Non-invasive Pressure Estimation in Patients with Pulmonary Arterial Hypertension: Data-Driven or Model-Based?

Yingyu Yang[1](✉), Stephane Gillon[2], Jaume Banus[1], Pamela Moceri[2], and Maxime Sermesant[1]

[1] Inria, Université Côte d'Azur, Sophia Antipolis, France
`yingyu.yang@inria.fr`
[2] Nice University Hospital, Université Côte d'Azur, Nice, France

Abstract. Right heart catheterisation is considered as the gold standard for the assessment of patients with suspected pulmonary hypertension. It provides clinicians with meaningful data, such as pulmonary capillary wedge pressure and pulmonary vascular resistance, however its usage is limited due to its invasive nature. Non-invasive alternatives, like Doppler echocardiography could present insightful measurements of right heart but lack detailed information related to pulmonary vasculature. In order to explore non-invasive means, we studied a dataset of 95 pulmonary hypertension patients, which includes measurements from echocardiography and from right-heart catheterisation. We used data extracted from echocardiography to conduct cardiac circulation model personalisation and tested its prediction power of catheter data. Standard machine learning methods were also investigated for pulmonary artery pressure prediction. Our preliminary results demonstrated the potential prediction power of both data-driven and model-based approaches.

Keywords: Cardiac modelling · Machine learning · Pulmonary hypertension

1 Introduction

Pulmonary arterial hypertension (PAH) is a pathological hemodynamic condition defined as mean pulmonary arterial pressure (mPAP) at rest >25 mmHg, measured by gold standard - right heart catheterisation (RHC) [12]. Pulmonary arterial hypertension can originate in lungs, heart, pulmonary artery and blood, and eventually leads to right heart failure or death. Standard diagnostic procedure requires clinical evaluation, non-invasive imaging and right heart catheterisation [8].

However, some patients do not receive RHC as part of their diagnostic routine and this may be related to lack of training or the potential perception of

RHC invasive risk, especially in the pediatric population [13]. This phenomenon increases the possibility of incomplete diagnosis, which diminishes the effect of targeted therapies [4]. In reality, echocardiography and catheterisation are usually conducted in separated labs. In order to combine the hemodynamic information provided by RHC and echocardiography, in our work, we explored the possibility of incorporating catheter-based data prediction, specifically, mean pulmonary artery pressure (mPAP) and pulmonary vascular resistance (PVR), into routine echocardiography diagnosis.

There exists very simple ways to estimate PVR [11] and mPAP [3] but most of them only rely on one or two echocardiographic measurements, which largely propagates measurement uncertainty to prediction and constrains their usage under different physiological conditions. Recently, with the advance of machine learning techniques, data-driven algorithms demonstrated good performance in cardiac tasks [5]. Besides, numerical modeling of pulmonary circulation also showed the ability to assess hemodynamic values non-invasively [9]. In our work, we used a simplified cardiac lumped model which can be easily personalised from clinical data in order to simulate cardiac indicators. In addition, machine-learning based regression methods were also tested for their prediction power.

2   Methods

2.1   Data Presentation

Our retrospective dataset was collected from the records of Nice University Hospital in 123 patients with known or suspected pulmonary hypertension. Echocardiography-based cardiac indicators, such as ejection fraction, end-diastolic left and right ventricular volumes, were extracted by an experienced cardiologist. Complete or incomplete catheterisation measurement records (44% received both echocardiography and catheterisation within 48 h) are available for all the patients (see detailed data description in Table 1[*]). Specifically, RAP in echocardiography data is estimated from inferior vena cava (IVC) diameter and its respirophasic variations, which leads to an ordinal value with possible values from {5,10,15,20}. sPAP is then calculated by $sPAP = 4 * TRV_{max}^2 + RAP$,

---

[*]Abbreviations: Body Surface Area (BSA), Pulmonary Artery HyperTension (PAHT), Heart Rate (HR), Brain Natriuretic Peptide (BNP), Blood Pressure (BP), Left Ventricle Ejection Fraction (LVEF), Left Ventricle Outflow Track Diameter (D_{LVOT}), Velocity Time Integral of Left Ventricle Outflow Tract (VTI_{LVOT}), Left Ventricle End-Diastolic Diameter (LVEDD), Left Ventricle End-Systolic Diameter (LVESD), Left Ventricle End-Diastolic volume (LVEDV), Right Ventricle Ejection Fraction 3D (RVEF 3D), Right Ventricle Outflow Tract Diameter (D_{RVOT}), Velocity Time Integral of Right Ventricle Outflow Tract (VTI_{RVOT}), Right Ventricle End-Systolic Diameter (RVESD), Right Ventricle End-Diastolic Volume (RVEDV), Systolic Pulmonary Artery Pressure (sPAP), Tricuspid Annular Plane Systolic Excursion (TAPSE), Right Atrium Pressure (RAP), Mean Pulmonary Artery Pressure (mPAP), Pulmonary Capillary Wedge Pressure (Pcap), Pulmonary Vascular Resistance (PVR), Cardiac Output (CO), Cardiac Index (CI)

Table 1. Detailed description of patient data

Feature	Missing	Statistics	Feature	Missing	Statistics
Clinical information					
Age	0	62 ± 18	Sex	0	50.5% female
Height (cm)	1	166.2 ± 9.5	Weight (kg)	1	68.2 ± 16.6
BSA	5	1.75 ± 0.23	BNP (ng/L)	7	275.4 ± 508.6
HR	24	76 ± 13	BP (mmHg)	72	132 ± 22 & 81 ± 16
		61.0% (1), 22.1% (3)			36.8%, (2), 46.3% (3)
PAHT group	1	13.6% (4), 2.1% (5)	NYHA	3	13.6% (4)
		1.2% (NAN)			3.3% (NAN)
Echocardiography data					
LVEF (%)	2	67.5 ± 8.9	RVEF 3D (%)	8	35.0 ± 9.8
D_{LVOT} (mm)	59	18.9 ± 4.6	D_{RVOT} (mm)	85	26.3 ± 5.2
VTI_{LVOT} (cm)	59	19.4 ± 5.7	VTI_{RVOT} (cm)	3	14.6 ± 4.5
LVEDD (mm)	18	43.7 ± 6.7	RVEDD (mm)	21	46.2 ± 6.4
LVESD (mm)	42	26.1 ± 6.3	RVEDV (mL)	8	98.1 ± 39.7
sPAP (mmHg)	1	73.5 ± 23.5	RAP (mmHg)	0	10.6 ± 4.6
TAPSE (mm)	1	19.9 ± 5.5	S'Wave (cm/s)	1	11.3 ± 3.0
Catheter data					
mPAP (mmHg)	0	43.5 ± 13.0	Pcap (mmHg)	1	11.4 ± 4.0
RAP (mmHg)	0	8.9 ± 4.3	PVR (UW)	4	7.5 ± 4.0
CO (L/min)	4	4.7 ± 1.4	CI (L/min/m^2)	6	2.7 ± 0.7

where TRV_{max} refers to tricuspid regurgitation maximum velocity. In our analysis, records of 95 patients were included. The other 28 records were discarded because of lack of catheter measurement.

2.2   Modeling-Based Prediction

Cardiovascular 0D Model. To incorporate cardiovascular dynamics into the prediction model, we consider a 0D model of the whole cardiovascular circulation system [2]. Derived from a 3D cardiac electromechanical model, the 0D model not only consists of less ordinary differential equation but also preserves the capacity to describe the important properties of the heart. Under the assumption of the spherical ventricle symmetry in 0D model, the inner radius (R_0) is directly related to the myocardial size. Reduced deformation and stress tensors demonstrate good representation of important cardiac characteristics, such as heart contractility (σ_0) and stiffness (C_1).

This 0D model has manifested its modeling potential in solving personalisation problems [10]. Consider a 0D model M, with a set of parameters P_M and model states O_M. We take a subset $\theta \subseteq P_M$, which contains parameters such as heart contractility (σ_0) and myocardial stiffness (C_1), and fix all the other parameters with default values. Interesting model states $O \subseteq O_M$, such as pulmonary artery pressure and ejection fraction, present cardiac indicators of the heart model. Given a set of clinical observations $\hat{O}$, the aim of personalisation is to find suitable varying parameters $\hat{\theta}$ so that the corresponding output of the fitted 0D model is as close as possible to clinical references, i.e. $O(\hat{\theta}) \approx \hat{O}$ (Fig. 1).
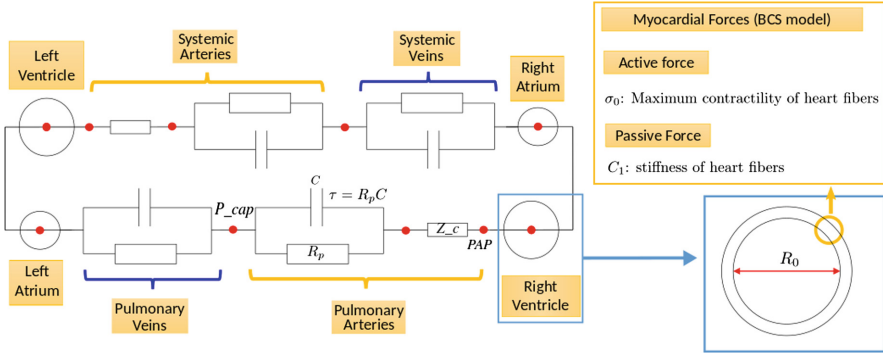
Fig. 1. Schema of used cardiac 0D model (Adapted from [1]).

We assume Gaussian distribution priors for both interested parameters θ and model states O, i.e. $\theta \sim \mathcal{N}(\mu, \Sigma)$ and $O|\theta \sim \mathcal{N}(\hat{O}(\theta), \Delta)$. Essentially, the personalisation problem equals to Maximum A Posterior. With Gaussian distribution, the objective function is derived as:

$$min\{(O(\theta) - \hat{O})^T \Delta^{-1}(O(\theta) - \hat{O}) + \gamma(\theta - \mu)^T \Sigma^{-1}(\theta - \mu)\}$$

where $\hat{O}$ refers to observed model states and Δ, a diagonal covariance matrix, represents the tolerance interval for each dimension of model states. The second term is regarded as a regulariser. γ controls to what extent we forces the parameter θ to follow the prior distribution, which helps to attenuate the non-unique solution effect of this ill-posed inverse problem.

We solve this high-dimensional and non-convex problem by applying a nonparametric evolutionary strategy CMA-ES [6]. Iteratively Updated Prior (IUP) method, as defined in [10], is deployed to iteratively update the prior distribution based on former population personalisation results.

Experiments. We first investigate the intrinsic prediction power of the 0D cardiovascular model. Based on the available clinical data, we chose the following 5 features extracted from echocardiography for personalisation: systolic pulmonary artery pressure (sPAP), right ventricle ejection fraction (RVEF), right ventricle end-diastolic volume (RVEDV), left ventricle ejection fraction (LVEF). In order to assure equal stroke volume of left and right heart, the left ventricle end-diastolic volume (LVEDV) is calculated from available data: $LVEDV = \frac{RVEDV*RVEF}{LVEF}$. Considering the uncertainty of measurement, we assign a tolerance interval for every selected feature: 200 Pa for sPAP, 5% for LVEF and RVEF and 10 mL for RVEDV and LVEDV. Available RAP values are not included in our setting. Finally, parameters of both left heart and right heart are selected for personalisation: left and right heart contractility (σ_0), left and right myocardial stiffness (c_1), right ventricle inner radius (R_0), pulmonary proximal resistance (Z_c) and pulmonary distal resistance(R_p). Left ventricle radius

is set by $(LVEDD + LVESD)/4$ if both LVEDD and LVESD are available. Or else it is set to 18 mm, the mean value in the population. Since every patient possesses at least one target feature, we fit the model on the whole dataset of 95 patients. We assume the covariance Σ of varying parameters is a full matrix and γ is selected from $\{0.1, 0.5, 1, 2\}$.

We follow the same protocol as the cardiologist to extract mPAP and PVR from model output curves: mPAP is calculated as the mean value of pulmonary pressure-time integral during one cardiac cycle, and PVR is calculated as PVR (UW) = (mPAP − Pcap)/CO, where CO comes from flow-time integral and heart rate and Pcap is fixed at 10 mmHg.

A supervised method is also proposed based on personalisation. We split our dataset into training data and test data with a configuration of 5-fold cross-validation. In the training phase, echocardiography and catheter features are fitted iteratively with $\gamma = 0.5$ for 10 iterations. Then the fitted parameter distribution of the 10th iteration of training data is used as test prior. We then perform one iteration of personalisation to fit only echocardiography features for test data with $\gamma \in \{0.5, 1, 2\}$.

The optimisation of 0D model personalisation is performed over the logarithm of the parameter values (Table 2).

Table 2. Selected features and parameters for 0D model personalisation

Echo	Catheter	Varying parameters
sPAP	mPAP	Left heart contractility σ_0
RVEF	CO	Left heart stiffness c_1
LVEF	Pcap	Right heart contractility σ_0
RVEDV	PVR	Right heart stiffness c_1
LVEDV		Right ventricular radius R_0
		Pulmonary proximal resistance Z_p
		Pulmonary distal resistance R_p

Model Implementation. Our cardiac 0D model is originally implemented in CellML language. It was exported into C language and incorporated into a Python program which enables flexible experiments. The 0D model is very fast and it takes less than 1 s to output cardiac curves. CMA package implemented by Hansen et al. [7] is used in our optimisation. With parallel computation, optimal parameters for one patient can be found in 3 min on a computer with 8 cores (Intel i7-8650U CPU 1.90 GHz).

2.3   Learning-Based Prediction

5 regression methods implemented in $scikit-learn$ 0.21.2 were tested using echocardiographic cardiac features to predict catheter data: lasso regression, ridge regression (RR), k-nearest neighbour regression (KNN), partial least-square regression (PLR) and ada-boosting decision tree regression

(ADAT). Optimal hyper-parameters of different estimators were determined through nested 10-fold cross-validation grid search. Specifically, we search $\alpha \in \{10^{-3}, 10^{-2}, ...10^2\}$ for Lasso and Ridge, number of neighbors $N \in \{2, 3, ...10\}$ for KNN , number of components $N \in \{1, 2, ...15\}$ for PLR and number of estimators $N \in \{2, 4, 8, 16, 50, 100, 200\}$ for ADAT.

We use all the data except catheter data to perform regression analysis. Category data, such as NYHA, Group PAH, and columns with more than 40% missing values (VTI_{LVOT}, D_{RVOT}, D_{LVOT}, LVESD and BP) were eliminated. From available data, we are able to calculate TRV_{max} and $\frac{TRV_{max}}{VTI_{RVOT}}$, the later of which is reported correlated with PVR [11]. A correlation analysis on the 18 predictors shows linearity between some predictors (correlation coefficient larger than 0.6) and finally we have 11 predictors left for regression analysis: age, BSA, LVEF, RVEF, HR, RAP, sPAP, LVEDD, RVEDV,VTI_{RVOT}, TAPSE. Considering the missing value problem of our dataset, simple and multiple imputation methods implemented in $scikit-learn$ 0.21.2 are also conducted before every regression learning: mean imputation, median imputation, Bayesian ridge regression iterative imputation, k-nearest neighbour iterative imputation, decision tree regression iterative imputation and extra-tree iterative imputation. We report R^2 score (coefficient of determination) and root mean squared error (RMSE) for each regression method based on a 5-fold cross validation.

We also test simple estimation (SIMPLE) methods for mPAP and PVR based on formulas $mPAP = 0.61 * PAPs + 2(mmHg)$ following the work of [3] and $PVR = 29.7 * (TRV_{max}/VTI_{RVOT}) - 0.29$ following the work of [11].
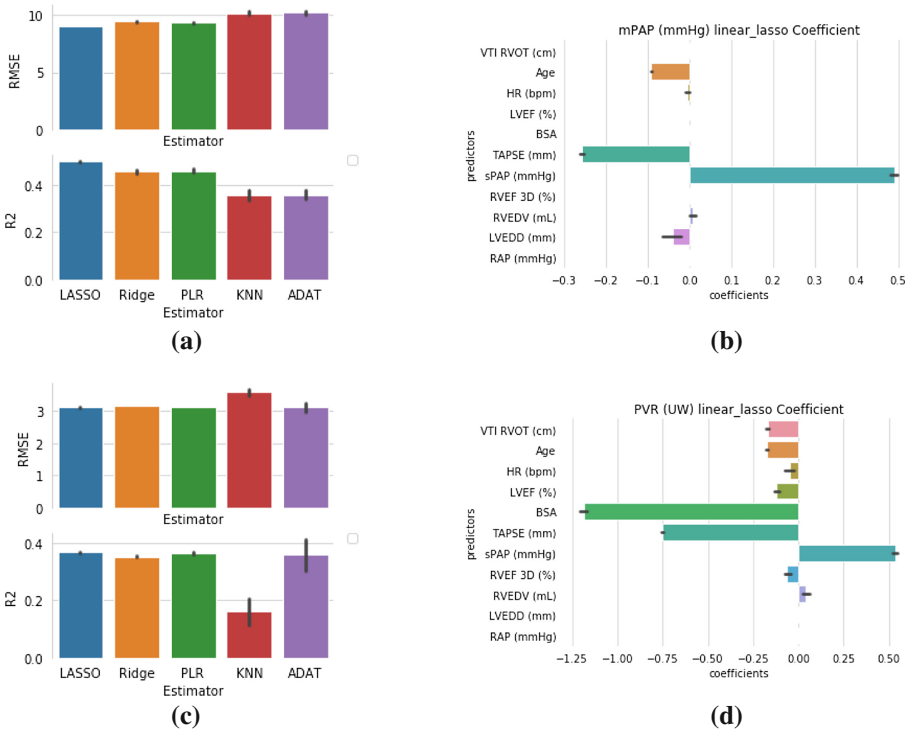
3   Results

Modelling-Based Prediction. With only echocardiography-based indicators, our result of 0D model personalisation indicate that a reasonable γ improves prediction accuracy. A large γ will nominate objective function and forces varying parameter to follow prior distribution, while a small γ enables more accurate feature fitting. In our case, with $\gamma = 0.5$, estimated mPAP correlates modestly with ground truth ($r = 0.65, p < 0.0001$) and demonstrates a reasonable error (shown in Table 3: MF0.5). With $\gamma = 1$, estimated PVR has the lowest error and correlates slightly with ground truth ($r = 0.40, p < 0.001$).

Table 3. 0D cardiac model based prediction results

State	Metrics	MF0.1	MF0.5	MF1	MF2	MF-CV0.5	MF-CV1	MF-CV2
mPAP	$RMSE$	11.08	10.61	10.83	11.44	13.01	12.76	12.92
	R^2	0.26	0.32	0.30	0.21	-0.05	-0.01	-0.05
PVR	$RMSE$	6.17	5.72	4.84	5.00	10.55	9.44	9.34
	R^2	-1.36	-1.04	-0.45	-0.55	-7.33	-6.30	-5.53

However, in modelling-based supervised method (MF-CV), when echocardiography and catheter data are mixed for personalistion, the discrepancy between ECHO and CAT data mislead parameter prior direction. After training phase, we obtain prior distribution from last iteration of group personalisation. When new test data comes, personalisation is moving to a biased direction.

Learning-Based Prediction. In Fig. 2, we observe that LASSO and PLR estimators not only demonstrate less prediction error, but also are more stable to various imputed data. Lasso coefficients show that both sPAP and TAPSE are significant factors for mPAP and PVR regression. This is consistent with the fact that mPAP and PVR are highly correlated ($r = 0.81, p < 0.01$).

Fig. 2. Mean Pulmonary Artery Pressure (mPAP) and Pulmonary Vascular Resistance (PVR) prediction results (Data-driven methods). Results of models with different imputation methods are averaged to distinguish the performance of estimators. Results shown in mean ± std. (a) RMSE and R2 metric of different estimator for mPAP. (b) Lasso regression coefficient (alpha = 0.1) for mPAP prediction. (c) RMSE and R2 metric of different estimator for PVR. (d) Lasso regression coefficient (alpha = 0.01) for PVR prediction.

Prediction Summary. We present the averaged metric value (based on different imputation methods) and involved features for all estimators. With lasso regression result, we exclude the features with normalized coefficient smaller than 0.01 for mPAP and 0.2 for PVR, e.t. we have sPAP, TAPSE, LVEDD and age for mPAP and BSA, sPAP, TAPSE for PVR. We then redo LASSO RIDGE and PLR with those selected features. Here mPAP's best prediction is with tuned hyperparameter: $\alpha = 0.01$ for Lasso, $\alpha = 0.1$ for RR, $N = 8$ for KNN, $N = 2$ for PLR and $N = 500$ for ADAT. PVR best result is with hyperparameters: $\alpha = 0.01$ for Lasso, $\alpha = 0.5$ for RR, $N = 7$ for KNN, $N = 2$ for PLR and $N = 500$ for ADAT (Table 4).

Table 4. Best regression results of different estimators for mPAP and PVR

State	Metrics	LASSO	RR	PLR	KNN	ADAT	SIMPLE	MF-CV1	MF1
mPAP	$RMSE$	8.75	8.72	8.76	10.12	10.18	11.52	12.76	10.83
	R^2	0.52	0.53	0.52	0.35	0.36	0.20	-0.01	0.30
	Features	TAPSE sPAP Age LVEDD			All 11 Features		sPAP	sPAP LVEF LVEDV RVEF RVEDV	
PVR	$RMSE$	2.99	2.98	2.94	3.56	3.09	3.96	9.44	4.84
	R^2	0.41	0.41	0.42	0.16	0.35	0.04	-6.30	-0.45
	Features	TAPSE sPAP BSA			All 11 Features		$\frac{TRV_{max}}{VTI_{RVOT}}$	sPAP LVEF LVEDV RVEF RVEDV	

Using LASSO regression, we average the coefficient from different imputation methods and get the following estimation formula:

$$mPAP = 0.32 * sPAP - 0.65 * TAPSE - 0.12 * Age - 0.12 * LVEDD + 45.83 \quad (1)$$

$$PVR = 0.05 * sPAP - 0.33 * TAPSE - 4.94 * BSA + 18.83 \quad (2)$$

Supervised 0D model prediction (MD-CV1) fails to retain a good parameter prior for prediction however, echocardiography-based group optimisation demonstrates a prediction potential, which reveals the regularizing effect of population-based prior distribution. Here, best result is reached at $\gamma = 1$.

SIMPLE methods provide simple approximation of mPAP and PVR but their validity is restricted due to their dependence on one single measurement. Besides, regression methods surpass model-based estimation approaches. There may be two main reasons for their difference. First, we are not using all the available information for 0D model personalisation. For example, TAPSE, who is of significance in regression, are difficult to incorporate into 0D personalisation system.
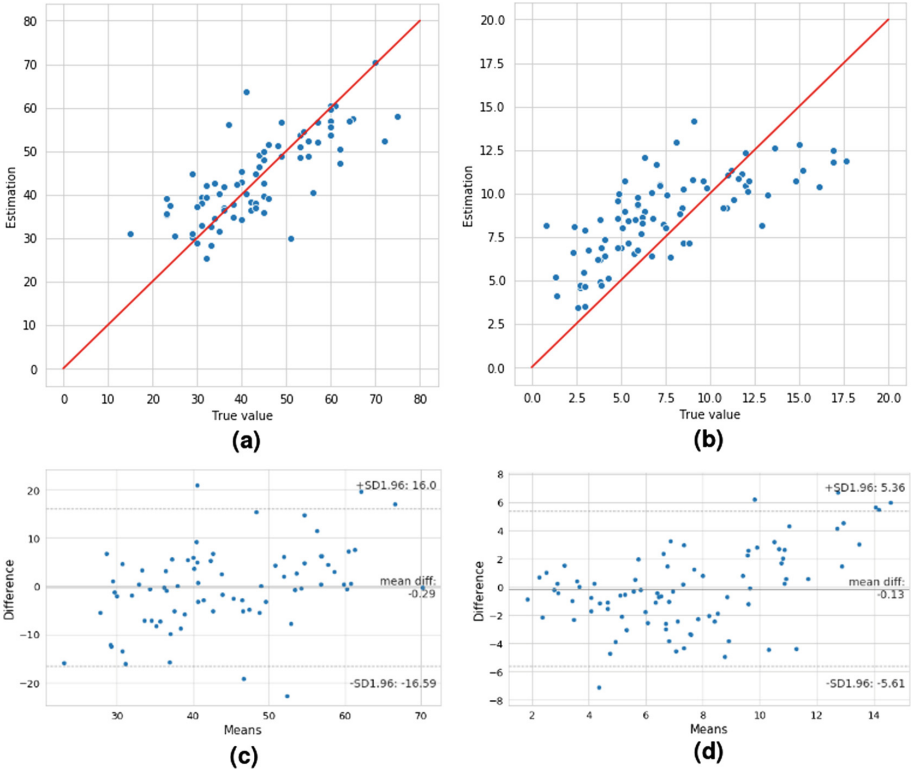
Fig. 3. Estimated value and ground truth comparison (lasso formulas). (a) The plot of mPAP ground truth and its estimated value using Eq. 1. (b) The plot of PVR ground truth and its estimated value using Eq. 2. (c) Bland-Altman analysis demonstrating the limits of agreement between invasive mPAP and mPAP determined via echocardiography, using Eq. 1. (d) Bland-Altman analysis demonstrating the limits of agreement between invasive PVR and PVR determined via echocardiography, using Eq. 2.

Secondly, our 0D model is highly reduced, some important measurements like VTI_{RVOT} and TR_{max} which exhibit important hemodynamic characteristics, is not compatible. Whereas, unlike the imperative demand of complete data for regression methods, 0D model personalisation can deal with missing data issue naturally [10] (Table 4).

4   Conclusion

Our preliminary results show a good potential of using data-driven methods and model-based approaches for estimating pulmonary pressure in pulmonary hypertension patients. Data-driven method is fast, simple and give good approximation of pulmonary pressure, but it strongly demands complete observation. Model-based approach captures complex hemodynamics from observed data and

deals with missing data issue naturally. Compared with data-driven methods, it exhibits a slightly poorer prediction accuracy. Based on current exploration, there are two directions of future work. One is to extend 0D model personalisation method so as to integrate more observed data into system. The other is adopting data-driven methods to predict accurate parameter distribution for personalisation.

References

1. Banus, J., Lorenzi, M., Camara, O., Sermesant, M.: Large scale cardiovascular model personalisation for mechanistic analysis of heart and brain interactions. In: Coudière, Y., Ozenne, V., Vigmond, E., Zemzemi, N. (eds.) Functional Imaging and Modeling of the Heart, FIMH 2019. Lecture Notes in Computer Science, vol. 11504, pp. 285–293. Springer, Cham (2019)
2. Caruel, M., Chabiniok, R., Moireau, P., Lecarpentier, Y., Chapelle, D.: Dimensional reductions of a cardiac model for effective validation and calibration. Biomech. Model. Mechanobiology **13**(4), 897–914 (2014)
3. Chemla, D., et al.: New formula for predicting mean pulmonary artery pressure using systolic pulmonary artery pressure. Chest **126**(4), 1313–1317 (2004)
4. Deaño, R.C., et al.: Referral of patients with pulmonary hypertension diagnoses to tertiary pulmonary hypertension centers. JAMA Intern. Med. **173**(10), 887 (2013)
5. Gudigar, A., et al.: Global weighted LBP based entropy features for the assessment of pulmonary hypertension. Pattern Recogn. Lett. **125**, 35–41 (2019)
6. Hansen, N.: The CMA evolution strategy: a comparing review. Towards New Evol. Comput. **102**(2006), 75–102 (2016)
7. Hansen, N., Akimoto, Y., Baudis, P.: CMA-ES/pycma on Github, February 2019
8. Howard, L.S., et al.: Echocardiographic assessment of pulmonary hypertension: standard operating procedure. Eur. Respir. Rev. **21**(125), 239–248 (2012)
9. Kheyfets, V.O., O'Dell, W., Smith, T., Reilly, J.J., Finol, E.A.: Considerations for numerical modeling of the pulmonary circulation-a review with a focus on pulmonary hypertension. J. Biomech. Eng. **135**(6), 061011 (2013)
10. Molléro, R., Pennec, X., Delingette, H., Ayache, N., Sermesant, M.: Population-based priors in cardiac model personalisation for consistent parameter estimation in heterogeneous databases. Int. J. Numer. Methods Biomed. Eng. **35**, e3158 (2018)
11. Rajagopalan, N., et al.: Noninvasive estimation of pulmonary vascular resistance in pulmonary hypertension. Echocardiography **26**(5), 489–494 (2009)
12. Rosenkranz, S., Preston, I.R.: Right heart catheterisation: best practice and pitfalls in pulmonary hypertension. Eur. Respir. Rev. **24**(138), 642–652 (2015)
13. Zuckerman, W.A.: Safety of cardiac catheterization at a center specializing in the care of patients with pulmonary arterial hypertension. Pulm. Circul. **3**(4), 831–839 (2013)