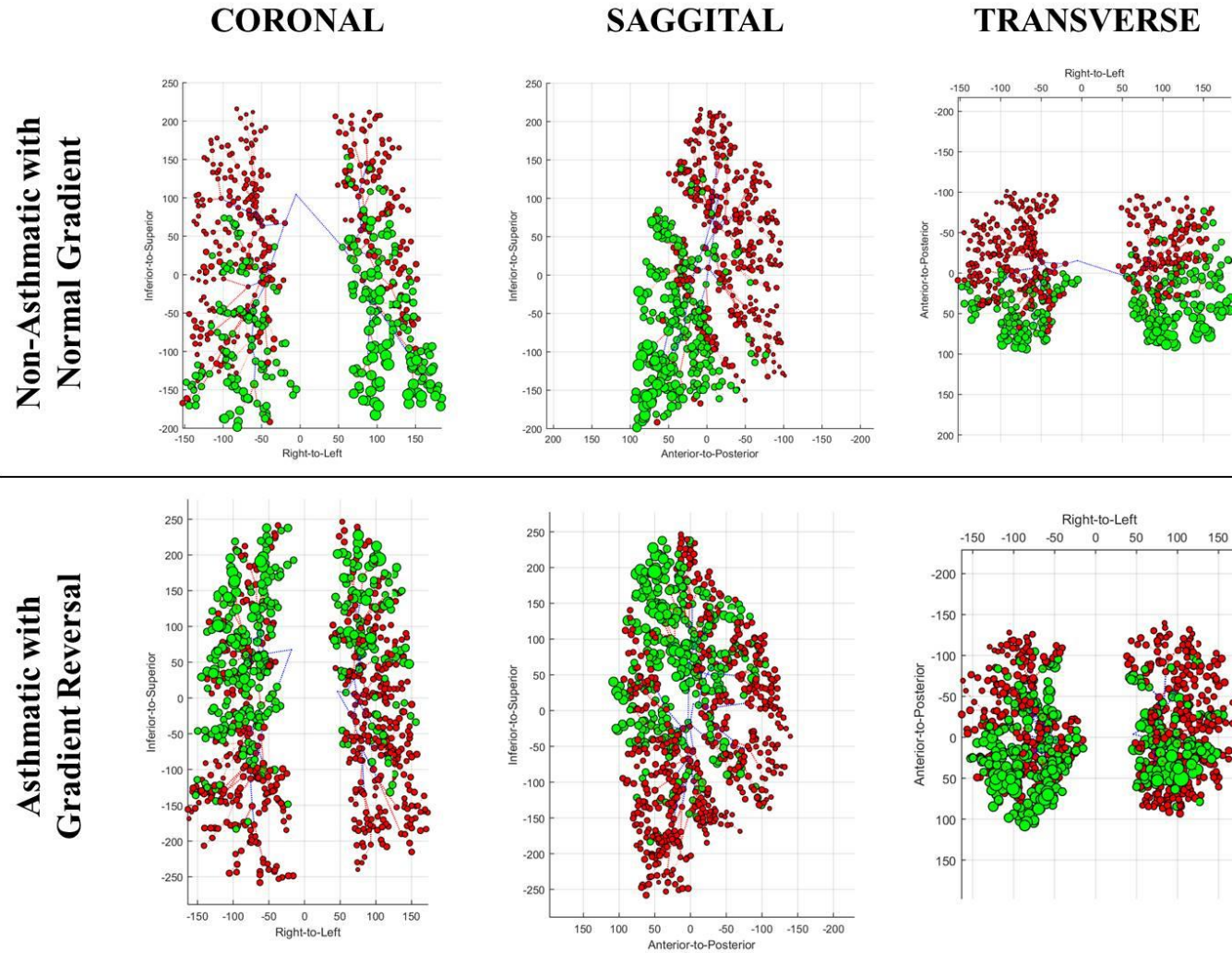
Airway radii and assigned ΔHU distribution in a subject (non-asthmatic), illustrating radii change due to increasing constriction in proportion to ΔHU and to an extent determined by weighting variable w . The top row (larger images) depicts heat maps of radii (mm) and ΔHU (HU) in a coronal cross section perceived posterior to the subject. The bottom row (smaller images) depicts heat maps of radii (mm) as constriction is applied moving left to right, with weight w annotated. $w=0$ means no change, $w=0.4$ is the weight presented as an optimal adjustment, and $w=0.9$ is a very strong constriction, note the similarity this now bares to the ΔHU distribution relative to $w=0$.

Figure 3.3-4: Histogram of ΔHU distribution for all and SO1 proximal voxels across all subjects.



Histogram of ΔHU voxel distribution for all lung segmented voxels (blue) and the subset of points within a fixed distance of some terminal (Strahler order 1) branch point (red), across all subjects ($n=23$). $\Delta HU=0$ annotated as vertical dashed line. It should be noted that $\Delta HU > 0$ is expected in general for a region that releases gas on expiration (expiration density > inspiration density), and it is clear that the voxels captured about the terminal branch points satisfy this to a greater extent relative to the total voxel set.

Figure 3.3-5: Branch assigned Δ HU visualisations.



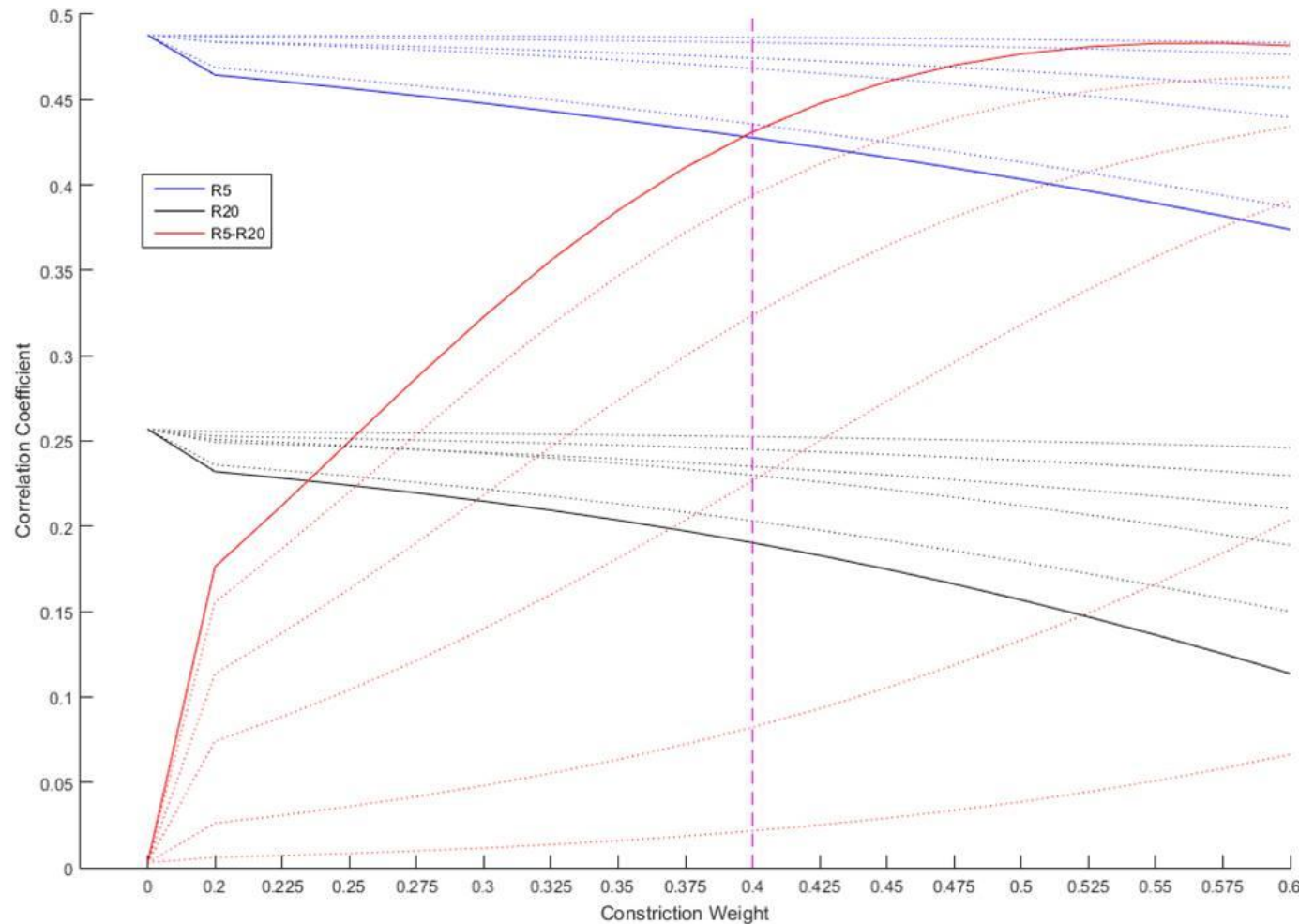
Branch Δ HU visualisations demonstrating gradient reversal.

Each row is a subject, selected to illustrate a healthy subject with biologically normal inferior-superior ventilation gradient (top row), and asthmatic subject with gradient reversal (bottom row).

Branches with a Strahler order between 5 and 12 are plotted.

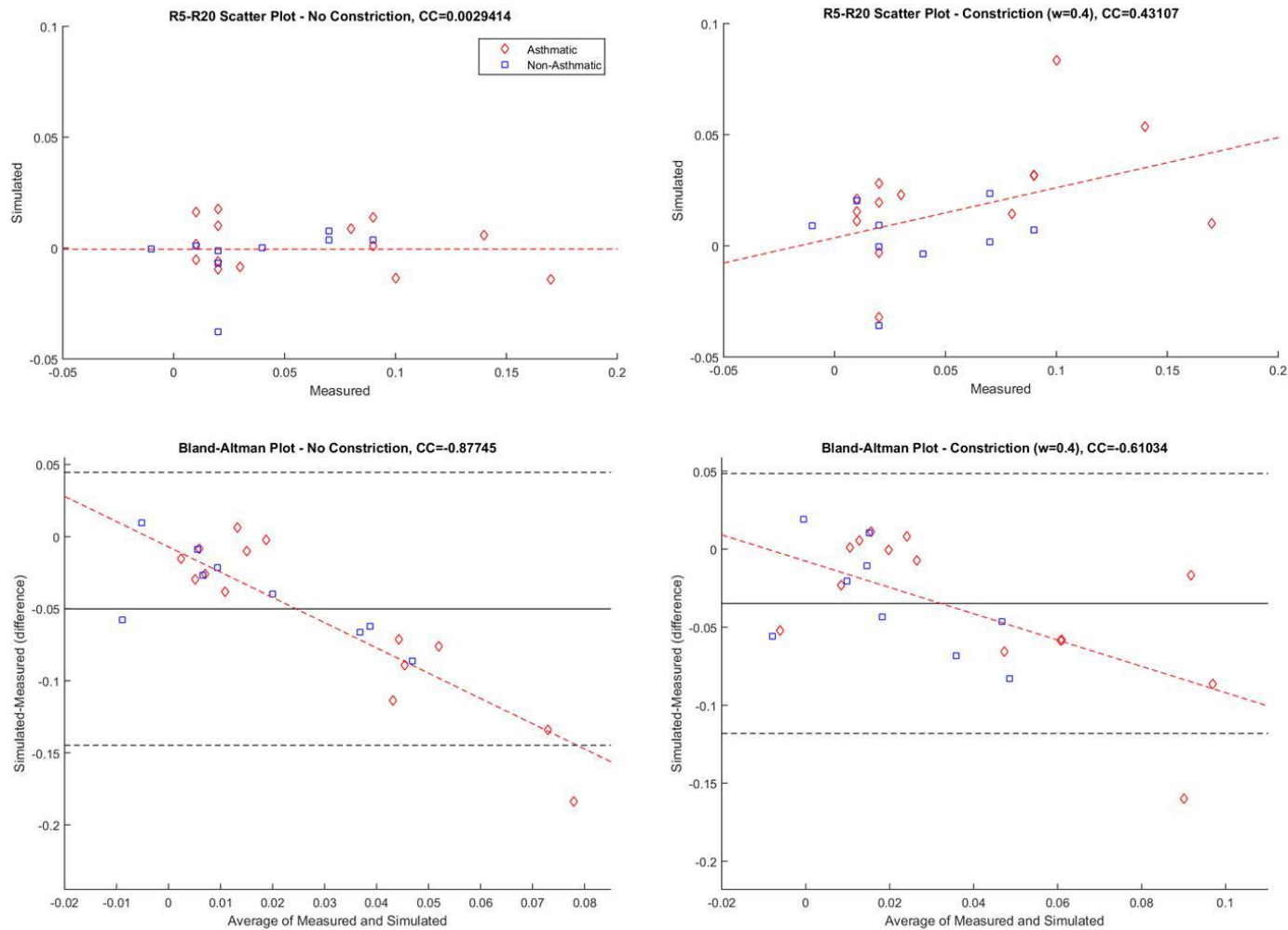
Branch Δ HU is represented by a disc positioned at the distal node with radius directly proportional to Δ HU value, and coloured green or red if the Δ HU is above or below the mean Δ HU (over all plotted branches) respectively.

Figure 3.3-6: FOT resistance correlations across varying constriction intensities and SO ranges.



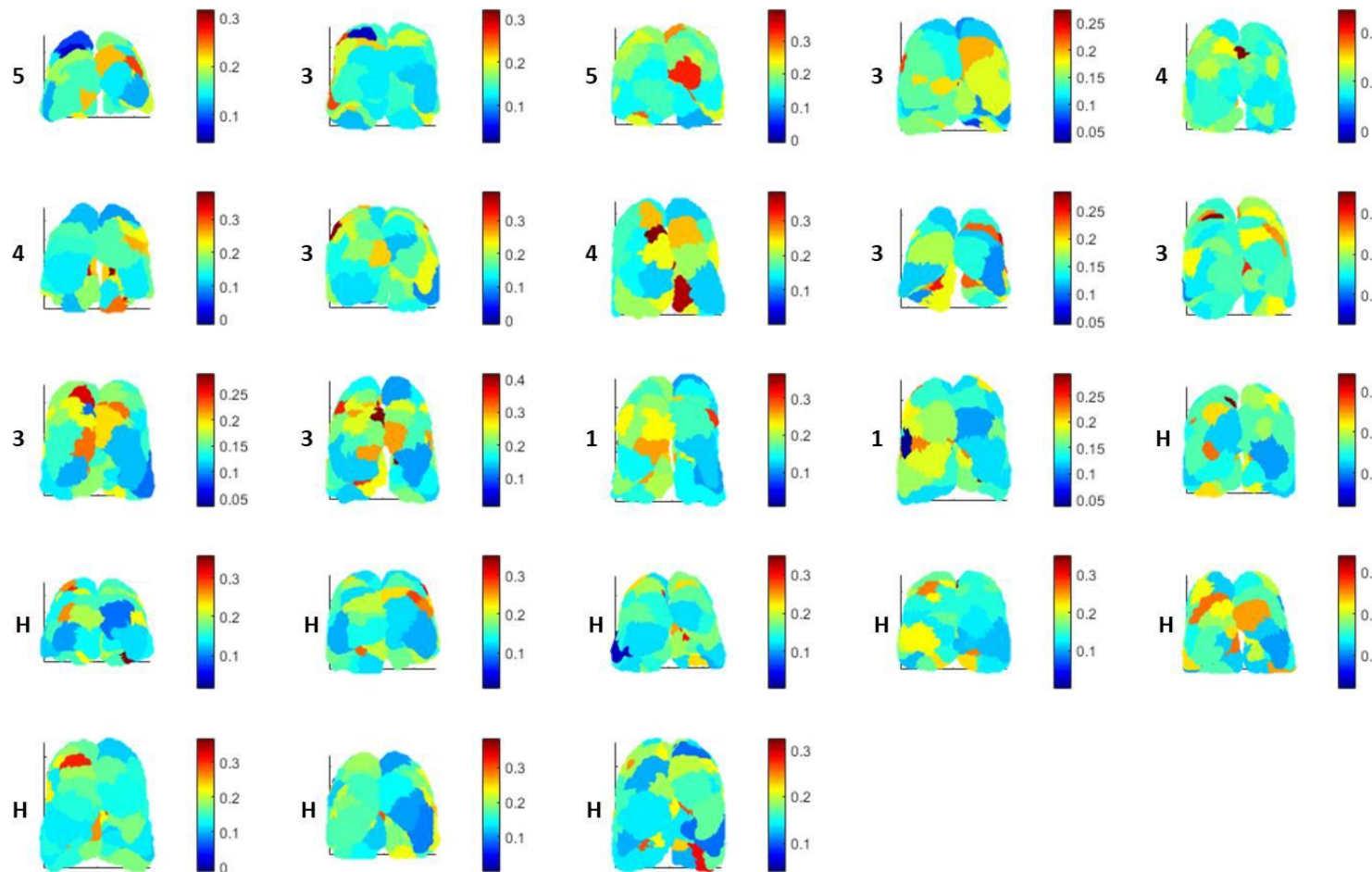
Pearson correlation coefficient (vertical axis) between clinically measured and computationally simulated R5-R20 across a range of constriction weightings (horizontal axis). The solid lines indicate constriction applied across all branches with a Strahler order of less than 11. The dashed lines illustrate the same except across a subset of Strahler orders from 1 to n, where $n=1, 2, 4, 6$ and 8 . It can be seen that R5-R20 correlation improves significantly with increasing constriction, whilst the individual features R5 and R20 individually lose correlation with a shallower gradient. The point $w=0.4$, indicated by the dashed line, is the smallest weight at which R5 correlation falls below R5-R20 correlation, and is presented as an optimal weighting for constriction.

Figure 3.3-7: Scatter and Bland-Altman plots for selected measured and simulated R5-R20 distributions.



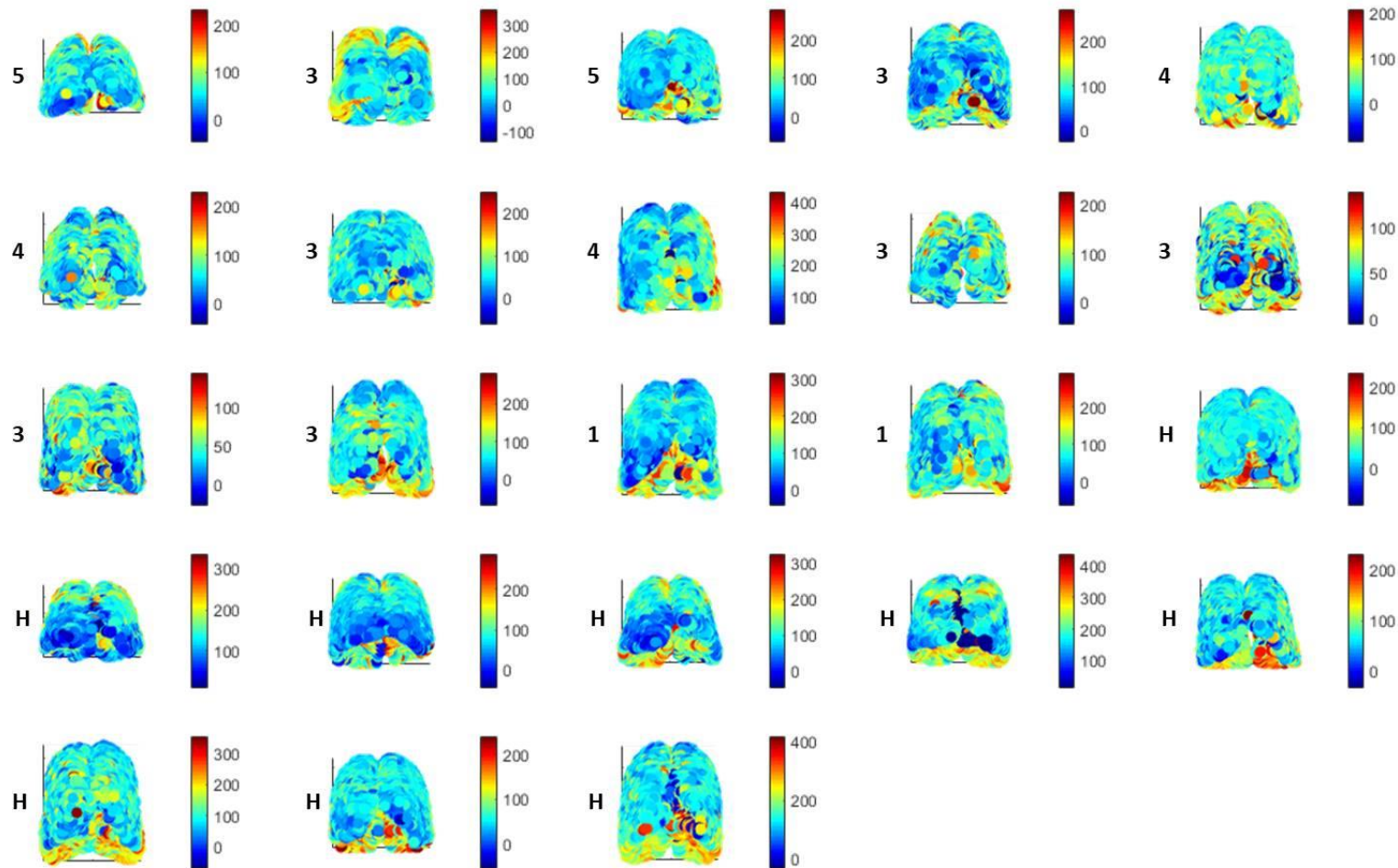
Scatter and Bland-Altman plots for measured and simulated R5-R20 distributions, in unaltered data (left) and after constriction ($w=0.4$) right. CC, correlation coefficient (Pearson's). Before constriction, there is no correlation (0.00 to 2 d.p.) between measured and simulated R5-R20 with the given FOT model, and a strong negative bias in measurement. After constriction is applied, positive correlation is achieved with a coefficient of 0.43 (2 d.p.), and the bias in measurement is decreased (CC: -0.88 to -0.61, 2 d.p.).

Figure 3.3-8: Heat maps of terminal branch radii across all subjects.



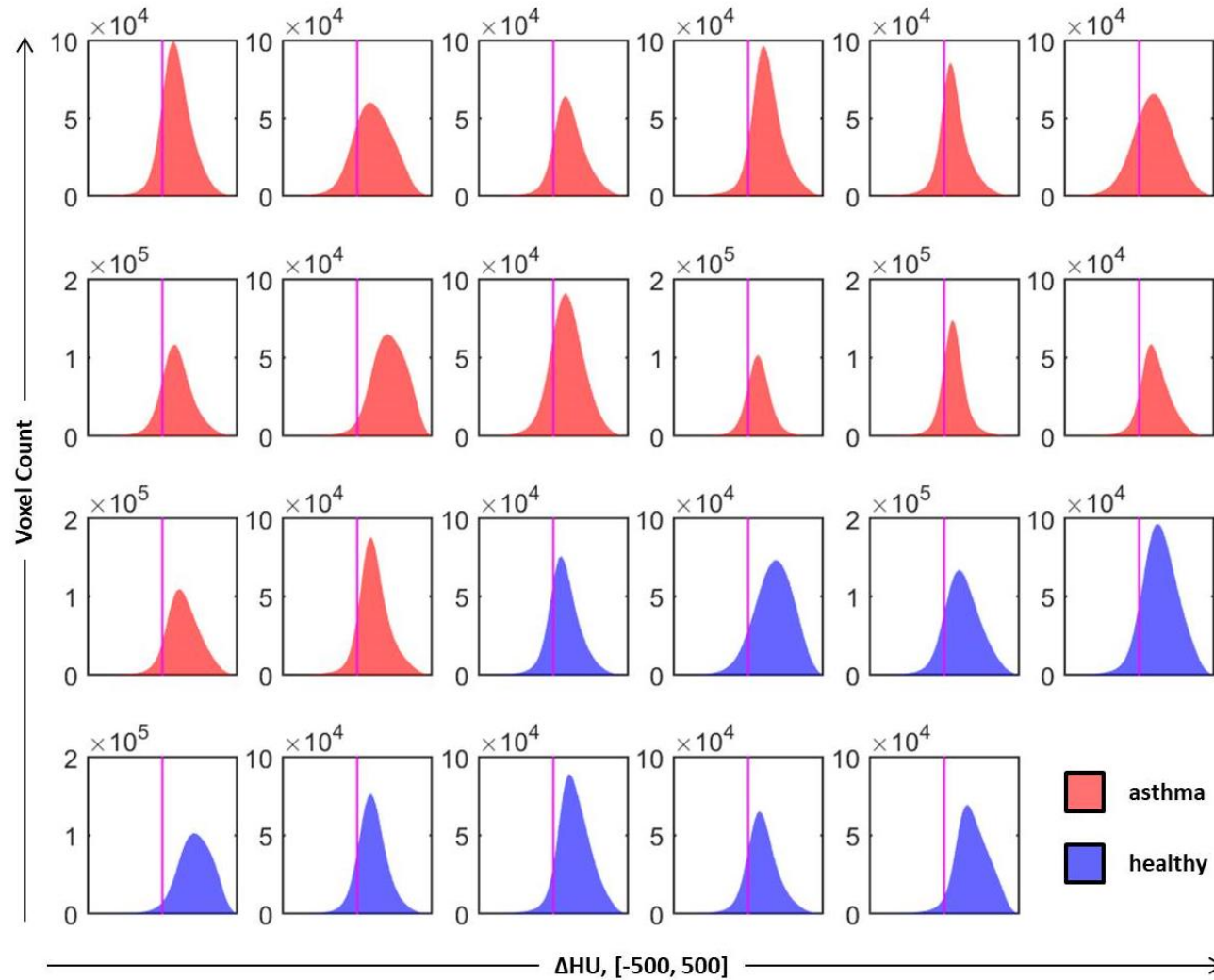
Heat maps of terminal (Strahler order 1) branch radii across all subjects. Perspective is orthogonal to the coronal plane, from the anterior direction. Branch radii in mm forms the colour axis with limits given by minimum and maximum occurring radii in a subject; value then determines face colour of a disc, of constant radius, centred at the distal branch end. Colour bars to the right of each subject indicate value (mm) range with associated colour. Subject GINA step is annotated to the left of each image, with a 'H' for non-asthmatic subjects.

Figure 3.3-9: Heat maps of terminal branch assigned ΔHU across all subjects.



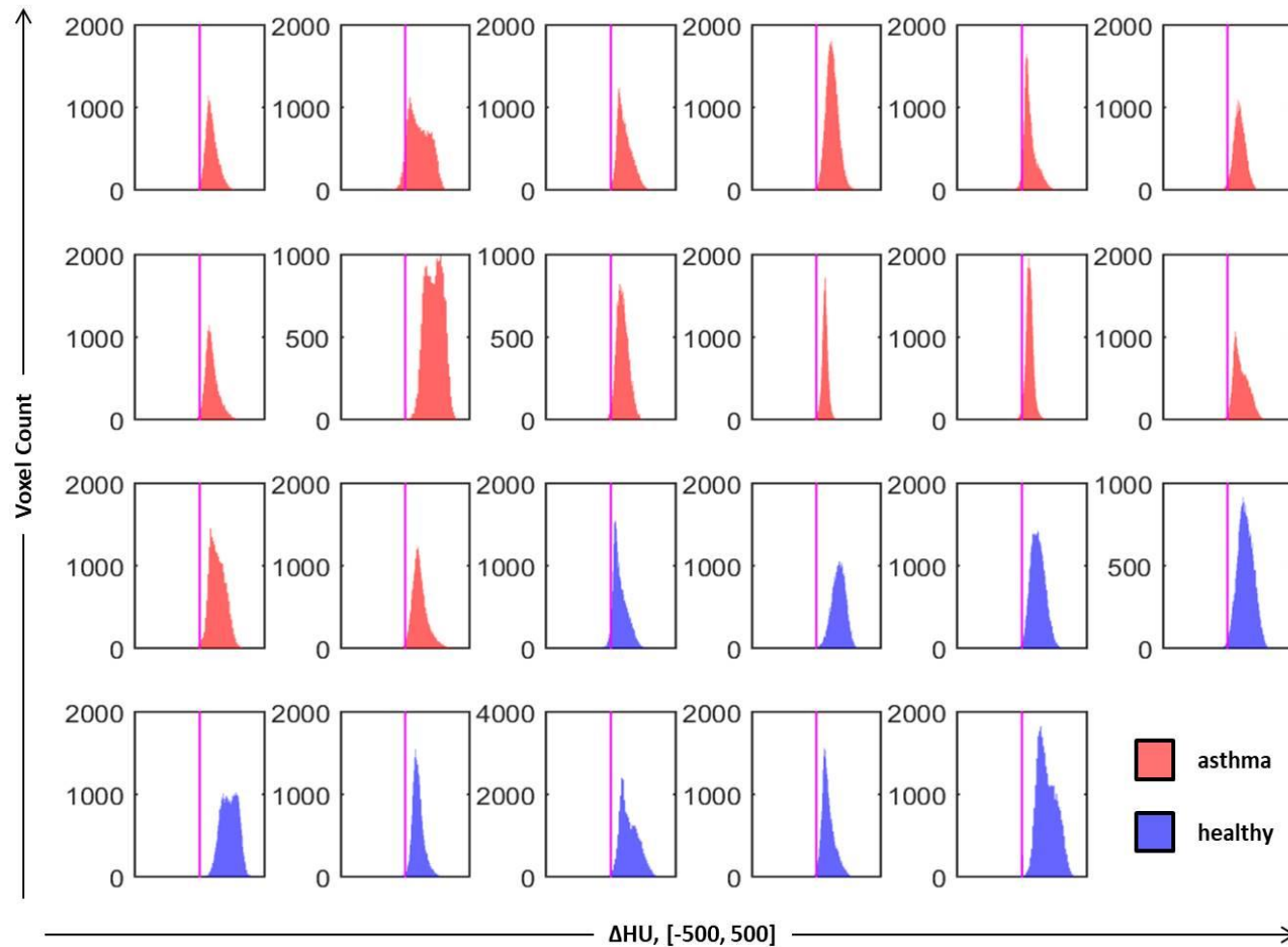
Heat maps of terminal (Strahler order 1) branch assigned ΔHU across all subjects. Perspective is orthogonal to the coronal plane, from the anterior direction. Branch ΔHU in Hounsfield Units forms the colour axis with limits given by minimum and maximum occurring ΔHU in a subject; value then determines face colour of a disc, of constant radius, centred at the distal branch end. Colour bars to the right of each subject indicate value (ΔHU) range with associated colour. Subject GINA step is annotated to the left of each image, with a 'H' for non-asthmatic subjects.

Figure 3.3-10: Histograms of ΔHU over all segmented lung volume voxels across all subjects.



Histograms of ΔHU over all segmented (Yacta) lung volume voxels across all subjects. Histogram coloured by asthmatic (red) and non-asthmatic (blue) as indicated bottom right. Vertical axis ranges from 0 to the indicated limit for all subjects, measuring voxel count falling within a range of ΔHU . Horizontal axis marks ΔHU , and ranges from -500 to 500, with 0 dead centre, annotated by a vertical line (magenta), for all subjects.

Figure 3.3-11: Histograms of ΔHU over all SO1 proximal voxels across all subjects.



Histograms of ΔHU over all SO1 assigned voxels, i.e. those identified within a specified distance of some terminal branch point in the airway tree, across all subjects. Histogram coloured by asthmatic (red) and non-asthmatic (blue) as indicated bottom right. Vertical axis ranges from 0 to the indicated limit for all subjects, measuring voxel count falling within a range of ΔHU . Horizontal axis marks ΔHU , and ranges from -500 to 500, with 0 dead centre, annotated by a vertical line (magenta), for all subjects.

Table 3.3-1: Clinical characteristics across all subjects grouped by asthma diagnosis.

	Asthmatic (n=14)	Non-asthmatic (n=9)
Clinical/Demographics		
Age (years)	53.5 (\pm 14.4)	53.3 (\pm 14.6)
Sex [male/female]	[5/9]	[5/4]
BMI (Kg.m ²)	25.42 (\pm 4.87)	28.16 (\pm 4.91)
Atopic [Yes/No]	[9/5]	[2/7]
Smoking (pack years)	1.9 (\pm 2.9)	1.4 (\pm 3.0)
GINA step [1/2/3/4/5]	[2/0/7/3/2]	-
No. of exacerbations. (past 12 months)	2.4 (\pm 2.7)	-
AQLQ	167 (\pm 38)	-
Asthma Duration (years)	22.0 (\pm 16.7)	-
Physiology		
Post-BD FEV ₁ (L)*	2.65 (\pm 0.96)	3.74 (\pm 1.04)
Post-BD FEV ₁ %	97.39 (\pm 19.86)	115.23 (\pm 19.63)
Post-BD FEV ₁ /FVC (%)	74.66 (\pm 11.41)	80.30 (\pm 3.88)
BD Response (% FEV ₁)	7.07 (\pm 8.69)	3.40 (\pm 2.97)
RV/TLC (%)	40.43 (\pm 10.41)	32.30 (\pm 8.75)
K _{CO} % pred.	99.86 (\pm 15.99)	94.44 (\pm 7.97)

BMI, body mass index; GINA, global initiative for asthma; AQLQ, asthma quality of life questionnaire; BD, bronchodilator; FEV, forced expiratory volume; FVC, forced vital capacity; RV, residual volume; TLC, total lung capacity; K_{CO}, index of gas exchange efficiency. Data expressed as mean (\pm standard deviation). Attribute normality was tested using one-sample Kolmogorov-Smirnov test over all subjects. Asthma / non-asthma group comparisons performed using two sample t-tests for parametric variables, and Mann-Whitney U-tests for non-parametric variables. Statistically significant difference, at the 5% significance level ($p < 0.05$), is indicated by an asterisk (*).

Table 3.3-2: Branch radii, Δ HU and resistance across all subjects, grouped by asthma diagnosis.

	Asthmatic (n=14)	Non-asthmatic (n=9)
Radii/ΔHU		
Radii (raw)	0.238 (\pm 0.026)	0.235 (\pm 0.007)
Δ HU (raw)	104.47 (\pm 38.22)	137.70 (\pm 47.51)
Δ HU (SO1)	105.48 (\pm 40.31)	139.55 (\pm 49.97)
Measured FOT		
R5	0.384 (\pm 0.094)	0.323 (\pm 0.065)
R20	0.326 (\pm 0.065)	0.287 (\pm 0.050)
R5-R20	0.058 (\pm 0.053)	0.037 (\pm 0.033)
Simulated FOT		
R5*	0.265 (\pm 0.054)	0.230 (\pm 0.012)
R20	0.263 (\pm 0.059)	0.234 (\pm 0.018)
R5-R20	0.001 (\pm 0.011)	-0.003 (\pm 0.014)
Simulated FOT (constr.)		
R5	0.429 (\pm 0.126)	0.356 (\pm 0.050)
R20	0.407 (\pm 0.106)	0.352 (\pm 0.054)
R5-R20*	0.022 (\pm 0.026)	0.003 (\pm 0.017)

HU, Hounsfield Unit; Δ HU, HU at expiration minus HU at inspiration (HU change); SO, Strahler order; R5, resistance at 5 Hz; R20, resistance at 20Hz. Data expressed as mean (\pm standard deviation). Attribute normality was tested using one-sample Kolmogorov-Smirnov test over all subjects. Asthma / non-asthma group comparisons performed using two sample t-tests for parametric variables, and Mann-Whitney U-tests for non-parametric variables. Statistically significant difference, at the 5% significance level ($p < 0.05$), is indicated by an asterisk (*).

4 Conclusions

In this closing section a summary of the results arising from work across all three case studies is given, with consideration of the stated aims and hypotheses, and with respect to both the methods developed and the outcomes reported from their application. Future directions arising from the work are discussed, and a closing statement regarding the nature of the thesis and its outcome is made.

4.1 Discussion of results

The purpose of this thesis was to develop and apply methods, including a novel mathematical approach to data analysis with TDA mapper, to visualise and extract information from CT imaging data to provide insights into the spatial mechanisms driving VH markers measured at the mouth and simulated in FOT modelling. Data visualisation has been extensively developed and illustrated in all three studies, with TDA applied in 3.1 to identify phenotypes relevant to VH discovered through SAA, and in 3.2 to generate powerful visualisations of the lungs with spatial, anatomical and functional information linked to airway tree depth. The importance of functional gradient in the inferior-to-superior direction across the lung has provided insight into this spatial mechanism affecting global VH indices, and the utility of small airway HU changes, identified by combined airway tree modelling and image registration methods, has been demonstrated in 3.3 through improvement in the modelling of small airway resistance.

In response to aim 1.6.1, many novel algorithms were developed, tested and implemented to extract features summarising information from multiple perspectives from the raw CT data. Thorough statistical testing and reporting was conducted on the resulting data set, with group comparisons based on asthma treatment intensity (GINA step) and VH magnitude. PRM registered CT data was shown to contain information on lung VH as CT feature sets best discriminated VH markers S_{acin} and R5-R20 in LDA, complimenting information from other domains. JDH based features associated strongest with airway obstruction reported from spirometry (FEV_1/FVC) whilst SAA based features, particularly inferior-to-superior ΔHU gradient, discriminated extremes of

VH markers from MBW and IOS as hypothesised. TDA mapper was applied to CT, VH and clinical data to further illustrate that data driven phenotypes driven by SAA gradient existed with relevance to asthma characterisation.

A significant point of controversy running throughout the thesis was the use of ΔHU as a marker of ‘ventilation’, as it is present in all three studies and really originated in the first. There is reason to suspect it is not always the case, as perfusion (blood flow) will have a significant effect on tissue density variation, and indeed the anterior-to-posterior ΔHU gradients being positive over all subjects in study 3.1 can be explained by blood pooling due to gravity (see **Figure 3.1-2 C**). In defence of statements treating ΔHU as a marker of voxel ventilation in the case of inferior-to-superior averaging of differences ($\overline{\Delta HU}^{IS*}$) the given direction is approximately perpendicular to the direction of gravity as a subject is lying supine on the CT scanner bed, see **Figure 3.1-2 B**, so minimizing the effect of the anterior-to-posterior perfusion gradient, and study of $\overline{\Delta HU}^{IS*}$ correlation with subject weight provided statistically significant ($p < 0.05$) evidence corroborating established theory of ventilation gradients (309). In general it may be conceded that the use of ΔHU as a marker of ventilation should be treated with caution, and more work is needed to support or reject this claim, e.g. spatial comparison of PRM registered voxels from CT with 3He MRI attenuation as a better established marker of ventilation in imaging; past studies have compared registered CT and 3He MRI at the lobar level with promising results (33).

Regarding specifically the method of SAA, it has been demonstrated to effectively summarise the heterogeneity of voxel functional value (ΔHU in this case) in a particular direction across the lungs, and proof of concept of its relevance to clinical phenotyping and investigating mechanisms underpinning VH in the lungs and asthma has been presented. Currently significant limitations of SAA developed in this thesis include i) the constriction of application to cardinal axes determined by the orientation of the subject within a scanner, the scanner bed and real spatial axis traversed by the rotating emitter; ii) arbitrary extent of voxels eliminated from the ends of ranges determined by minimum and maximum spatial values which may be sensitive to outliers, and iii) arbitrary selection of the number of intervals and interval widths for averaging of voxel functions. To address (i), the agreement with proposed and real subject orientation, and methods for generalising directional axis for evaluation should be investigated, and to address (ii) and (iii) more study of the effects of parameters variation and justification

for value selection should be made, e.g. should interval width be uniform with respect to axis distance or voxel set mass (in this study it was the former).

The major outcome of study 3.1 in response to aim 1.6.1 was the identification and visualisation, through decile-wise SAA mean comparisons, min-max visualisations and group comparison testing with S_{acin} and R5-R20 splitting about the mean, of the significant heterogeneity of inferior-superior lung ventilation in the given population and close association to VH markers measured in clinic; specifically the differences in the lower regions of the lung (**Figure 3.1-3 A**). A secondary outcome was the proof of VH relevant information that can be (and in this study was) extracted from CT data of the lungs registered at inspiration and expiration. Evidence supporting all hypothesis stated in 1.6.1 was provided. However the relatively small sample size of 52 subjects limits the statistical power to infer results observed here to a larger population, and really observations made should be validated and replicated in studies of other and ideally larger populations. The relationship between age, obesity, VH and presented gradients across the lung needs to be better understood to provide better understanding of their interaction and relationship with asthma; whilst it is evident there exists asthmatics with reversed inferior-superior ventilation gradient, and basal PRM^{fSAD} concentrations, there are healthy subjects with gradient reversal due to obesity, and some asthmatic subjects with normal ventilation gradient.

In response to aim 1.6.2 TDA mapper was applied, using Ayasdi software, to data sets of lung voxels from 4 subjects chosen to represent varying magnitudes of VH based on MBW measurements LCI and S_{acin} . This work, presented in study 3.2, was largely focused on the development of a new methodology incorporating PRM registration, Chaste extended centreline simulation, extracting lung function and airway depth measures respectively, with TDA mapper for data visualisation. Alignment of point sets was a significant problem that was sufficiently solved to enable this workflow. k-NN SO score assignment to voxels was studied for $k=1, 5$ and 10 , with $k=1$ used in the final visualisations, i.e. a voxel was associated with the SO of the nearest airway branch post set-alignment. Then a 20-NN algorithm was applied with Ayasdi to produce compressed visualisations of the high dimensional data sets, and presented in diagrams with a post-hoc node group selection process isolating the small airways based on SO.

The first clear limitation that arises on review of study 3.2 is the small sample size of subjects ($n=4$); this should be kept in mind when interpreting results as the statistical strength of observations is limited to exploratory notes and clearly further study is really required to prove existence of patterns observed in larger populations. The main result in this work is thus qualitative and at the subject specific level, providing very rich visualisations of the lungs (left and right individually), and illustration of the application of mapper for organ visualisation from imaging data in general. In addition the subjects chosen for study were arbitrarily all male, to eliminate variation due to sex at the cost of unaccountability for the effects of sex-differences (298) on patterns observed. However it may be argued that the human lung is similar enough between the sexes that it is reasonable to expect results observed here may apply in general, again requiring replication in larger cohort(s).

A significant component of this work and for study 3.3 was the set alignment algorithm, developed experimentally from sample input data and assessed in reasonableness of the output behaviour, especially from ΔHU assignment (see **Figure 3.3-5**). As stated above the alignment algorithm presented appeared to be sufficient for the purpose of these studies, however suboptimal in general for a number of reasons, including i) lack of automation, ii) poorly defined alignment quality criteria and iii) uncontrolled inter-subject variability. These problems arise largely out of the scenario of fitting together data points generated by different software processing the same input data for different purposes. The general problem of aligning point sets of ‘similar’ shape is an interesting one; however time constraints necessitated the ad-hoc alignment scheme developed here. To address (i), (ii) should first be resolved, i.e. a definition for alignment quality implemented, and then (iii) will be addressed by consistent treatment of alignment across all subjects. Evidence is presented in both 3.2 and 3.3 corroborating observations in 3.1 and supporting the statement that implemented alignment algorithm was fit for purpose here in any case.

The main result from analysis in study 3.2 is threefold and arising from visual observation of networks presented in **Figures 3.2-2 to 3.2-5**; i) inferior-to-superior gradient reversal of ΔHU with increasing VH is observed comparing the Z axis coordinate and ΔHU heat mappings, ii) there is a strong negative spatial correlation with PRM^{fSAD} and X axis (anterior-to-posterior) coordinate that appears to break with increasing VH, and iii) ΔHU appears more spatially fragmented with increasing VH.

Immediately (iii) suggests application of clustering to test and provide a quantitative outcome measure, i.e. link number of clusters in ΔHU node distribution with 'heterogeneity of ΔHU ' in the network. All three outcomes discriminated between asthmatic and non-asthmatic subjects studied as hypothesised. An additional result from this work was the innate segmentation potential of TDA networks, as it has been demonstrated that the 20-NN approach can segment high density tissues like airway walls, lobe pleura and blood vessels from the imaging data, demonstrated visually in **Figure 3.2-7**.

Finally in response to aim 1.6.3 another set of methods were developed and applied in study 3.3, across a cohort of 23 subjects, i.e. those from the total population with both PRM registered CT data and viable extended centreline tree structures for use, the latter being the main limiting factor. The primary aim in this study was to demonstrate that using ΔHU about the small airways based on extended centreline simulation could improve correlation between measured and simulated small airway resistance, represented by R5-R20, in FOT computational modelling. Secondary aims of improving the ventilation signal of ΔHU and potentially improved visualisation of lung ventilation using tree structures with assigned ΔHU to branches arose naturally on route to this goal. As with previous studies a significant portion of the work was dedicated to development of novel methods and then application to achieve the given aims.

A significant problem that had to be solved to make this final study feasible was the identification of points within a certain distance, by the standard Euclidean metric, to a given point, where the point set for checking was large (order 10^7). This was solved using cell array optimization described in 2.5.6, as part of an algorithm implementing ΔHU assignment to terminal tree branches, followed by averaging to assign ΔHU to every branch in a tree. A limitation of the method as implemented is that it costs at least 10 minutes per subject to compute on a standard machine, which is acceptable for local use on the given sub-population ($n=23$), but should be optimized really for application to larger cohorts, e.g. use of recursive programming to take advantage of the innate recursive structure of airway trees N.B. this does not include extended centreline tree calculation and PRM image registration times, the latter being most costly at present. It is clear from visual output (see **Figure 3.3-5**) that the series of algorithms were implemented successfully and to a sufficient level of accuracy for analysis.

The primary aim in 1.6.3 was achieved through demonstrated increase in the correlation between measured and simulated R5-R20, see **Figure 3.3-6**, supporting the hypothesis that using ΔHU from PRM registered CT data significantly improves modelling of small airway resistance, essentially incorporating into the FOT model a finer measure of regional compliance as affected by radii adjustment as a function of ΔHU . It should be noted however that there exists a negative bias in the model presented, even post ΔHU based adjustment, and that the output does not follow $y=x$. This suggests limitation in the FOT model for accurate simulation of R5-R20 that should be investigated further to produce more viable simulations in future work.

An important secondary outcome was the demonstration of ΔHU distributions shifting right of zero, as this shows isolation of voxels that exhibit behaviour more in line with parenchymal tissue through gas release on expiration ($\Delta HU > 0$), see 2.5.2. ΔHU distributions did not shift significantly with respect to the mean, and demonstrated bimodal behaviour in many cases, likely due to the strong anterior-posterior ΔHU gradient observed across all subjects due to blood pooling / supine lung density gradients. Thus evidence was provided supported the hypothesis that utilization of voxels about terminal branch points of extended centreline trees identifies those voxels demonstrating more functional behaviour relative to the whole voxel set.

Finally an algorithm for visualisation of airway trees with the assigned ΔHU in branches across a specified range of SO was developed and implemented successfully to once again show inferior-superior ΔHU gradient reversal in subjects, see **Figure 3.3-5**, and arguably a cleaner presentation relative to min-max projection in 3.1 and 20-NN networks in 3.2; supporting the hypothesis that airway trees with branch ΔHU assigned can show a clear picture of the distribution of ventilation in the lungs from CT data, an outcome which may be useful for clinical application where knowledge of major airway ventilation potential due to connected small airway function can be applied, e.g. bronchoscopy and surgery targeting specific airways based on such CT analysis.

4.2 Future work

There are numerous clear paths for development arising from the work presented across all studies, primarily through improving and extending the methodologies developed, and through future validation studies in larger cohorts that seek to demonstrate replication and the extent of observed results, and provide more statistical evidence for statements made. It is clear for example that ventilation gradient reversal exists in numerous asthmatic subjects within the studied population, but the level of prevalence of this abnormal ventilation pattern in the general asthmatic population is as yet unestablished; a future study examining $\overline{\Delta HU}^{IS*}$ or a similar metric based on imaging data from a large population of asthmatic subjects is required.

The SAA algorithm may be further developed and studied in a number of ways. The direction of analysis may be generalised beyond the three Cartesian axes defined by the CT scanner, to be parameterised by three coordinates and provided a much larger space for study of directional gradients of voxel functional values; this may be achieved, for example, by defining the line through the mean point and another in the chosen direction, partitioning then a segment of the line capturing an ROI, and at defined intervals on every partition using planes orthogonal to the line to identify voxels for analysis by proximity to the given plane. Another improvement beyond what was done in this thesis would be to study gradients of functional values besides ΔHU , such as PRM^{fSAD} , to see how other quantities vary across the lung. Generalising the SAA algorithm in this way will then necessitate computationally efficient algorithms for analysing the parameter space, including variation in interval size and number, gradient directions in three dimensions and extent of boundary trimming if any.

The relationship of ΔHU to ventilation and perfusion can be further studied, and methods to provide other measures of ventilation explored. The Jacobian of the geometric deformation mapping generated in image registration provides a direct measure of volume change about a voxel, and looks promising to study in place of and perhaps augmenting HU changes in future studies. Validation against a measure of ‘true ventilation’ may be achieved through spatially aligned comparisons of 3He MRI attenuation coefficients in space, essentially matching lung volumes of approximate (but certainly not exactly) equal shape, involving cubic lattices of disparate resolution (MRI typically producing larger cubes).

An improved method for similar point set alignment may be developed and automated for carrying out the work presented in 3.2 and 3.3 in significantly larger data sets. The clear first step is initial alignment of point sets based on mean or median centralisation and matching variance or ranges. Matching rotational orientation automatically is less clear but can be avoided if input data is consistent; in general one set may need to be rotated and reflected to match orientation to the other. Final alignment may be achieved through perturbations in translation and enlargement incorporating a measure of alignment quality based on point by point proximity measure to the complimentary set. There again exists the computational difficulty of handling large numbers of points in image volumes, which can be assisted using cell array methods and coarse graining of the high volume point set, i.e. replace high density cloud with lower volume grid prior to perturbation and alignment quality assessment.

There is a great deal of potential for development in the TDA mapper approach to lung visualisation presented in 3.2, and strong requirement for automation to facilitate application to larger numbers of subjects for statistical analysis. A significant factor in the current workflow is the difficulty and computational cost of data preparation and quality control, requiring expertise in image registration, airway centreline modelling and lung volume modelling with various software developed by distinct groups; there is a clear need for these methods to be centralised in more general and encompassing software packages that can perform the pre-processing automatically with interfaces facilitating rapid QC/QA by the researcher, i.e. medical imaging data in, anatomical structures and registered VOI out. The TDA mapper analysis then has a huge parameter space for exploration, varying clustering methods, resolution, gain and perhaps most significantly lenses and input feature sets; in the current workflow this is severely encumbered by the order 10^7 input point set, and aforementioned coarse graining or other volume reduction techniques should be explored for improving rapidity.

A significant question arising from the work was the absolute accuracy of extended airway centreline structures with respect to the true airway tree, which cannot be accurately segmented due to resolution limitation in CT due to limit on radiation dosage a subject may experience. In this and other studies it is assumed that the initial segmentation is essentially exact, up to approximately eight generations from the trachea, thereafter there will be error in specific branch locations and certainly out to the terminal branches there is little reason to expect exact alignment. The key question is

whether on average the branch dimensions and reported SO are accurate, and quantitatively to what extent they match. Lung transplants provide excised lung samples with airways that may then be subjected to very high resolution CT scanning (196) for potential calculation of true airway trees. There is thus clear potential to answer the stated questions in a future study combining simulated airway centrelines with airways calculated from high resolution CT scans of excised lung volumes; ideologically this comparison exercise is simple, but obtaining accurate tree models from lungs removed precisely due to major dysfunction, e.g. lung cancer, seems likely to prove challenging.

The Δ HU branch assignment algorithm developed in 3.3 can be further explored through testing with varying sphere sizes and other methods such as k-NN for associating voxels proximal to terminal branch points. Perhaps the most significant imminent further application of this work is the isolation of functional voxel sets based on simulated tree structures to improve algorithms developed in 3.1 and other imaging features based on registered lung voxels; i.e. run SAA and other intra and inter lung quantitative measures on the sets of voxels identified as being proximal to the acinar airways, hypothetically providing a cleaner signal of ventilation relative to the total voxel set identified in prior segmentation of lung volumes for registration. The effects of airway obstruction at various depths in the airway tree should also be explored, as intermediate airway closure may be detectable due to the resultant lack of function (volume change) in the connected smaller airways caused by obstruction.

4.3 Closing statement

In this thesis the problem of extracting and evaluating information from CT imaging using novel mathematical methods, including TDA mapper, to investigate VH within the human lung and asthma has been presented and explored through the development of a plethora of methods and analyses. Evidence of the heterogeneity and association of directional gradients of ventilation across the lung with VH has been presented and visually elucidated from multiple perspectives. The application of TDA mapper in the phenotyping and visualisation of medical imaging data has been investigated and demonstrated at both population and organ specific resolutions. The potential for CT imaging metrics, based on PRM, to support FOT modelling of small airway resistance has been examined and evidence of model improvement presented.

This work presents a particular case of mathematical ideas and problem solving applied to the highly complex challenge of disease characterisation and study in an interdisciplinary environment arising from the clinical setting. It is a combination of routine medical research methods and statistical analyses, with creative and exotic experimentation using cutting edge software and topological methods to examine contemporary challenges in asthma and respiratory research, specifically supporting understanding of the mechanisms of VH in asthma using CT imaging data.

5 Appendices

All code within these appendices is provided in the Matlab programming language. Iterative and conditional statements, and comments, typically colour coded blue / green in the default Matlab IDE (integrated development environment) settings have been formatted in bold type, for monochromatic presentation.

Code is sampled from actual scripts used in this thesis. N.B. for brevity these are generally only pieces of larger sections that would be functional, purposed to demonstrate the significant functions used and exemplify algorithmic structures through which data was processed to obtain the results presented.

Appendix 1: JDH ellipse minor / major vector calculation and annotation

```
C = cov(M);  
[V,D] = eig(C);  
centr = [mean(M(:,1)) mean(M(:,2))];  
V1 = V(:,1); V2 = V(:,2);  
V1 = V1./norm(V1); V2 = V2./norm(V2);  
ev = [sqrt(D(1,1)) sqrt(D(2,2))];  
V1 = -ev(1)*V1; V2 = -ev(2)*V2;  
quiver3(centr(1),centr(2),maxM+10,V1(1),V1(2),0,'black','LineWidth',2)  
quiver3(centr(1),centr(2),maxM+10,V2(1),V2(2),0,'black','LineWidth',2)
```

M is an $k \times 2$ matrix of points (x,y) of cells in the JDH exceeding the described threshold ($0.25 \cdot \max(\text{voxel frequency})$), cov.m computes covariance matrix of these points, eig.m computes eigenvalues and quiver3.m graphically plots vectors as arrows.

Appendix 2: SAA feature extraction in the inferior-superior (z) direction

```
M = sortrows(M,5); % sort into ascending z value
rngZ = range(M(:,5)); % range SAA acts over pre-trim
trm = 0.15*rngZ; % length to be trimmed
start = floor(min(M(:,5)) + trm); fin = floor(max(M(:,5)) - trm);
pnts = floor(linspace(start,fin,11)); % dividing points
trck = 2; % index of dividing point to test
VALS = zeros(s(1),10); % to record interval values
valCnts = zeros(1,10); % counts (to compute stats)
for j = 1 : s(1)
    coord = M(j,5);
    if coord < start
        continue
    end
    if coord >= fin % using intervals of form [start,fin)
        break
    end
    if coord >= pnts(trck)
        trck = trck + 1; % move onto dividing point
    end
    valCnts(trck-1) = valCnts(trck-1) + 1;
    VALS(valCnts(trck-1),trck-1) = M(j,1) - M(j,2); % add dHU
end
valMeans = zeros(1,10); valStds = valMeans;
for j = 1 : 10
    valMeans(j) = mean(VALS(1:valCnts(j),j));
    valStds(j) = std(VALS(1:valCnts(j),j));
end
pMean = polyfit(1:10,valMeans,1); pStd = polyfit(1:10,valStds,1);
toWrite(20) = pStd(1); % std( $\Delta HU$ )IS
toWrite(21) = pMean(1); % mean( $\Delta HU$ )IS
G = zeros(1,9);
for j = 1 : 9
    G(j) = valMeans(j+1)-valMeans(j);
end
toWrite(23) = mean(G); % mean( $\Delta HU$ )IS*
```

(supporting text overleaf)

M is a data matrix of registered lung voxels with columns [eHU iHU x y z]. In the preamble note 0.15 defines 15% of the range of the inferior-superior (z) length to be avoided at each end. Dividing points are determined between these inner boundaries (with trimmed exterior removed). An index then flags when a dividing point has been crossed, moving calculation on to the next interval. In the first iterative statement ΔHU values and point counts are summed and recorded respectively. Then means and standard deviations are calculated on the ten intervals, and polynomial fitting (dimension one) is applied to calculate features as the gradients of fitted lines, and $\overline{\Delta HU}^{IS*}$ is computed as the mean of the differences between intervals.

Appendix 3: Inter-lung feature extraction

```
L1vcnt = length(L1(:,1));
L2vcnt = length(L2(:,1));
L1dhu = L1(:,4)-L1(:,5);
L2dhu = L2(:,4)-L2(:,5);
L1avg = mean(L1dhu); L1std = std(L1dhu);
L2avg = mean(L2dhu); L2std = std(L2dhu);
fprintf(fID,'%f ',abs(L1avg-L2avg)); %RLmeanDiff
fprintf(fID,'%f ',abs(L1std-L2std)); %RLstdDiff
fprintf(fID,'%f ',1-min(L1vcnt/L2vcnt,L2vcnt/L1vcnt)); %RLsizeRat
```

N.B. ‘fprintf’ here prints to text file in floating point number (%f) format; simple method to output data structure to pass into storage (MS Excel / other database).

Appendix 4: k-means clustering, k=2, to identity left and right lungs

```
idx = kmeans(M(:,1:3),k,'Start', ...
    [mean(M(:,1))mean(M(:,2))-2*std(M(:,2)) mean(M(:,3)) ;...
    mean(M(:,1)) mean(M(:,2))+2*std(M(:,2)) mean(M(:,3))]);
```

Where k=2, and M is an $m \times 5$ matrix with columns [x y z eHU iHU].

Appendix 5: Inter-lung feature extraction

```
tmp = knnsearch(N,M,'K',k);
for i = 1 : n
    SOsum = 0;
    for j = 1 : k
        SOsum = SOsum + BRC(tmp(i,j),4);
    end
    RES(i) = SOsum/k; %store SO score for this voxel
end
```

N is the set of simulated tree node coordinates, M is the set of voxel coordinates, and BRC the branch matrix with SO in the 4th column:

Appendix 6: Calculating branch Δ HU from voxels proximal to leaf node (v)

```
ep2 = ep*ep; % squared distance
sz = size(C); % obtain dimensions of C
xlim = sz(1); ylim = sz(2); zlim = sz(3); % limits on size of cell arr
fx = fs{1}; fy = fs{2}; fz = fs{3};
centr = [fx(v(1)) fy(v(2)) fz(v(3))]; % center cell ID to check
bag = []; % list to build for averaging

for x = -1 : 1
    for y = -1 : 1
        for z = -1 : 1
            c = [centr(1)+x centr(2)+y centr(3)+z]; % cell to chk
            if c(1)>0 && c(2)>0 && c(3)>0 ...
                && c(1)<xlim && c(2)<ylim && c(3)<zlim %is valid cell ref?

                toChk = C{c(1),c(2),c(3)}; %obtain cell

                if ~(isempty(toChk))
                    m = length(toChk(:,1));
                    for i = 1 : m
                        if d22(v,toChk(i,1:3)) < ep2
                            bag = [bag toChk(i,4)];
                        end
                    end
                end
            end
        end
    end
end
```

```

if isempty(bag)
    out = 0; % point outside proximity of CT point cloud
else
    out = nanmean(bag); % return mean of points within ep of v
end

```

N.B. dotted line abbreviates for a series of **end** statements closing all of the above.

Code for efficiently identifying proximal points to average over for ΔHU ; for a terminal node with positional vector \mathbf{v} , cell array object \mathbf{C} as described, coordinate mapping function set \mathbf{fs} , and distance \mathbf{ep} (i.e. radius r of sphere) short for epsilon (Greek letter). N.B. r^2 is calculated to avoid expensive square root operation in distance checking. Essentially the ‘bag’ variable collects points falling within the sphere and returns either the average ΔHU of these points, or, if it happens there are no points within r of this terminal node (assumed to be very rare if it occurs at all), the algorithm implemented defaults to returning $\overline{\Delta\text{HU}} = 0$.

Appendix 7: Statistical group comparison algorithm with normality testing

```
if kstest((1/nanstd(M(:,index(i))))*(M(:,index(i))- ...
nanmean(M(:,index(i)))*ones(52,1))) % 1 => non-normal
    if s(2) == 2 % man whitney U test
        p = ranksum(G{1}(:,i),G{2}(:,i));
        TR(i) = p*ones(size(TR(i))); % p-value
    else
        [p,~,stats] = kruskalwallis(dat,ind,'off');
        TR(i) = p; % p-value
        c = multcompare(stats);
    end
else % normal distr. => use ANOVA
    if s(2) == 2 % t test
        [h,p] = ttest2(G{1}(:,i),G{2}(:,i));
        TR(i) = p*ones(size(TR(i))); % p-value
    else
        [p,~,stats] = anova1(dat,ind,'off');

        TR(i) = p; % p-value
        c = multcompare(stats);
    end
end
```

M is a data matrix of subjects (rows) and features (columns), in this study asthmatics / healthy volunteers and attributes from described methods, TR is the variable ‘to return’, and functions ‘nanmean.m’ and ‘nanstd.m’ calculate mean and standard deviation ignoring missing data (NaN values), to allow inclusion of analysis of incomplete features such as sputum cell counts (not all subjects could produce a sufficient sample at the time of testing). G is a cell array of non-intersecting subsets of M defining the groups being tested. The c variable carries forward statistical outcome of pairwise group comparisons in the case of $n > 2$ groupings.

Appendix 8: SAA interval/decile-wise statistical analysis algorithm

```
for j = 1 : n
    [p,h] = ranksum(L(:,j),U(:,j)); % p-value
    y1avg(j) = mean(L(:,j)); y1std(j) = std(L(:,j));
    y2avg(j) = mean(U(:,j)); y2std(j) = std(U(:,j));
end
plotCompGraph(y1avg,y1std,y2avg,y2std) % plot visualization
```

U is a set of subjects drawn from upper tertile of a selected feature; L from the lower.

Appendix 9: LDA implementation with Tikhonov regularization

```
M = zscore(M); % zero mean and unit variance for all features
c1 = M(1:g,attr); c2 = M(g+1:end,attr);
S1 = g*cov(c1); S2 = (length(M(:,1))-g)*cov(c2);
Sw = S1+S2;
cnt = 0;
while true
    cnt = cnt + 1;
    if rank(Sw) == length(attr)
        break; % okay to proceed
    else % Tikhonov regularization
        Sw = Sw + (0.01*rand(size(Sw))).*eye(size(Sw));
    end
    if cnt > 10
        error('failed to un-singularize matrix')
    end
end

SwInv = inv(Sw); mu1 = mean(c1); mu2 = mean(c2);
v = SwInv*(mu1-mu2)'; % directional vector for projection
Y1 = v'*c1'; Y2 = v'*c2';
h1 = histogram(Y1); hold on; h2 = histogram(Y2);
```

(supporting text overleaf)

Function ‘zscore.m’ maps input data by column to zscore value, i.e. zero mean and unit variance, ‘cov.m’ is a built-in function calculating the covariance matrix, ‘rank.m’ calculates the rank of a matrix, and ‘inv.m’ calculates matrix inverse. M is the data matrix of subjects and features. N.B. a matrix must be non-singular to be invertible, thus a requirement to calculating the projection vector \mathbf{p} as defined above is that the matrix S_W be non-singular. The purpose of the while loop is to check this property, and in the case of singularity apply Tikhonov regularization, a small random perturbation of the diagonal of S_W , ‘rand.m’ being a built-in random number generator (RNG) algorithm and ‘eye.m’ returning the required identity matrix here, intended to shift a singular matrix into a non-singular state with minimal disruption to the data; note a scaling factor of 0.01 (1%) applied to the RNG vector.

Appendix 10: Branch Δ HU visualisation across selected SO range

```
for i = 1 : n % iterate over all branches
    dHU = BRC(i,6);
    zsc = (dHU-xmean)/xstd;
    if i > 1
        if zsc >= 0
            nodFC{i} = 'green';
        end
        if zsc < 0
            nodFC{i} = 'red';
        end
    end
    plotSz(i) = 155 + zsc*50; % record size
end

figure
hold on
for i = 1 : n % plotting branches
    if ismember(BRC(i,5),SOs) % check if branch to plot
        proxID = BRC(i,1); distID = BRC(i,2);
        a = NOD(proxID,:); b = NOD(distID,:);
        SO = BRC(i,5);
        for j = 1 : length(SOs) % find SO index in SOs list
            if SOs(j)==SO
                break
            end
        end
        if brcPlotFlag
            plot3([a(1);b(1)],[a(2);b(2)],[a(3);b(3)],brcChar{j}, ...
                'linewidth',brcWid{j})
        end
        nodMarkSf = 0.04; % internal scale factor
        plot3(b(1),b(2),b(3),nodChar{j},'MarkerSize', ...
            nodMarkSf*plotSz(i),'MarkerFaceColor',nodFC{i})
    end
end

xlabel('X ~ anterior-posterior')
ylabel('Y ~ left-right')
zlabel('Z ~ inferior-superior')
```

The first loop assigns colours green and red to ΔHU above and below the mean ΔHU respectively, mean taken over all branches selected for plotting (through SO specification). Size of circular markers to plot at distal branch points is set at 155 points (1 point = 1/72 inches) + 50z where z is the z-score of ΔHU for that branch, i.e. increases or decreases final size in direct proportion to ΔHU . Proximal and distal nodes are identified for all branches to plot (having SO within the given SO list SOs), and used to determine positional vectors for plotting with plot3.m, where a disc is placed at the **distal** branch point, with colour and size determined as described.

Appendix 11: Airway radii adjustment factor calculation from ΔHU

```
%pre-processing
k=3;
dhuM=mean(dHU); stdM=std(dHU);
for j=1:length(dHU)
    if dHU(j)<dhuM-k*stdM
        dHU(j)=dhuM+k*stdM;
    end
    if dHU(j)>dhuM+k*stdM
        dHU(j)=dhuM+k*stdM;
    end
end

%rescaling radii
for j=1:length(SO_list)
    scaling = dHU(SO==SO_list(j));
    maxHU = max(scaling);
    minHU = min(scaling);
    scaling=scaling-maxHU*ones(size(scaling));
    scaling=scaling*(1/(maxHU-minHU));
    scaling=w*scaling;
    scaling=scaling+ones(size(scaling));
    radii(SO==SO_list(j))=radii(SO==SO_list(j)).* ...
        scaling;
end
```

Stepwise calculation of radii adjustment factor f_w performed on the ‘scaling’ variable.

Appendix 12: Branch Δ HU visualisation across selected SO range

```

GHpair = [1.22 5.8]; %FOT parameters G and H
sc=10^(-3); %airway scaling (conversion to m)
BRCws = [0 0.2:0.025:0.6]; %weights
SO_LIST = {1,1:2,1:4,1:6,1:8,1:10}; %Strahler order ranges
nSO = length(GHpair(:,1)); nBC = length(BRCws);
FOT_RES = cell(3,nSO,nBC); %cell array to store results
n=length(MEAS(:,1)); %number of subjects
for i=1:nSO %iterate SO
    for j=1:nBC %iterate weights
        chk=memory;
        D2=brcRescaler(D,SO_LIST{1},BRCws(j)); %rescale radii
        FOT_RES{2,i,j}=zeros(n,6);
        for x=1:n %run FOT code over all subjects
            FOT_RES{2,i,j}(x,:)=FOT_sim( ...
                [D2{x,4}{1}(:,1:3) sc*D2{x,4}{1}(:,4)],...
                D2{x,4}{6},D2{x,4}{2},5:1:20,...
                GHpair(i,1),GHpair(i,2));
        end
        X=[MEAS(:,1),MEAS(:,2),...
            MEAS(:,1)-MEAS(:,2)]; %measured values
        Y=[FOT_RES{2,i,j}(:,1),FOT_RES{2,i,j}(:,2),...
            FOT_RES{2,i,j}(:,1)-FOT_RES{2,i,j}(:,2)]; %simulated values
        FOT_RES{3,i,j}=zeros(3,4); %store R^2, corr. coef. and pval
        for x=1:3
            mdl = fitlm(X(:,x),Y(:,x));
            Rs=mdl.Rsquared;
            [R,P]=corrcoef(X(:,x),Y(:,x));
            FOT_RES{3,i,j}(x,:)= [Rs.Ordinary,Rs.Adjusted,...
                R(1,2),P(1,2)];
        end
    end
end
end
end

```

(supporting text overleaf)

GHpair is a vector specifying parameters for tissue dampening and elastance (see section 2.4.3), airways are initially rescaled by 10^{-3} as the model records diameters in mm, and the FOT simulation script, 'FOT_sim' in the above, was designed using Système international (SI) units, i.e. meters for length. 'brcRescaler' effects the described radii rescaling operation. FOT_RES is a cell array for storing results from experiments. Built in functions corrcoef.m and fitlm.m are used for calculation of correlation coefficients and fitting of a linear model respectively. Correlation coefficient (r-value), associated p-value, and both ordinary and adjusted R-squared from linear modelling are recorded.

6 References

- 1) Siddiqui S, Brightling CE. Airways Disease: Phenotyping Heterogeneity Using Measures of Airway Inflammation. *Allergy Asthma Clin Immunol*. 2007 Jun 15;3(2):60.
- 2) Carr TF, Bleecker E. Asthma heterogeneity and severity. *World Allergy Organ J*. 2016 Dec;9(1):41.
- 3) Asthma statistics | British Lung Foundation [Internet]. Available from: <https://statistics.blf.org.uk/asthma>
- 4) Masoli M, Fabian D, Holt S, Beasley R, Global Initiative for Asthma (GINA) Program. The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy*. 2004 May;59(5):469–78.
- 5) Akinbami LJ, Moorman JE, Bailey C, Zahran HS, King M, Johnson CA, et al. Trends in asthma prevalence, health care use, and mortality in the United States, 2001-2010. *NCHS Data Brief*. 2012 May;(94):1–8.
- 6) Burr ML, Butland BK, King S, Vaughan-Williams E. Changes in asthma prevalence: two surveys 15 years apart. *Arch Dis Child*. 1989 Oct 1;64(10):1452–6.
- 7) Soriano JB, Kiri VA, Maier WC, Strachan D. Increasing prevalence of asthma in UK primary care during the 1990s. *Int J Tuberc Lung Dis*. 2003 May 1;7(5):415–21.
- 8) Mukherjee M, Stoddart A, Gupta RP, Nwaru BI, Farr A, Heaven M, et al. The epidemiology, healthcare and societal burden and costs of asthma in the UK and its member nations: analyses of standalone and linked national databases. *BMC Med*. 2016 Aug 29;14(1):113.
- 9) Siddiqui S, Gonem S, Wardlaw AJ. Advances in the Management of Severe Asthma. *Semin Respir Crit Care Med*. 2012 Dec;33(6):666–84.

- 10) Murray CJL, Lopez AD. Measuring the Global Burden of Disease. *N Engl J Med*. 2013 Aug 1;369(5):448–57.
- 11) Humphreys E, Upton J, Walker S, Price D. Quality of life and asthma control: A UK population survey. *Eur Respir J*. 2014 Sep 1;44(Suppl 58):P4090.
- 12) Lloyd A, Price D, Brown R. The impact of asthma exacerbations on health-related quality of life in moderate to severe asthma patients in the UK. *Prim Care Respir J*. 2007 Feb 3;16(1):22–7.
- 13) Schmier JK, Manjunath R, Halpern MT, Jones ML, Thompson K, Diette GB. The impact of inadequately controlled asthma in urban children on quality of life and productivity. *Ann Allergy Asthma Immunol*. 2007 Mar 1;98(3):245–51.
- 14) Accordini S, Bugiani M, Arossa W, Gerzeli S, Marinoni A, Olivieri M, et al. Poor Control Increases the Economic Cost of Asthma. *Int Arch Allergy Immunol*. 2006;141(2):189–98.
- 15) Bahadori K, Doyle-Waters MM, Marra C, Lynd L, Alasaly K, Swiston J, et al. Economic burden of asthma: a systematic review. *BMC Pulm Med*. 2009 May 19;9(1):24.
- 16) Teague WG, Tustison NJ, Altes TA. Ventilation heterogeneity in asthma. *J Asthma*. 2014 Sep 1;51(7):677–84.
- 17) Farah CS, King GG, Brown NJ, Peters MJ, Berend N, Salome CM. Ventilation heterogeneity predicts asthma control in adults following inhaled corticosteroid dose titration. *J Allergy Clin Immunol*. 2012 Jul 1;130(1):61–8.
- 18) Downie SR, Salome CM, Verbanck S, Thompson B, Berend N, King GG. Ventilation heterogeneity is a major determinant of airway hyperresponsiveness in asthma, independent of airway inflammation. *Thorax*. 2007 Aug 1;62(8):684–9.
- 19) Paiva M, Engel LA. Theoretical studies of gas mixing and ventilation distribution in the lung. *Physiol Rev*. 1987 Jul;67(3):750–96.

- 20) Robinson PD, Latzin P, Verbanck S, Hall GL, Horsley A, Gappa M, et al. Consensus statement for inert gas washout measurement using multiple- and single- breath tests. *Eur Respir J*. 2013 Mar 1;41(3):507–22.
- 21) Bickel S, Popler J, Lesnick B, Eid N. Impulse oscillometry: interpretation and practical applications. *Chest*. 2014 Sep;146(3):841–7.
- 22) Desiraju K, Agrawal A. Impulse oscillometry: The state-of-art for lung function testing. *Lung India Off Organ Indian Chest Soc*. 2016;33(4):410–6.
- 23) Galant SP, Komarow HD, Shin H-W, Siddiqui S, Lipworth BJ. The case for impulse oscillometry in the management of asthma in children and adults. *Ann Allergy Asthma Immunol*. 2017 Jun 1;118(6):664–71.
- 24) Soares M, Bordas R, Thorpe J, Timmerman B, Brightling C, Kay D, et al. Validation of impulse oscillometry R5-R20 as a small airways dysfunction detection tool in adult asthma. *Eur Respir J*. 2016 Sep 1;48(suppl 60):OA4968.
- 25) Walker C, Gupta S, Raj V, Siddiqui S, Brightling CE. Imaging advances in asthma. *Expert Opin Med Diagn*. 2011 Sep;5(5):453–65.
- 26) Ash SY, Diaz AA. The role of imaging in the assessment of severe asthma. *Curr Opin Pulm Med*. 2017;23(1):97–102.
- 27) Tzeng Y-S, Lutchen K, Albert M. The difference in ventilation heterogeneity between asthmatic and healthy subjects quantified using hyperpolarized ³He MRI. *J Appl Physiol*. 2009 Mar 1;106(3):813–22.
- 28) Niimi A, Matsumoto H, Amitani R, Nakano Y, Mishima M, Minakuchi M, et al. Airway Wall Thickness in Asthma Assessed by Computed Tomography. *Am J Respir Crit Care Med*. 2000 Oct 1;162(4):1518–23.
- 29) Palágyi K, Tschirren J, Hoffman EA, Sonka M. Quantitative analysis of pulmonary airway tree structures. *Comput Biol Med*. 2006 Sep 1;36(9):974–96.
- 30) Lee SM, Seo JB, Lee SM, Kim N, Oh SY, Oh Y-M. Optimal threshold of subtraction method for quantification of air-trapping on coregistered CT in COPD patients. *Eur Radiol*. 2016 Jul 1;26(7):2184–92.

- 31) Galbán CJ, Chenevert TL, Meyer CR, Tsien C, Lawrence TS, Hamstra DA, et al. The parametric response map is an imaging biomarker for early cancer treatment outcome. *Nat Med*. 2009 May;15(5):572–6.
- 32) Galbán CJ, Han MK, Boes JL, Chughtai KA, Meyer CR, Johnson TD, et al. Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat Med*. 2012 Nov;18(11):1711–5.
- 33) Tahir BA, Van Holsbeke C, Ireland RH, Swift AJ, Horn FC, Marshall H, et al. Comparison of CT-based Lobar Ventilation with 3He MR Imaging Ventilation Measurements. *Radiology*. 2015 Aug 28;278(2):585–92.
- 34) Pu J, Gu S, Liu S, Zhu S, Wilson D, Siegfried JM, et al. CT based computerized identification and analysis of human airways: A review. *Med Phys*. 2012;39(5):2603–16.
- 35) van Ginneken B, Baggerman W, van Rikxoort EM. Robust Segmentation and Anatomical Labeling of the Airway Tree from Thoracic CT Scans. In: Metaxas D, Axel L, Fichtinger G, Székely G, editors. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*. Springer Berlin Heidelberg; 2008. p. 219–26. (Lecture Notes in Computer Science).
- 36) Reynisson PJ, Scali M, Smistad E, Hofstad EF, Leira HO, Lindseth F, et al. Airway Segmentation and Centerline Extraction from Thoracic CT – Comparison of a New Method to State of the Art Commercialized Methods. *PLOS ONE*. 2015 Dec 11;10(12):e0144282.
- 37) Bordas R, Lefevre C, Veeckmans B, Pitt-Francis J, Fetita C, Brightling CE, et al. Development and Analysis of Patient-Based Complete Conducting Airways Models. *PLOS ONE*. 2015 Dec 11;10(12):e0144105.
- 38) Foy BH, Soares M, Bordas R, Richardson M, Bell A, Singapuri A, Hargadon B, Brightling C, Burrowes K, Kay D, Owers-Bradley J. Lung computational models and the role of the small airways in asthma. *Am J Respir Crit Care Med*. 2019 May 18(ja).

- 39) Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, et al. Cluster Analysis and Clinical Asthma Phenotypes. *Am J Respir Crit Care Med*. 2008 Aug 1;178(3):218–24.
- 40) Hilvering B, Vijverberg S, Houben L, Schweizer R, Lammers J-W, Koenderman L. The identification of asthma phenotypes by categorical PCA: Combinatorial analysis of clinical parameters and dysfunctional blood eosinophils. *Eur Respir J*. 2014 Sep 1;44(Suppl 58):P3006.
- 41) Bousquet J, Anto JM, Sterk PJ, Adcock IM, Chung KF, Roca J, et al. Systems medicine and integrated care to combat chronic noncommunicable diseases. *Genome Med*. 2011 Jul 6;3(7):43.
- 42) Hood L, Flores M. A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. *New Biotechnol*. 2012 Sep 15;29(6):613–24.
- 43) Wu X, Zhu X, Wu G, Ding W. Data mining with big data. *IEEE Trans Knowl Data Eng*. 2014 Jan;26(1):97–107.
- 44) Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med*. 2016 Sep 29;375(13):1216–9.
- 45) Verleysen M, François D. The Curse of Dimensionality in Data Mining and Time Series Prediction. In: Cabestany J, Prieto A, Sandoval F, editors. *Computational Intelligence and Bioinspired Systems*. Springer Berlin Heidelberg; 2005. p. 758–70. (Lecture Notes in Computer Science).
- 46) Hubert M, Engelen S. Robust PCA and classification in biosciences. *Bioinformatics*. 2004 Jul 22;20(11):1728–36.
- 47) Geiger BC, Kubin G. Relative information loss in the PCA. In: 2012 IEEE Information Theory Workshop. 2012. p. 562–6.
- 48) Carlsson G. Topology and data. *Bull Am Math Soc*. 2009;46(2):255–308.
- 49) Carlsson G. Topological pattern recognition for point cloud data. *Acta Numerica*. 2014 May;23:289–368.

- 50) Li L, Cheng W-Y, Glicksberg BS, Gottesman O, Tamler R, Chen R, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci Transl Med*. 2015 Oct 28;7(311):311ra174-311ra174.
- 51) Nicolau M, Levine AJ, Carlsson G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proc Natl Acad Sci*. 2011 Apr 26;108(17):7265–70.
- 52) Soares M, Mirgorodskaya E, Koca H, Viklund E, Richardson M, Gustafsson P, et al. Particles in exhaled air (PExA): non-invasive phenotyping of small airways disease in adult asthma. *J Breath Res*. 2018 Sep;12(4):046012.
- 53) Bateman ED, Hurd SS, Barnes PJ, Bousquet J, Drazen JM, FitzGerald M, et al. Global strategy for asthma management and prevention: GINA executive summary. *Eur Respir J*. 2008 Jan 1;31(1):143–78.
- 54) Peters SP. Heterogeneity in the pathology and treatment of asthma. *Am J Med*. 2003 Aug 18;115(3, Supplement 1):49–54.
- 55) Lima JJ. Treatment Heterogeneity in Asthma. *Mol Diagn Ther*. 2007 Mar 1;11(2):97–104.
- 56) Koshak EA. Classification of asthma according to revised 2006 GINA: Evolution from severity to control. *Ann Thorac Med*. 2007;2(2):45–6.
- 57) Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma—Summary Report 2007. *J Allergy Clin Immunol*. 2007 Nov 1;120(5):S94–138.
- 58) Antonicelli L, Bucca C, Neri M, Benedetto FD, Sabbatani P, Bonifazi F, et al. Asthma severity and medical resource utilisation. *Eur Respir J*. 2004 May 1;23(5):723–9.
- 59) Lusuardi M, Lucioni C, De Benedetto F, Mazzi S, Sanguinetti CM, Donner CF. GOLD severity stratification and risk of hospitalisation for COPD exacerbations. *Monaldi Arch Chest Dis*. 2008;69(4).

- 60) Postma DS, Rabe KF. The Asthma–COPD Overlap Syndrome. *N Engl J Med*. 2015 Sep 24;373(13):1241–9.
- 61) Holgate ST. The epidemic of allergy and asthma. *Nature*. 1999 Nov 25;402:2–4.
- 62) Pekkanen J, Lampi J, Genuneit J, Hartikainen A-L, Järvelin M-R. Analyzing atopic and non-atopic asthma. *Eur J Epidemiol*. 2012 Apr 1;27(4):281–6.
- 63) Gould HJ, Sutton BJ. IgE in allergy and asthma today. *Nat Rev Immunol*. 2008 Mar;8(3):205–17.
- 64) Ali FR. Does this patient have atopic asthma? *Clin Med*. 2011 Jan 8;11(4):376–80.
- 65) Pearce N, Pekkanen J, Beasley R. How much asthma is really attributable to atopy? *Thorax*. 1999 Mar 1;54(3):268–72.
- 66) O’Byrne PM. Airway inflammation and asthma. *Aliment Pharmacol Ther*. 1996;10(Sup2):18–24.
- 67) Bousquet J, Chané P, Lacoste JY, Barnéon G, Ghavanian N, Enander I, et al. Eosinophilic Inflammation in Asthma. *N Engl J Med*. 1990 Oct 11;323(15):1033–9.
- 68) Fahy JV. Eosinophilic and Neutrophilic Inflammation in Asthma. *Proc Am Thorac Soc*. 2009 May 1;6(3):256–9.
- 69) George L, Brightling CE. Eosinophilic airway inflammation: role in asthma and chronic obstructive pulmonary disease. *Ther Adv Chronic Dis*. 2016 Jan 1;7(1):34–51.
- 70) Haldar P, Brightling CE, Hargadon B, Gupta S, Monteiro W, Sousa A, et al. Mepolizumab and Exacerbations of Refractory Eosinophilic Asthma. *N Engl J Med*. 2009 Mar 5;360(10):973–84.
- 71) Hoenderdos K, Condcliffe A. The neutrophil in chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol*. 2013 May;48(5):531–9.

- 72) Juniper EF, Guyatt GH, Epstein RS, Ferrie PJ, Jaeschke R, Hiller TK. Evaluation of impairment of health related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax*. 1992 Feb;47(2):76–83.
- 73) Juniper EF, O'byrne PM, Guyatt G h, Ferrie P j, King D r. Development and validation of a questionnaire to measure asthma control. *Eur Respir J*. 1999;14(4):902–7.
- 74) Juniper EF, Buist AS, Cox FM, Ferrie PJ, King DR. Validation of a Standardized Version of the Asthma Quality of Life Questionnaire. *Chest*. 1999 May 1;115(5):1265–70.
- 75) Juniper EF, Svensson K, Mörk A-C, Ståhl E. Measurement properties and interpretation of three shortened versions of the asthma control questionnaire. *Respir Med*. 2005 May 1;99(5):553–8.
- 76) Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. *Nat Rev Genet*. 2009 Mar;10(3):184–94.
- 77) McDougall CM, Blaylock MG, Douglas JG, Brooker RJ, Helms PJ, Walsh GM. Nasal Epithelial Cells as Surrogates for Bronchial Epithelial Cells in Airway Inflammation Studies. *Am J Respir Cell Mol Biol*. 2008 Nov 1;39(5):560–8.
- 78) Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol*. 1996 Dec;14(13):1675–80.
- 79) Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003 Feb 15;31(4):e15–e15.
- 80) Weibel ER. What makes a good lung? *Swiss Med Wkly*. 2009 Jul 11;139(27–28):375–86.

- 81) Horton RE. Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geol Soc Am Bull.* 1945 Mar 1;56(3):275-370.
- 82) Strahler AN. Hypsometric (area-altitude) analysis of erosional topography. *Geol Soc Am Bull.* 1952 Nov 1;63(11):1117-42.
- 83) Petersson J, Glenny RW. Gas exchange and ventilation–perfusion relationships in the lung. *Eur Respir J.* 2014 Oct 1;44(4):1023–41.
- 84) Siddiqui SH. Structure function relationships in adult asthma (Doctoral dissertation, University of Leicester). Available from: <http://europepmc.org/abstract/eth/509326>
- 85) Brightling CE, Gupta S, Gonem S, Siddiqui S. Lung damage and airway remodelling in severe asthma. *Clin Exp Allergy.* 2012;42(5):638–49.
- 86) Bryan AC, Milic-Emili J, Pengelly D. Effect of gravity on the distribution of pulmonary ventilation. *J Appl Physiol.* 1966 May 1;21(3):778–84.
- 87) Hopkins SR, Henderson AC, Levin DL, Yamada K, Arai T, Buxton RB, et al. Vertical gradients in regional lung density and perfusion in the supine human lung: the Slinky effect. *J Appl Physiol Bethesda Md* 1985. 2007 Jul;103(1):240–8.
- 88) Lai-Fook SJ, Rodarte JR. Pleural pressure distribution and its relationship to lung volume and interstitial pressure. *J Appl Physiol.* 1991 Mar 1;70(3):967–78.
- 89) Galvin I, Drummond GB, Nirmalan M. Distribution of blood flow and ventilation in the lung: gravity is not the only factor. *BJA Br J Anaesth.* 2007 Apr 1;98(4):420–8.
- 90) Tawhai MH, Nash MP, Lin C-L, Hoffman EA. Supine and prone differences in regional lung density and pleural pressure gradients in the human lung with constant shape. *J Appl Physiol.* 2009 Sep;107(3):912–20.

- 91) Henderson AC, Sá RC, Theilmann RJ, Buxton RB, Prisk GK, Hopkins SR. The gravitational distribution of ventilation-perfusion ratio is more uniform in prone than supine posture in the normal human lung. *J Appl Physiol Bethesda Md* 1985. 2013 Aug 1;115(3):313–24.
- 92) Crawford AB, Makowska M, Paiva M, Engel LA. Convection- and diffusion-dependent ventilation maldistribution in normal subjects. *J Appl Physiol Bethesda Md* 1985. 1985 Sep;59(3):838–46.
- 93) Berend N, Salome CM, King GG. Mechanisms of airway hyperresponsiveness in asthma. *Respirology*. 2008;13(5):624–31.
- 94) Verbanck S, Thompson BR, Schuermans D, Kalsi H, Biddiscombe M, Stuart-Andrews C, et al. Ventilation heterogeneity in the acinar and conductive zones of the normal ageing lung. *Thorax*. 2012 Sep 1;67(9):789–95.
- 95) Jarenbäck L, Ankerst J, Björmer L, Tufvesson E. Acinar ventilation heterogeneity in COPD relates to diffusion capacity, resistance and reactance. *Respir Med*. 2016 Jan 1;110:28–33.
- 96) Hardaker KM, Downie SR, Kermode JA, Berend N, King GG, Salome CM. Ventilation heterogeneity is associated with airway responsiveness in asthma but not COPD. *Respir Physiol Neurobiol*. 2013 Oct 1;189(1):106–11.
- 97) Engel LA, Landau L, Taussig L, Martin RR, Sybrecht G. Influence of bronchomotor tone on regional ventilation distribution at residual volume. *J Appl Physiol*. 1976 Mar 1;40(3):411–6.
- 98) Bryan AC, Bentivoglio LG, Beerel F, MacLeish H, Zidulka A, Bates DV. Factors affecting regional distribution of ventilation and perfusion in the lung. *J Appl Physiol*. 1964 May 1;19(3):395–402.
- 99) Milic-Emili J, Henderson JA, Dolovich MB, Trop D, Kaneko K. Regional distribution of inspired gas in the lung. *J Appl Physiol*. 1966 May 1;21(3):749–59.
- 100) James AL, Paré PD, Hogg JC. The Mechanics of Airway Narrowing in Asthma. *Am Rev Respir Dis*. 1989 Jan 1;139(1):242–6.

- 101) Elias JA, Zhu Z, Chupp G, Homer RJ. Airway remodeling in asthma. *J Clin Invest.* 1999 Oct 15;104(8):1001–6.
- 102) Redington AE, Howarth PH. Airway wall remodelling in asthma. *Thorax.* 1997 Apr;52(4):310–2.
- 103) Svenningsen S, Nair P, Guo F, McCormack DG, Parraga G. Is ventilation heterogeneity related to asthma control? *Eur Respir J.* 2016 Aug 1;48(2):370–9.
- 104) Criée CP, Sorichter S, Smith HJ, Kardos P, Merget R, Heise D, et al. Body plethysmography – Its principles and clinical use. *Respir Med.* 2011 Jul 1;105(7):959–71.
- 105) DuBois AB, Botelho SY, Comroe JH. A new method for measuring airway resistance in man using a body plethysmograph: values in normal subjects and in patients with respiratory disease. *J Clin Invest.* 1956 Mar 1;35(3):327–35.
- 106) Coates AL, Peslin R, Rodenstein D, Stocks J. Measurement of lung volumes by plethysmography. *Eur Respir J.* 1997 Jun 1;10(6):1415–27.
- 107) Mead J. Volume displacement body plethysmograph for respiratory measurements in human subjects. *J Appl Physiol.* 1960 Jul 1;15(4):736–40.
- 108) Jaeger MJ, Otis AB. Measurement of airway resistance with a volume displacement body plethysmograph. *J Appl Physiol.* 1964 Jul 1;19(4):813–9.
- 109) Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. *Eur Respir J.* 2005 Aug 1;26(2):319–38.
- 110) Pierce R. Spirometry: an essential clinical measurement. *Aust Fam Physician.* 2005 Jul;34(7):535.
- 111) Moore VC. Spirometry: step by step. *Breathe.* 2012 Mar 1;8(3):232–40.
- 112) Hyatt RE. Expiratory flow limitation. *J Appl Physiol.* 1983 Jul 1;55(1):1–7.
- 113) Wilson TA, Hyatt RE, Rodarte JR. The mechanisms that limit expiratory flow. *Lung.* 1980 Dec 1;158(1):193–200.

- 114) Dawson SV, Elliott EA. Wave-speed limitation on expiratory flow-a unifying concept. *J Appl Physiol*. 1977 Sep 1;43(3):498–515.
- 115) Hansen JE, Sun X-G, Wasserman K. Spirometric Criteria for Airway Obstruction: Use Percentage of FEV1/FVC Ratio Below the Fifth Percentile, Not < 70%. *Chest*. 2007 Feb 1;131(2):349–55.
- 116) Robertson JS, Siri WE, Jones HB. Lung ventilation patterns determined by analysis of nitrogen elimination rates; use of mass spectrometer as a continuous gas analyzer. *J Clin Invest*. 1950 May;29(5):577–90.
- 117) Newth CJ, Enright P, Johnson RL. Multiple-breath nitrogen washout techniques: including measurements with patients on ventilators. *Eur Respir J*. 1997 Sep 1;10(9):2174–85.
- 118) Jonmarker C, Jansson L, Jonson B, Larsson A, Werner O. Measurement of functional residual capacity by sulfur hexafluoride washout. *Anesthesiology*. 1985 Jul;63(1):89–95.
- 119) Guglani L, Kasi A, Starks M, Pedersen KE, Nielsen JG, Weiner DJ. Difference between SF6 and N2 Multiple Breath Washout kinetics is due to N2 back diffusion and error in N2 offset. *J Appl Physiol Bethesda Md* 1985. 2018 26;
- 120) Robinson PD, Goldman MD, Gustafsson PM. Inert Gas Washout: Theoretical Background and Clinical Utility in Respiratory Disease. *Respiration*. 2009;78(3):339–55.
- 121) Stuart-Andrews CR, Kelly VJ, Sands SA, Lewis AJ, Ellis MJ, Thompson BR. Automated detection of the phase III slope during inert gas washout testing. *J Appl Physiol*. 2011 Dec 15;112(6):1073–81.
- 122) Bouhuys A, Lichtneckert S, Lundgren C, Lundin G. Voluntary changes in breathing pattern and N2 clearance from lungs. *J Appl Physiol*. 1961 Nov;16:1039–42.

- 123) Verbanck S, Schuermans D, Van Muylem A, Paiva M, Noppen M, Vincken W. Ventilation distribution during histamine provocation. *J Appl Physiol*. 1997 Dec 1;83(6):1907–16.
- 124) Verbanck S, Schuermans D, Van Muylem A, Melot C, Noppen M, Vincken W, et al. Conductive and Acinar Lung-zone Contributions to Ventilation Inhomogeneity in COPD. *Am J Respir Crit Care Med*. 1998 May 1;157(5):1573–7.
- 125) Horáček J, Koucký V, Hladík M. Contribution of Interval Linear Algebra to the Ongoing Discussions on Multiple Breath Washout Test. *arXiv preprint arXiv:1902.09026*. 2019 Feb 24.
- 126) Hasler D, Anagnostopoulou P, Nyilas S, Latzin P, Schittny J, Obrist D. A multi-scale model of gas transport in the lung to study heterogeneous lung ventilation during the multiple-breath washout test. *PLOS Comput Biol*. 2019 Jun 17;15(6):e1007079.
- 127) DuBois AB, Brody AW, Lewis DH, Burgess BF. Oscillation Mechanics of Lungs and Chest in Man. *J Appl Physiol*. 1956 May 1;8(6):587–94.
- 128) Michaelson ED, Grassman ED, Peters WR. Pulmonary mechanics by spectral analysis of forced random noise. *J Clin Invest*. 1975 Nov 1;56(5):1210–30.
- 129) Smith HJ, Reinhold P, Goldman. Forced oscillation technique and impulse oscillometry. In 2005.
- 130) Olaguíbel JM, Alvarez-Puebla MJ, Anda M, Gómez B, García BE, Tabar AI, et al. Comparative analysis of the bronchodilator response measured by impulse oscillometry (IOS), spirometry and body plethysmography in asthmatic children. *J Investig Allergol Clin Immunol*. 2005;15(2):102–6.
- 131) Komarow HD, Myles IA, Uzzaman A, Metcalfe DD. Impulse oscillometry in the evaluation of diseases of the airways in children. *Ann Allergy Asthma Immunol*. 2011 Mar 1;106(3):191–9.

- 132) Naji N, Keung E, Kane J, Watson RM, Killian KJ, Gauvreau GM. Comparison of changes in lung function measured by plethymography and IOS after bronchoprovocation. *Respir Med*. 2013 Apr 1;107(4):503–10.
- 133) Oostveen E, MacLeod D, Lorino H, Farré R, Hantos Z, Desager K, et al. The forced oscillation technique in clinical practice: methodology, recommendations and future developments. *Eur Respir J*. 2003 Dec 1;22(6):1026–41.
- 134) Wei X, Shi Z, Cui Y, Mi J, Ma Z, Ren J, Li J, Xu S, Guo Y. Impulse oscillometry system as an alternative diagnostic method for chronic obstructive pulmonary disease. *Medicine*. 2017 Nov;96(46).
- 135) Gonem S, Natarajan S, Desai D, Corkill S, Singapuri A, Bradding P, et al. Clinical significance of small airway obstruction markers in patients with asthma. *Clin Exp Allergy*. 2014;44(4):499–507.
- 136) Filler AG. The history, development and impact of computed imaging in neurological diagnosis and neurosurgery: CT, MRI, and DTI. *Internet J Neurosurg*. 2010 May;7(1):1-69.
- 137) Mould RF. Röntgen and the discovery of X-rays. *Br J Radiol*. 1995 Nov 1;68(815):1145–76.
- 138) Brooks RA. A quantitative theory of the Hounsfield unit and its application to dual energy scanning. *J Comput Assist Tomogr*. 1977 Oct;1(4):487–93.
- 139) Herman GT. Fundamentals of computerized tomography: image reconstruction from projections. Springer Science & Business Media; 2009 Jul 14.
- 140) Mersereau RM, Oppenheim AV. Digital reconstruction of multidimensional signals from their projections. *Proc IEEE*. 1974 Oct;62(10):1319–38.
- 141) Dudgeon DE, Mersereau RM. Multidimensional digital signal processing. Prentice-Hall; 1984. 428 p.
- 142) British Thoracic Society Scottish Intercollegiate Guidelines Network. British Guideline on the Management of Asthma. *Thorax*. 2008 May;63 Suppl 4:iv1-121.

- 143) 2019 GINA Report: Global Strategy for Asthma Management and Prevention [Internet]. Global Initiative for Asthma - GINA. Available from: <https://ginasthma.org/reports/2019-gina-report-global-strategy-for-asthma-management-and-prevention/>
- 144) Paganin F, Trussard V, Seneterre E, Chanez P, Giron J, Godard P, et al. Chest radiography and high resolution computed tomography of the lungs in asthma. *Am Rev Respir Dis*. 1992 Oct;146(4):1084–7.
- 145) Lynch DA, Newell JD, Tschomper BA, Cink TM, Newman LS, Bethel R. Uncomplicated asthma in adults: comparison of CT appearance of the lungs in asthmatic and healthy subjects. *Radiology*. 1993 Sep;188(3):829–33.
- 146) Park CS, Müller NL, Worthy SA, Kim JS, Awadh N, Fitzgerald M. Airway obstruction in asthmatic and healthy individuals: inspiratory and expiratory thin-section CT findings. *Radiology*. 1997 May;203(2):361–7.
- 147) Khadadah M, Jayakrishnan B, Muquim A, Roberts O, Sinan T, Maradny N, et al. High resolution computed tomography in asthma. *Oman Med J*. 2012 Mar;27(2):145–50.
- 148) Gupta S, Siddiqui S, Haldar P, Raj JV, Entwisle JJ, Wardlaw AJ, et al. Qualitative analysis of high-resolution CT scans in severe asthma. *Chest*. 2009 Dec;136(6):1521–8.
- 149) Lo D, Maniyar A, Gupta S, Gaillard E. High prevalence of bronchiectasis on chest CT in a selected cohort of children with severe Asthma. *BMC Pulm Med*. 2019 Jul 26;19(1):136.
- 150) Yilmaz S, Ekici A, Ekici M, Keles H. High-resolution computed tomography findings in elderly patients with asthma. *Eur J Radiol*. 2006 Aug 1;59(2):238–43.
- 151) Laurent F, Latrabe V, Raheison C, Marthan R, Tunon-de-Lara JM. Functional significance of air trapping detected in moderate asthma. *Eur Radiol*. 2000 Aug 1;10(9):1404–10.

- 152) Niimi A, Matsumoto H, Amitani R, Nakano Y, Sakai H, Takemura M, et al.
Effect of short-term treatment with inhaled corticosteroid on airway wall thickening in asthma. *Am J Med.* 2004 Jun 1;116(11):725–31.
- 153) Kurashima K, Kanauchi T, Hoshi T, Takaku Y, Ishiguro T, Takayanagi N, et al.
Effect of early versus late intervention with inhaled corticosteroids on airway wall thickness in patients with asthma. *Respirology.* 2008;13(7):1008–13.
- 154) Goldin JG, Tashkin DP, Kleerup EC, Greaser LE, Haywood UM, Sayre JW, et al.
Comparative effects of hydrofluoroalkane and chlorofluorocarbon beclomethasone dipropionate inhalation on small airways: Assessment with functional helical thin-section computed tomography. *J Allergy Clin Immunol.* 1999 Dec 1;104(6):s258–67.
- 155) MRI: From Picture to Proton. *Radiology.* 2004 Aug 1;232(2):474–474.
- 156) Geva T. Magnetic Resonance Imaging: Historical Perspective. *J Cardiovasc Magn Reson.* 2006 Jan 1;8(4):573–80.
- 157) Bloch F, Hansen WW, Packard M. Nuclear Induction. *Phys Rev.* 1946 Feb 1;69(3–4):127–127.
- 158) Shampo MA, Kyle RA. Edward M. Purcell— Nobel Prize for Magnetic Resonance Imaging. *Mayo Clin Proc.* 1997 Jun 1;72(6):585.
- 159) Weishaupt D, Koechli VD, Marincek B. How does MRI work?: An Introduction to the Physics and Function of Magnetic Resonance Imaging. Springer Science & Business Media; 2008. 172 p.
- 160) Aronberg DJ, Glazer HS, Sagel SS. MRI and CT of the mediastinum: comparisons, controversies, and pitfalls. *Radiol Clin North Am.* 1985 Sep;23(3):439–48.
- 161) Bergin CJ, Pauly JM, Macovski A. Lung parenchyma: projection reconstruction MR imaging. *Radiology.* 1991 Jun;179(3):777–81.

- 162) Albert MS, Cates GD, Driehuys B, Happer W, Saam B, Springer CS, et al. Biological magnetic resonance imaging using laser-polarized ^{129}Xe . *Nature*. 1994 Jul;370(6486):199–201.
- 163) Kauczor H-U, Ebert M, Kreitner K-F, Nilgens H, Surkau R, Heil W, et al. Imaging of the lungs using ^3He MRI: Preliminary clinical experience in 18 patients with and without lung disease. *J Magn Reson Imaging*. 1997;7(3):538–43.
- 164) Fain S, Schiebler ML, McCormack DG, Parraga G. Imaging of lung function using hyperpolarized helium-3 magnetic resonance imaging: Review of current and emerging translational methods and applications. *J Magn Reson Imaging*. 2010;32(6):1398–408.
- 165) Qing K, Ruppert K, Jiang Y, Mata JF, Miller GW, Shim YM, et al. Regional mapping of gas uptake by blood and tissue in the human lung using hyperpolarized xenon-129 MRI. *J Magn Reson Imaging*. 2014;39(2):346–59.
- 166) Thomen RP, Sheshadri A, Quirk JD, Kozlowski J, Ellison HD, Szczesniak RD, et al. Regional Ventilation Changes in Severe Asthma after Bronchial Thermoplasty with ^3He MR Imaging and CT. *Radiology*. 2014 Aug 19;274(1):250–9.
- 167) Cox G, Miller JD, McWilliams A, FitzGerald JM, Lam S. Bronchial Thermoplasty for Asthma. *Am J Respir Crit Care Med*. 2006 May 1;173(9):965–9.
- 168) Cox PG, Miller J, Mitzner W, Leff AR. Radiofrequency ablation of airway smooth muscle for sustained treatment of asthma: preliminary investigations. *Eur Respir J*. 2004 Oct 1;24(4):659–63.
- 169) Svenningsen S, Eddy RL, Lim HF, Cox PG, Nair P, Parraga G. Sputum Eosinophilia and Magnetic Resonance Imaging Ventilation Heterogeneity in Severe Asthma. *Am J Respir Crit Care Med*. 2018 Jan 9;197(7):876–84.
- 170) Mugler JP, Altes TA. Hyperpolarized ^{129}Xe MRI of the human lung. *J Magn Reson Imaging JMRI*. 2013 Feb;37(2):313–31.

- 171) Norquay G, Parnell SR, Xu X, Parra-Robles J, Wild JM. Optimized production of hyperpolarized ^{129}Xe at 2 bars for in vivo lung magnetic resonance imaging. *J Appl Phys*. 2013 Jan 28;113(4):044908.
- 172) Venegas JG, Winkler T, Musch G, Melo MFV, Layfield D, Tgavalekos N, et al. Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature*. 2005 Apr;434(7034):777–82.
- 173) Harris RS, Venegas JG, Wongviriyawong C, Winkler T, Kone M, Musch G, et al. ^{18}F -FDG Uptake Rate Is a Biomarker of Eosinophilic Inflammation and Airway Response in Asthma. *J Nucl Med*. 2011 Nov 1;52(11):1713–20.
- 174) Vijayakumar J, Subramanian S, Singh P, Corsini E, Fontanez S, Lawler M, et al. Arterial inflammation in bronchial asthma. *J Nucl Cardiol*. 2013 Jun 1;20(3):385–95.
- 175) Farahi N, Loutsios C, Tregay N, Wright AKA, Berair R, Lok LSC, et al. In vivo imaging reveals increased eosinophil uptake in the lungs of obese asthmatic patients. *J Allergy Clin Immunol*. 2018 Nov 1;142(5):1659-1662.e8.
- 176) Newman S, Salmon A, Nave R, Drollmann A. High lung deposition of $^{99\text{m}}\text{Tc}$ -labeled ciclesonide administered via HFA-MDI to patients with asthma. *Respir Med*. 2006 Mar 1;100(3):375–84.
- 177) De Backer JW, Vos WG, Vinchurkar SC, Claes R, Drollmann A, Wulfrank D, et al. Validation of Computational Fluid Dynamics in CT-based Airway Models with SPECT/CT. *Radiology*. 2010 Dec 1;257(3):854–62.
- 178) Shaw TJ, Wakely SL, Peebles CR, Mehta RL, Turner JM, Wilson SJ, et al. Endobronchial ultrasound to assess airway wall thickening: validation in vitro and in vivo. *Eur Respir J*. 2004 Jun 1;23(6):813–7.
- 179) Soja J, Grzanka P, Śladek K, Okoń K, Ćmiel A, Mikoś M, et al. The Use of Endobronchial Ultrasonography in Assessment of Bronchial Wall Remodeling in Patients With Asthma. *Chest*. 2009 Sep 1;136(3):797–804.

- 180) Goorsenberg AWM, d'Hooghe JNS, de Bruin DM, van den Berk IAH, Annema JT, Bonta PI. Bronchial Thermoplasty-Induced Acute Airway Effects Assessed with Optical Coherence Tomography in Severe Asthma. *Respiration*. 2018;96(6):564–70.
- 181) Ostergaard L, Weisskoff RM, Chesler DA, Gyldensted C, Rosen BR. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I: Mathematical approach and statistical analysis. *Magn Reson Med*. 1996 Nov;36(5):715–25.
- 182) Rosen BR, Belliveau JW, Vevea JM, Brady TJ. Perfusion imaging with NMR contrast agents. *Magn Reson Med*. 1990;14(2):249–65.
- 183) Law M, Young RJ, Babb JS, Peccerelli N, Chheang S, Gruber ML, et al. Gliomas: Predicting Time to Progression or Survival with Cerebral Blood Volume Measurements at Dynamic Susceptibility-weighted Contrast-enhanced Perfusion MR Imaging. *Radiology*. 2008 May 1;247(2):490–8.
- 184) Meyer CR, Boes JL, Kim B, Bland PH, Zasadny KR, Kison PV, et al. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Med Image Anal*. 1997 Apr;1(3):195–206.
- 185) Galbán CJ, Mukherji SK, Chenevert TL, Meyer CR, Hamstra DA, Bland PH, et al. A Feasibility Study of Parametric Response Map Analysis of Diffusion-Weighted Magnetic Resonance Imaging Scans of Head and Neck Cancer Patients for Providing Early Detection of Therapeutic Efficacy. *Transl Oncol*. 2009 Aug 18;2(3):184–90.
- 186) Ma B, Meyer CR, Pickles MD, Chenevert TL, Bland PH, Galbán CJ, et al. Voxel-by-Voxel Functional Diffusion Mapping for Early Evaluation of Breast Cancer Treatment. In: Prince JL, Pham DL, Myers KJ, editors. *Information Processing in Medical Imaging*. Springer Berlin Heidelberg; 2009. p. 276–87. (Lecture Notes in Computer Science).

- 187) Reischauer C, Froehlich JM, Koh D-M, Graf N, Padevit C, John H, et al. Bone Metastases from Prostate Cancer: Assessing Treatment Response by Using Diffusion-weighted Imaging and Functional Diffusion Maps—Initial Observations. *Radiology*. 2010 Nov 1;257(2):523–31.
- 188) Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans Med Imaging*. 2001 Jun;20(6):490–8.
- 189) Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic Epidemiology of COPD (COPDGene) Study Design. *COPD J Chronic Obstr Pulm Dis*. 2011 Feb 1;7(1):32–43.
- 190) Boes JL, Hoff BA, Bule M, Johnson TD, Rehemtulla A, Chamberlain R, et al. Parametric Response Mapping Monitors Temporal Changes on Lung CT Scans in the Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS). *Acad Radiol*. 2015 Feb 1;22(2):186–94.
- 191) Boes JL, Bule M, Hoff BA, Chamberlain R, Lynch DA, Stojanovska J, et al. The Impact of Sources of Variability on Parametric Response Mapping of Lung CT Scans. *Tomography*. 2015 Sep;1(1):69–77.
- 192) Koning HD. ES 02.01 The Dutch-Belgian Lung Cancer Screening Trial (NELSON). *J Thorac Oncol*. 2017 Nov 1;12(11):S1611.
- 193) Pompe E, van Rikxoort EM, Schmidt M, Rühaak J, Estrella LG, Vliegenthart R, et al. Parametric Response Mapping Adds Value to Current Computed Tomography Biomarkers in Diagnosing Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2015 May 1;191(9):1084–6.
- 194) Martinez CH, Diaz AA, Meldrum C, Curtis JL, Cooper CB, Pirozzi C, et al. Age and Small Airway Imaging Abnormalities in Subjects with and without Airflow Obstruction in SPIROMICS. *Am J Respir Crit Care Med*. 2016 Aug 26;195(4):464–72.

- 195) Hayden LP, Hardin ME, Qiu W, Lynch DA, Strand MJ, van Beek EJ, et al. Asthma Is a Risk Factor for Respiratory Exacerbations Without Increased Rate of Lung Function Decline: Five-Year Follow-up in Adult Smokers From the COPDGene Study. *Chest*. 2018 Feb 1;153(2):368–77.
- 196) Hoff BA, Pompe E, Galbán S, Postma DS, Lammers J-WJ, Hacken NHT ten, et al. CT-Based Local Distribution Metric Improves Characterization of COPD. *Sci Rep*. 2017 Jun 7;7(1):2999.
- 197) Mantz H, Jacobs K, Mecke K. Utilizing Minkowski functionals for image analysis: a marching square algorithm. *J Stat Mech Theory Exp*. 2008 Dec;2008(12):P12015.
- 198) Bell AJ, Foy BH, Richardson M, Singapuri A, Mirkes E, van den Berge M, et al. Functional CT imaging for identification of the spatial determinants of small-airways disease in adults with asthma. *J Allergy Clin Immunol*. 2019 Jul 1;144(1):83–93.
- 199) Hunter Peter, Chapman Tara, Coveney Peter V., de Bono Bernard, Diaz Vanessa, Fenner John, et al. A vision and strategy for the virtual physiological human: 2012 update. *Interface Focus*. 2013 Apr 6;3(2):20130004.
- 200) Burrowes K. S., De Backer J., Smallwood R., Sterk P. J., Gut I., Wirix-Speetjens R., et al. Multi-scale computational models of the airways to unravel the pathophysiological mechanisms in asthma and chronic obstructive pulmonary disease (AirPROM). *Interface Focus*. 2013 Apr 6;3(2):20120057.
- 201) Tawhai MH, Hunter P, Tschirren J, Reinhardt J, McLennan G, Hoffman EA. CT-based geometry analysis and finite element models of the human and ovine bronchial tree. *J Appl Physiol*. 2004 Dec 1;97(6):2310–21.
- 202) Pitt-Francis J, Pathmanathan P, Bernabeu MO, Bordas R, Cooper J, Fletcher AG, et al. Chaste: A test-driven approach to software development for biological modelling. *Comput Phys Commun*. 2009 Dec 1;180(12):2452–71.
- 203) Benade AH. On the Propagation of Sound Waves in a Cylindrical Conduit. *J Acoust Soc Am*. 1968 Aug 1;44(2):616–23.

- 204) Thurston GB. Periodic Fluid Flow Through Circular Tubes. *J Acoust Soc Am*. 1952 Nov 1;24(6):653–6.
- 205) Kaczka DW, Massa CB, Simon BA. Reliability of Estimating Stochastic Lung Tissue Heterogeneity from Pulmonary Impedance Spectra: A Forward-Inverse Modeling Study. *Ann Biomed Eng*. 2007 Oct 1;35(10):1722–38.
- 206) Lutchen KR, Gillis H. Relationship between heterogeneous changes in airway morphometry and lung resistance and elastance. *J Appl Physiol*. 1997 Oct 1;83(4):1192–201.
- 207) Kaczka DW, Ingenito EP, Suki B, Lutchen KR. Partitioning airway and lung tissue resistances in humans: effects of bronchoconstriction. *J Appl Physiol*. 1997 May 1;82(5):1531–41.
- 208) Foy BH, Kay D, Bordas R. Modelling responses of the inert-gas washout and MRI to bronchoconstriction. *Respir Physiol Neurobiol*. 2017;235:8–17.
- 209) Foy BH, Gonem S, Brightling C, Siddiqui S, Kay D. Modelling the effect of gravity on inert-gas washout outputs. *Physiol Rep*. 2018;6(10):e13709.
- 210) Soares M, Owers-Bradley J, Foy B, Kay D, Siddiqui S. The evaluation of frequency dependence of resistance using a patient-specific 3D physical model and a computational model. *Eur Respir J*. 2017 Sep 1;50(suppl 61):PA2476.
- 211) Foy BH, Kay D. A computational comparison of the multiple-breath washout and forced oscillation technique as markers of bronchoconstriction. *Respir Physiol Neurobiol*. 2017 Jun 1;240:61–9.
- 212) Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. *Int J Inf Manag*. 2015 Apr 1;35(2):137–44.
- 213) Laney D. 3D data management: Controlling data volume, velocity and variety. META group research note. 2001 Feb 6;6(70):1.
- 214) Chen, Chiang, Storey. Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Q*. 2012;36(4):1165.

- 215) Kwon O, Lee N, Shin B. Data quality management, data usage experience and acquisition intention of big data analytics. *Int J Inf Manag.* 2014 Jun 1;34(3):387–94.
- 216) Luo J, Wu M, Gopukumar D, Zhao Y. Big Data Application in Biomedical Research and Health Care: A Literature Review. *Biomed Inform Insights.* 2016 Jan 1;8:BII.S31559.
- 217) Chen M, Mao S, Liu Y. Big Data: A Survey. *Mob Netw Appl.* 2014 Apr 1;19(2):171–209.
- 218) Barrett MA, Humblet O, Hiatt RA, Adler NE. Big Data and Disease Prevention: From Quantified Self to Quantified Communities. *Big Data.* 2013 Aug 22;1(3):168–75.
- 219) Ram S, Zhang W, Williams M, Pengetnze Y. Predicting Asthma-Related Emergency Department Visits Using Big Data. *IEEE J Biomed Health Inform.* 2015 Jul;19(4):1216–23.
- 220) Belgrave D, Henderson J, Simpson A, Buchan I, Bishop C, Custovic A. Disaggregating asthma: Big investigation versus big data. *J Allergy Clin Immunol.* 2017 Feb 1;139(2):400–7.
- 221) Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *N Engl J Med.* 2007 Nov 22;357(21):2189–94.
- 222) Lang TA, Lang T, Secic M. How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers. ACP Press; 2006. 512 p.
- 223) Sun X, Briel M, Busse JW, Akl EA, You JJ, Mejza F, et al. Subgroup Analysis of Trials Is Rarely Easy (SATIRE): a study protocol for a systematic review to characterize the analysis, reporting, and claim of subgroup effects in randomized trials. *Trials.* 2009 Nov 9;10(1):101.
- 224) Greenhalgh T. How to read a paper: Statistics for the non-statistician. I: Different types of data need different statistical tests. *BMJ.* 1997 Aug 9;315(7104):364–6.

- 225) Curran-Everett D. Multiple comparisons: philosophies and illustrations. *Am J Physiol-Regul Integr Comp Physiol*. 2000 Jul 1;279(1):R1–8.
- 226) Jr FJM. The Kolmogorov-Smirnov Test for Goodness of Fit. *J Am Stat Assoc*. 1951 Mar 1;46(253):68–78.
- 227) Smirnov N. Table for Estimating the Goodness of Fit of Empirical Distributions. *Ann Math Stat*. 1948;19(2):279–81.
- 228) Livingston EH. Who was student and why do we care so much about his t-test?1. *J Surg Res*. 2004 May 1;118(1):58–65.
- 229) Box JF. Guinness, Gosset, Fisher, and Small Samples. *Stat Sci*. 1987 Feb;2(1):45–52.
- 230) Mann HB, Whitney DR. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann Math Stat*. 1947 Mar;18(1):50–60.
- 231) Anderson TW. R. A. Fisher and Multivariate Analysis. *Stat Sci*. 1996;11(1):20–34.
- 232) Kruskal WH, Wallis WA. Use of Ranks in One-Criterion Variance Analysis. *J Am Stat Assoc*. 1952 Dec 1;47(260):583–621.
- 233) McKight PE, Najab J. Kruskal-wallis test. *The corsini encyclopedia of psychology*. 2010 Jan 30:1-1.
- 234) Xu R, Wunsch D. *Clustering*. John Wiley & Sons; 2008. 370 p.
- 235) Everitt BS. Unresolved Problems in Cluster Analysis. *Biometrics*. 1979;35(1):169–81.
- 236) Everitt BS, Landau S, Leese M. *Cluster Analysis*. Taylor & Francis; 2001. 228 p.
- 237) MacQueen J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 1967 Jun 21 (Vol. 1, No. 14, pp. 281-297).

- 238) Jain AK. Data clustering: 50 years beyond K-means. *Pattern Recognit Lett*. 2010 Jun 1;31(8):651–66.
- 239) Ullman JB, Bentler PM. Structural equation modeling. *Handbook of psychology*. 2003 Apr 15:607-34.
- 240) Siroux V, Basagaña X, Boudier A, Pin I, Garcia-Aymerich J, Vesin A, et al. Identifying adult asthma phenotypes using a clustering approach. *Eur Respir J*. 2011 Aug 1;38(2):310–7.
- 241) Xanthopoulos P, Pardalos PM, Trafalis TB. Linear discriminant analysis. In: *Robust data mining 2013* (pp. 27-33). Springer, New York, NY.
- 242) Fisher RA. The Use of Multiple Measurements in Taxonomic Problems. *Ann Eugen*. 1936;7(2):179–88.
- 243) Ghebre MA, Bafadhel M, Desai D, Cohen SE, Newbold P, Rapley L, et al. Biological clustering supports both “Dutch” and “British” hypotheses of asthma and chronic obstructive pulmonary disease. *J Allergy Clin Immunol*. 2015 Jan 1;135(1):63-72.e10.
- 244) Barnes PJ. Against the Dutch Hypothesis: Asthma and Chronic Obstructive Pulmonary Disease Are Distinct Diseases. *Am J Respir Crit Care Med*. 2006 Aug 1;174(3):240–3.
- 245) F.R.S KP. LIII. On lines and planes of closest fit to systems of points in space. *Lond Edinb Dublin Philos Mag J Sci*. 1901 Nov 1;2(11):559–72.
- 246) Burgel P-R, Paillasseur J-L, Caillaud D, Tillie-Leblond I, Chanez P, Escamilla R, et al. Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. *Eur Respir J*. 2010 Sep 1;36(3):531–9.
- 247) Juniper EF, Wisniewski ME, Cox FM, Emmett AH, Nielsen KE, O’Byrne PM. Relationship between quality of life and clinical status in asthma: a factor analysis. *Eur Respir J*. 2004 Feb 1;23(2):287–91.

- 248) Gorban AN, Zinovyev AY. Principal graphs and manifolds. In: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. 2010 (pp. 28-59). IGI Global.
- 249) Euler L. Solutio problematis ad geometriam situs pertinentis. Commentarii academiae scientiarum Petropolitanae. 1741:128-40.
- 250) Shields R. Cultural Topology: The Seven Bridges of Königsburg, 1736. Theory Cult Soc. 2012 Jul 1;29(4-5):43-57.
- 251) Richeson DS. Euler's Gem: The Polyhedron Formula and the Birth of Topology. Edition Unstated edition. Princeton, NJ: Princeton University Press; 2012. 336 p.
- 252) Siersma D. Poincare and Analysis Situs, the beginning of algebraic topology. Nieuw archief voor wiskunde. Serie 5. 2012;13(3):196-200.
- 253) Wasserman L. Topological Data Analysis. Annu Rev Stat Its Appl. 2018;5(1):501-32.
- 254) Camara PG, Rosenbloom DIS, Emmett KJ, Levine AJ, Rabadan R. Topological Data Analysis Generates High-Resolution, Genome-wide Maps of Human Recombination. Cell Syst. 2016 Jul 27;3(1):83-94.
- 255) Singh G, Mémoli F, Carlsson GE. Topological methods for the analysis of high dimensional data sets and 3d object recognition. InSPBG 2007 Sep (pp. 91-100).
- 256) Dey TK, Wang Y. Reeb Graphs: Approximation and Persistence. Discrete Comput Geom. 2013 Jan 1;49(1):46-73.
- 257) Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, et al. Extracting insights from the shape of complex data using topology. Sci Rep. 2013 Feb 7;3:1236.
- 258) Lohr S. Ayasdi: A Big Data Start-Up With a Long History [Internet]. Bits Blog. 2013. Available from: <https://bits.blogs.nytimes.com/2013/01/16/ayasdi-a-big-data-start-up-with-a-long-history/>

- 259) Kim HE. Evaluating ayasdi's topological data analysis for big data (Doctoral dissertation, Master Thesis, Advisors S. Trahasch and RV Zicari, Offenburg University of Applied Sciences Goethe University Frankfurt).
- 260) Nielson JL, Paquette J, Liu AW, Guandique CF, Tovar CA, Inoue T, et al. Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. *Nat Commun.* 2015 Oct 14;6:8581.
- 261) Siddiqui S, Shikotra A, Richardson M, Doran E, Choy D, Bell A, et al. Airway pathological heterogeneity in asthma: Visualization of disease microclusters using topological data analysis. *J Allergy Clin Immunol.* 2018 Nov 1;142(5):1457–68.
- 262) Klein S, Staring M, Murphy K, Viergever MA, Pluim J. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Trans Med Imaging.* 2010 Jan;29(1):196–205.
- 263) Shamonin DP, Bron EE, Lelieveldt BPF, Smits M, Klein S, Staring M, et al. Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front Neuroinformatics.* 2013;7:50.
- 264) Achenbach T, Weinheimer O, Buschsieweke C, Heussel CP, Thelen M, Kauczor HU. [Fully automatic detection and quantification of emphysema on thin section MD-CT of the chest by a new and dedicated software]. *ROFO Fortschr Geb Rontgenstr Nuklearmed.* 2004 Oct;176(10):1409–15.
- 265) Heaney LG, Brightling CE, Menzies-Gow A, Stevenson M, Niven RM, Network on behalf of the BTSDA. Refractory asthma in the UK: cross-sectional findings from a UK multicentre registry. *Thorax.* 2010 Sep 1;65(9):787–94.
- 266) MacIntyre N, Crapo RO, Viegi G, Johnson DC, Grinten CPM van der, Brusasco V, et al. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *Eur Respir J.* 2005 Oct 1;26(4):720–35.
- 267) Pavord ID, Pizzichini MM, Pizzichini E, Hargreave FE. The use of induced sputum to investigate airway inflammation. *Thorax.* 1997 Jun;52(6):498–501.

- 268) Zysman-Colman Z, Lands LC. Whole Body Plethysmography: Practical Considerations. *Paediatr Respir Rev*. 2016 Jun 1;19:39–41.
- 269) Wanger J, Clausen JL, Coates A, Pedersen OF, Brusasco V, Burgos F, et al. Standardisation of the measurement of lung volumes. *Eur Respir J*. 2005 Sep 1;26(3):511–22.
- 270) Horsley AR, Gustafsson PM, Macleod KA, Saunders C, Greening AP, Porteous DJ, et al. Lung clearance index is a sensitive, repeatable and practical measure of airways disease in adults with cystic fibrosis. *Thorax*. 2008 Feb 1;63(2):135–40.
- 271) Weissheimer A, Menezes LM de, Sameshima GT, Enciso R, Pham J, Grauer D. Imaging software accuracy for 3-dimensional analysis of the upper airway. *Am J Orthod Dentofacial Orthop*. 2012 Dec 1;142(6):801–13.
- 272) Bhatawadekar SA, Leary D, Maksym GN. Modelling resistance and reactance with heterogeneous airway narrowing in mild to severe asthma. *Can J Physiol Pharmacol*. 2015 Jan 21;93(3):207–14.
- 273) Pennati F, Salito C, Baroni G, Woods J, Aliverti A. Comparison Between Multivolume CT-Based Surrogates of Regional Ventilation in Healthy Subjects. *Acad Radiol*. 2014 Oct 1;21(10):1268–75.
- 274) Choi S, Hoffman EA, Wenzel SE, Tawhai MH, Yin Y, Castro M, et al. Registration-based assessment of regional lung function via volumetric CT images of normal subjects vs. severe asthmatics. *J Appl Physiol Bethesda Md* 1985. 2013 Sep 1;115(5):730–42.
- 275) Lui JK, Lutchen KR. The role of heterogeneity in asthma: a structure-to-function perspective. *Clin Transl Med*. 2017 Aug 3;6(1):29.
- 276) Zha W, Kruger SJ, Johnson KM, Cadman RV, Bell LC, Liu F, et al. Pulmonary ventilation imaging in asthma and cystic fibrosis using oxygen-enhanced 3D radial ultrashort echo time MRI. *J Magn Reson Imaging*. 2018;47(5):1287–97.

- 277) Venegas J, Winkler T, Harris RS. Lung Physiology and Aerosol Deposition Imaged with Positron Emission Tomography. *J Aerosol Med Pulm Drug Deliv.* 2012 Aug 2;26(1):1–8.
- 278) Farrow CE, Salome CM, Harris BE, Bailey DL, Berend N, King GG. Peripheral ventilation heterogeneity determines the extent of bronchoconstriction in asthma. *J Appl Physiol.* 2017 Aug 10;123(5):1188–94.
- 279) Verbanck S, Paiva M. Gas mixing in the airways and airspaces. *Compr Physiol.* 2011 Jan;1(2):809-34.
- 280) Gonem S, Hardy S, Buhl N, Hartley R, Soares M, Kay R, et al. Characterization of acinar airspace involvement in asthmatic patients by using inert gas washout and hyperpolarized (3)helium magnetic resonance. *J Allergy Clin Immunol.* 2016 Feb;137(2):417–25.
- 281) Farah CS, King GG, Brown NJ, Downie SR, Kermode JA, Hardaker KM, et al. The role of the small airways in the clinical expression of asthma in adults. *J Allergy Clin Immunol.* 2012 Feb 1;129(2):381-387.e1.
- 282) Thompson BR, Douglass JA, Ellis MJ, Kelly VJ, O’Hehir RE, King GG, et al. Peripheral lung function in patients with stable and unstable asthma. *J Allergy Clin Immunol.* 2013 May 1;131(5):1322–8.
- 283) Galvin I, Drummond GB, Nirmalan M. Distribution of blood flow and ventilation in the lung: gravity is not the only factor. *BJA Br J Anaesth.* 2007 Apr 1;98(4):420–8.
- 284) Gupta S, Hartley R, Khan UT, Singapuri A, Hargadon B, Monteiro W, et al. Quantitative computed tomography–derived clusters: Redefining airway remodeling in asthmatic patients. *J Allergy Clin Immunol.* 2014 Mar 1;133(3):729-738.e18.
- 285) Shim SS, Schiebler ML, Evans MD, Jarjour N, Sorkness RL, Denlinger LC, et al. Lumen area change (Delta Lumen) between inspiratory and expiratory multidetector computed tomography as a measure of severe outcomes in asthmatic patients. *J Allergy Clin Immunol.* 2018 Dec 1;142(6):1773-1780.e9.

- 286) Bhatt SP, Soler X, Wang X, Murray S, Anzueto AR, Beaty TH, et al. Association between Functional Small Airway Disease and FEV1 Decline in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2016 Jan 25;194(2):178–84.
- 287) Reddel HK, Taylor DR, Bateman ED, Boulet L-P, Boushey HA, Busse WW, et al. An Official American Thoracic Society/European Respiratory Society Statement: Asthma Control and Exacerbations. *Am J Respir Crit Care Med*. 2009 Jul 1;180(1):59–99.
- 288) Dutrieue B, Vanholsbeeck F, Verbanck S, Paiva M. A human acinar structure for simulation of realistic alveolar plateau slopes. *J Appl Physiol*. 2000 Nov 1;89(5):1859–67.
- 289) Cauberghs M, Van de Woestijne KP. Mechanical properties of the upper airway. *J Appl Physiol*. 1983 Aug 1;55(2):335–42.
- 290) Boudewijn IM, Postma DS, Telenga ED, Hacken NHT ten, Timens W, Oudkerk M, et al. Effects of ageing and smoking on pulmonary computed tomography scans using parametric response mapping. *Eur Respir J*. 2015 Jun 25;ERJ-00094-2015.
- 291) Norman P. The CXCR5 receptor, a new target in inflammation? *Expert Opin Ther Pat*. 2008 Oct 1;18(10):1209–10.
- 292) Campana L, Kenyon J, Zhalehdoust-Sani S, Tzeng Y-S, Sun Y, Albert M, et al. Probing airway conditions governing ventilation defects in asthma via hyperpolarized MRI image functional modeling. *J Appl Physiol*. 2009 Apr 1;106(4):1293–300.
- 293) Verbanck S, King GG, Zhou W, Miller A, Thamrin C, Schuermans D, et al. The quantitative link of lung clearance index to bronchial segments affected by bronchiectasis. *Thorax*. 2018 Jan 1;73(1):82–4.
- 294) Partridge MR, van der Molen T, Myrseth S-E, Busse WW. Attitudes and actions of asthma patients on regular maintenance therapy: the INSPIRE study. *BMC Pulm Med*. 2006 Jun 13;6(1):13.

- 295) Hwang S, Son S-W, Kim SC, Kim YJ, Jeong H, Lee D. A protein interaction network associated with asthma. *J Theor Biol.* 2008 Jun 21;252(4):722–31.
- 296) Farrow CE, Salome CM, Harris BE, Bailey DL, Bailey E, Berend N, et al. Airway closure on imaging relates to airway hyperresponsiveness and peripheral airway disease in asthma. *J Appl Physiol.* 2012 Jul 26;113(6):958–66.
- 297) Postma DS, Brightling C, Fabbri L, Molen T van der, Nicolini G, Papi A, et al. Unmet needs for the assessment of small airways dysfunction in asthma: introduction to the ATLANTIS study. *Eur Respir J.* 2015 Jun 1;45(6):1534–8.
- 298) Bellemare F, Jeanneret A, Couture J. Sex Differences in Thoracic Dimensions and Configuration. *Am J Respir Crit Care Med.* 2003 Aug 1;168(3):305–12.
- 299) Brashier B, Salvi S. Measuring lung function using sound waves: role of the forced oscillation technique and impulse oscillometry system. *Breathe.* 2015 Mar 1;11(1):57–65.
- 300) Tomalak W, Radlinski J. Reference values for IOS derived R5-R20 and AX assessed from healthy polish children. *Eur Respir J.* 2016 Sep 1;48(suppl 60):PA3614.
- 301) Achenbach T, Weinheimer O, Biedermann A, Schmitt S, Freudenstein D, Goutham E, et al. MDCT assessment of airway wall thickness in COPD patients using a new method: correlations with pulmonary function tests. *Eur Radiol.* 2008 Dec 1;18(12):2731–8.
- 302) Bell AJ. MD Simulation for Head-on Collision of Liquid Nanodroplets Obeying Modified LJ Potential. *arXiv preprint arXiv:1507.01070.* 2015 Jul 4.
- 303) Politi AZ, Donovan GM, Tawhai MH, Sanderson MJ, Lauzon A-M, Bates JHT, et al. A multiscale, spatially distributed model of asthmatic airway hyper-responsiveness. *J Theor Biol.* 2010 Oct 21;266(4):614–24.
- 304) Tawhai MH, Hunter PJ. Multibreath washout analysis: modelling the influence of conducting airway asymmetry. *Respir Physiol.* 2001 Sep 1;127(2):249–58.

- 305) Swan AJ, Clark AR, Tawhai MH. A computational model of the topographic distribution of ventilation in healthy human lungs. *J Theor Biol.* 2012 May 7;300:222–31.
- 306) Raghavan D, Jain R. Increasing awareness of sex differences in airway diseases. *Respirology.* 2016;21(3):449–59.
- 307) Vinchurkar S, Backer LD, Vos W, Holsbeke CV, Backer JD, Backer WD. A case series on lung deposition analysis of inhaled medication using functional imaging based computational fluid dynamics in asthmatic patients: effect of upper airway morphology and comparison with in vivo data. *Inhal Toxicol.* 2012 Jan 1;24(2):81–8.
- 308) De Backer W, Devolder A, Poli G, Acerbi D, Monno R, Herpich C, et al. Lung Deposition of BDP/Formoterol HFA pMDI in Healthy Volunteers, Asthmatic, and COPD Patients. *J Aerosol Med Pulm Drug Deliv.* 2010 Jan 29;23(3):137–48.
- 309) Holley HS, Milic-Emili J, Becklake MR, Bates DV. Regional Distribution of Pulmonary Ventilation and Perfusion in Obesity. *J Clin Invest.* 1967 Apr 1;46(4):475–81.