Improving Assessment in Kidney Transplantation by Multitask General Path Model

Qing Lan[1§], Xiaoyu Chen[2§], Murong Li[3], John Robertson[4], Yong Lei[3], Ran Jin[1]

[1] Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA
[2] Department of Industrial Engineering, University of Louisville, Louisville, KY 40292, USA
[3] State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou 310013, China
[4] School of Biomedical Engineering and Sciences, Virginia Tech, Blacksburg, VA 24061, USA
[§] Qing Lan and Xiaoyu Chen contributed equally to this work.

Correspondence Author: Qing Lan, 250 Durham Hall, 1145 Perry Street, Blacksburg, VA 24061, USA. E-mail: qing83@vt.edu. Tel: 404-580-1940

ABSTRACT

Kidney transplantation helps end-stage patients regain health and quality-of-life. The decisions for matching donor kidneys and recipients affect success of transplantation. However, current kidney matching decision procedures do not consider viability loss during preservation. The objective here is to forecast heterogeneous kidney viability, based on historical datasets to support kidney matching decision-making. Six recently procured porcine kidneys were used to conduct viability assessment experiments to validate the proposed multitask general path model. The model forecasts kidney viability by transferring knowledge from learning the commonality of all kidneys and the heterogeneity of each kidney. The proposed model provides exactly accurate kidney viability forecasting results compared to the state-of-the-art models including a multitask learning model, a general path model, and a general linear model. The proposed model provides satisfactory kidney viability forecasting accuracy because it quantifies the degradation information from trajectory of a viability loss path. It transfers knowledge of common effects from all kidneys and identifies individual effects of each kidney. This method can be readily extended to other decision-making scenarios in kidney transplantation to improve overall assessment performance. For example, analytical generalizations gained by modeling have been validated based on needle biopsy data targeting the improvement of tissue extraction accuracy. The proposed model applied in multiple kidney assessment processes in transplantation can potentially reduce the kidney discard rate by providing effective kidney matching decisions. Thus, the increased kidney utilization rate will benefit more patients and prolong their lives.

Keywords: data-driven decision-making, multitask general path model, kidney assessment in transplantation, kidney viability forecasting, needle biopsy

1. INTRODUCTION

Kidney transplantation is the best treatment option for patients who have kidney failure, because it can create the opportunity of living a longer and healthier life [1]. However, the demand for transplantable kidneys far exceeds the supply. Currently, <25% of total patients on the national kidney waiting list will ever receive a donated kidney; the median waiting time for a kidney is 3.6 years [2]. Even worse, the pool of kidney donors has deceased in recent years [3]. To match the donor kidney and recipient, United Network for Organ Sharing (UNOS) makes the decision based on three factors: distance, biological matching, and urgency [4]. However, the viability loss during transportation has never been considered as a key factor in decision-making and loss of viability may lead to >20% kidney discard rate [5]. Moreover, the lack of accurate methods to assess kidney viability causes geographic variation in kidney discard rate, with more discards in areas that have long distances between donated organs and potential recipients [6]. Therefore, a method of forecasting kidney viability is needed to support the decision-making to find the best recipient for kidney transplantation.

Forecasting kidney viability is very challenging for two reasons: 1) surgeons need to assess donor organ viability, but there are no universal guidelines or procedures for objectively and reliably assessing viability [7]. Currently, two methods, visual inspection and biopsy are used [8], and both are either subjective and/or invasive. As a result, the assessment results are very likely to be inconsistent for different surgeons due to different experience, research and education backgrounds; 2) kidneys lose viability rapidly once they are procured, and the degradation rate is individual-specific [9]. Because assessment cannot be performed during preservation or transportation, forecasting kidney viability needs to depend on historical data (i.e., how long can most kidneys remain viable in static cold storage, based on 1- and 5-year graft survival data). However, how to determine the information to be transferred from historical data of existing similar-but-non-identical kidneys to a new kidney is not a precise science and many donor factors and preservation conditions can affect viability data.

The objective of this research is to develop and test a method to forecast kidney viability during preservation by transferring knowledge from similar-but-non-identical kidneys to support decision-making of donor kidneys and recipients matching for predicting the best-fit patient. Moreover, to improve the overall kidney assessment process by extracting tissues more accurately from target areas in the needle biopsy process, the location and number of post-procurement biopsies can be optimized to reduce potential tissue damage due to repetitive biopsy procedures. Therefore, in this research, we formulate the viability forecasting problem as a multitask general path model (MT-GPM), where each task is one pre-defined region on the kidney, based on the vascular structure of the kidney, and the reality that a partial kidney can be transplanted for children [10]. This model incorporates advantages over multitask learning models and general path models by decomposing model coefficients into common effects and individual effects. Therefore, the heterogeneity across tasks can be captured to improve the forecasting accuracy while the commonality is derived from the aggregate of all tasks. More specifically, common effects of all similar-but-non-identical kidneys quantified by shared information are used to depict the trajectory of viability loss, and individual effects are used to identify the viability loss path difference of each kidney. The proposed method is validated by porcine kidney experiments using biopsy scores as the metric of the viability loss path. Compared to benchmark models including a multitask learning model (MTL) [11], a general path model (GPM) [12], and a general linear model (GLM) [13], MT-GPM has the best forecasting accuracy so that it can support decision-making.

The proposed MT-GPM can be applied to two procedures to improve the overall kidney assessment performance. Firstly, it can be used to accurately forecast kidney viability aiming to help the kidney matching decision-making by selecting recipients that are more urgent but at greater distance, based on the estimation of viability loss. Secondly, it can be applied to needle biopsy in forecasting the magnitude of deformation path of needle biopsies used to support biopsy-based decision-making.

2. RELATED WORK

There is a substantial body of research on kidney preservation and biopsy methods. Simple cold storage (CS), while widely used and effective, defines the preservation boundaries to generally less than 24 hours [14]. To prevent organ

deterioration, the kidney preservation process has been improved by using machine perfusion systems (MPS), typically with flush solutions that buffers harsh molecular conditions that evolve during ischemia [15, 16]. On measure, MPS is the better solution for organ preservation because it significantly reduces the risk of delayed kidney function and potentially allows evaluating the viability of kidney before transplantation [17, 18]. Despite MPS offering a reliable platform for kidney preservation, in its current available iterations, it does not provide kidney viability measurements. To assess organ viability, biopsy is the most widely used method [19]. Biopsy is most accurate (predictive) when used with quantitative analysis of whole-side tissue images and sampling of several sites [20, 21]. Moreover, biopsy results are used extensively to predict organ quality for 'non-standard' donors; approximately 75% of extended criteria donor kidneys (ECD) are biopsied [22]. However, repeating biopsies may damage the kidney tissue permanently and this may potentially affect the utilization rate [23]. With the availability of MPS as platform and biopsy as validation method, non-invasive methods can be used effectively to assess kidney viability. For example, computer-assisted diagnosis (CAD) methods can be applied to offer a quantitative diagnosis of the viability of the kidney. CAD has become one of the major research areas in medical diagnosis due to high speed, accuracy, and reliability [24]. Some researchers have used machine learning to predict the outcome of kidney transplantation, but the performance did not meet expectations [25]. Although machine learning is becoming increasingly sophisticated and more widely used, it requires a large sample size to keep training and modifying the model [26]. Because medical data is not always available, due to concerns for respecting the privacy of patients and the confidentiality of healthcare data [27, 28], statistical modeling is used to evaluate medical data for classification and prediction of small data sets [29].

To ensure reliability, data-driven degradation models are widely used to predict the useful lifetime of an engineering system or a biomedical system [12]. GPM, a degradation path model, was first proposed by Lu and Meeker [12, 30] for making inferences about the distribution of failure time for degradation data. One issue is that GPM is very sensitive to outliers. The unit-to-unit variability was modeled by adding random effects in the degradation model, and then developed with Bayesian updating to adjust the model performance [31, 32]. However, although the Markov Chain Monte Carlo (MCMC) procedure can be analytically used for intractable joint posterior density function estimation, it is computationally intensive for complex model structures. Zhou et al. introduced a degradation model in a nonparametric modeling framework, instead of using a common parametric model with random effects [33]. The limitation is that the non-parametric model is less efficient than the parametric model in some cases. Bagdonavicius et al. [34] investigated multiple failure modes in the degradation model. Hong et al. [35, 36] proposed dynamic covariates in the degradation model, using a linear random effects model, and then using a nonlinear random effects model, but the general path model would only have good performance when the degradation curves are monotone or nearly monotone. GPM has been adopted from industrial reliability scenarios, assuming all the deployed equipment are identical in quality. However, the kidneys (biological 'machines') are heterogeneous due to size, shape, donor age, and molecular condition. Thus, GPM may not be applicable or appropriate to apply to kidney viability forecasting. Recently, a spatial-temporal degradation model was proposed to describe complex degradation items which show different degradation patterns at different locations [37, 38]. This model assumes the propagation of degradation at a constant speed. However, kidney heterogeneity may significantly affect outcomes from the use of this model.

Multi-task learning (MTL) is an approach for transferring information and improving generalization by using related tasks as an inductive bias [11]. The models define commonality simultaneously from all the tasks and exploit differences across tasks, so that acquired information can improve modeling performance [11]. MTL is widely used in many real-world applications, including medical and clinical fields. In organ transplantation, MTL logistic regression has been conducted to predict the liver viability based on infrared imaging data [39]. However, this method is not able to forecast the liver viability at a future time. Data Shared Lasso (DSL) [40] investigated the continuum between individual models for each group and one model for all groups, but DSL is not a forecasting solution and is not similar to MTL. In summary, the needs of kidney viability forecasting, and the heterogeneity of different kidneys gave us the motivation to integrate GPM and MTL modeling methods to investigate the viability loss path. Thus, we propose MT-GPM to forecast the viability of heterogeneous kidneys, in order to improve the kidney matching decision-making process, and extended MT-GPM to needle biopsy forecasting as another validation example.

3. MATERIALS AND METHODS

In this research, we apply a data-driven model to forecast kidney viability, in terms of biopsy scores to support matching decision-making in order to reduce kidney discard rate. The assumptions of the model are: 1) the kidney viability loss path can be described in a general function form, for instance, the viability loss path in Case Study 1 can be approximated by quadratic functions based on historical data; 2) the kidney viability loss paths are similar-but-non-identical, because they share commonality among locations in different kidneys while maintaining individual differences.

3.1 The Proposed Method

We propose the MT-GPM that treat each location of four pre-defined locations on the kidney as one task. The biopsy score y of i-th viability loss path for task g at time t_j is given by:

$$y_{(i,g,t_{i,g,j})} = \eta\left(t_j, \boldsymbol{\beta}_{(i,g)}\right) + \varepsilon_{(i,g,t_{i,g,j})}, \tag{1}$$

where t_j is the time of the j-th biopsy score as measurement, $j = 1, \dots, J$; $\varepsilon_{(i,g,t_{i,g,j})}$ is the measurement error which is assumed to independently and identically distribute as $N(0, \sigma_\varepsilon^2)$; $\eta_{(i,g)}$ is approximated quadratic path from assumption of the general form for i-th viability loss path in g-th task parameterized by $\boldsymbol{\beta}_{(i,g)}, i = 1, \dots, I, g = 1, \dots, G$. In each task, we have I viability loss paths representing the number of kidneys used in experiment. Because kidneys are similar-but-non-identical by sharing commonality and having heterogeneity. Therefore, we further decompose the model coefficient $\boldsymbol{\beta}_{(i,g)}$ to:

$$\boldsymbol{\beta}_{(i,g)} = \boldsymbol{\phi}_{(g)} + \boldsymbol{\theta}_{(i,g)}, \tag{2}$$

where $\boldsymbol{\phi}_{(g)} \in \mathbb{R}^{p(I+1)\times 1}$, p is the number of common-effects parameters representing commonality for all replicates; $\boldsymbol{\theta}_{(i,g)}$ is a vector of the i-th viability loss path individual-effects parameters in the g-th task representing heterogeneity. Thus, common-effects parameters are transferring knowledge from all kidneys, and individual-effects parameters are identifying the kidney differences. Motivated by data shared lasso [40], our model is assumed as second order polynomial model with $\boldsymbol{x}_{(i,g,t_{i,g,j})}$ representing timestamp of i-th viability loss path in g-th task as model inputs, so the $y_{(i,g,t_{i,g,j})}$ can be generated as:

$$y_{(i,g,t_{i,g,j})} = \boldsymbol{x}_{(i,g,t_{i,g,j})}(\boldsymbol{\phi}_{(g)} + \boldsymbol{\theta}_{(i,g)}) + \varepsilon_{(i,g,t_{i,g,j})}. \tag{3}$$

The model parameters $\boldsymbol{\beta}_{(i,g)} = \boldsymbol{\phi}_{(g)} + \boldsymbol{\theta}_{(i,g)}$ can be estimated by optimizing Equation (4), and r_g is given to control the amount of pooling, and we recommend $r_g = 1$ in this model:

$$\left(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}\right) \coloneqq \underset{\boldsymbol{\phi},\boldsymbol{\theta}}{\operatorname{argmin}} \frac{1}{2} \sum_{i,g,j} \left(y_{(i,g,t_{i,g,j})} - \boldsymbol{x}_{(i,g,t_{i,g,j})}\left(\boldsymbol{\phi}_{(g)} + \boldsymbol{\theta}_{(i,g)}\right)\right)^2 + \lambda\left(\left\|\boldsymbol{\phi}_{(g)}\right\|_1 + \sum_{g=1}^{i} r_g \left\|\boldsymbol{\theta}_{(i,g)}\right\|_1\right), \tag{4}$$

where the first part is the least square loss function to ensure the forecasting accuracy, and the second part is lasso penalty that penalizes on the sparsity of the coefficients [41]. Directly optimizing Equation (4) by using convex programming methods will leads to high computation workload and non-satisfactory time latency in forecasting due to multiple non-differentiable l_1 norms of model coefficients. Denoting $\boldsymbol{W}_g = (\boldsymbol{\phi}_g^T, \boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \dots, \boldsymbol{\theta}_I^T)^T \in \mathbb{R}^{p(I+1)\times G}$ as the model coefficient vector, and $\boldsymbol{\varepsilon}_{ig}$ as the residual term, we show that optimizing Problem (4) can be approximated by optimizing Equation (6) by defining $\boldsymbol{Z}_g$ as:

$$\boldsymbol{Z}_g = \begin{pmatrix} \boldsymbol{X}_{1g} & r_1\boldsymbol{X}_{1g} & 0 & \cdots & 0 \\ \boldsymbol{X}_{2g} & 0 & r_2\boldsymbol{X}_{2g} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ \boldsymbol{X}_{Ig} & 0 & 0 & \cdots & r_I\boldsymbol{X}_{Ig} \end{pmatrix}, \tag{5}$$

where we separate the common effects and individual effects as the first column representing common effects and each $r_i \boldsymbol{X}_{ig}$ is the covariates to estimate individual effects in one task among different viability loss paths.

$$\widehat{\boldsymbol{W}} = \underset{\boldsymbol{W}}{\operatorname{argmin}} \sum_{i=1}^{g} \left\| \boldsymbol{Z}_g \boldsymbol{W}_g - \boldsymbol{Y}_g \right\|_{\mathrm{F}}^2 + \rho \| \boldsymbol{\kappa} \circ \boldsymbol{W} \|_{2,1}, \tag{6}$$

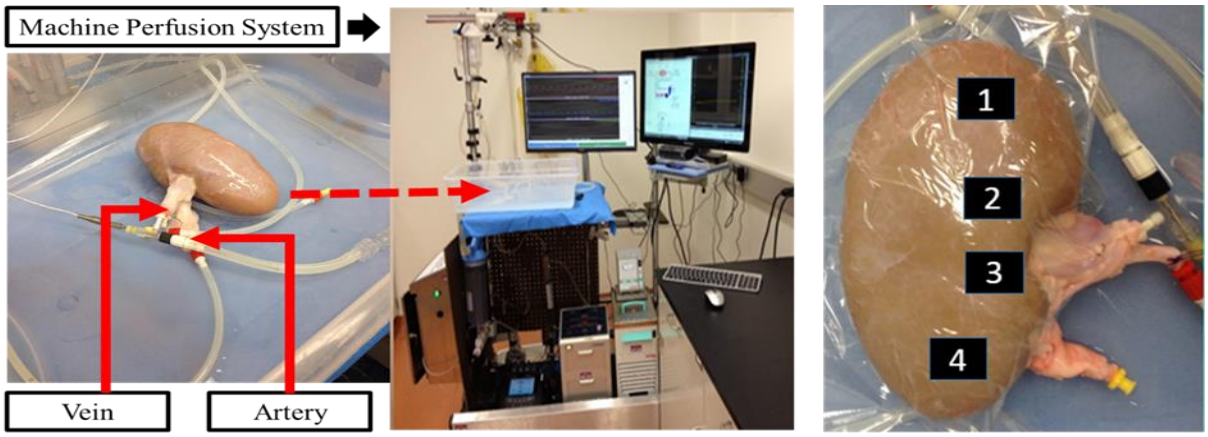
where $\boldsymbol{\kappa} \in \mathbb{R}^{p(I+1)\times G}$ is a known weights matrix, and the weights are data dependent $\boldsymbol{k} = 1/\left|\widehat{\boldsymbol{\beta}}_{OLS}\right|^{\gamma}$. If the weight is data-dependent, then the weighted lasso can have the oracle properties [42]. The $l_{2,1}$-norm regularized regression model is proposed to encourage multiple predictors to share similar sparsity patterns, which can quantify between-task similarities (i.e., common effects) and within task similarities (i.e., individual effects). The convexity of Equation (6) admits a globally optimal solution guaranteed by K.K.T. condition [43]; and ρ is the tuning parameter to control the model sparsity which leads to lower sample requirement [44]. The adaptive lasso achieves the oracle properties, and the shrinkage leads to a near-minimax-optimal estimator [42]. Therefore, it is applied to further improve the forecasting accuracy by adding weights to penalize different coefficients in the $l_{2,1}$ penalty. The optimization problem showing in Equation (6) can be efficiently solved by accelerated gradient descent [45].

We use leave-one-kidney-out cross validation method to validate the model. Five kidneys out of six are used iteratively as training data and the remaining one as testing data. The initial model forecasts the viability by only using common effects, and then the model is updated by adding testing data one at a time to incorporate individual effects. The remaining test data is used to test forecasting accuracy.

In summary, we used a parametric model to represent the viability loss path. MT-GPM can forecast kidney viability in transplantation by simultaneously learning knowledge from all kidneys while identifying the individual differences. It can be extended to other domains, where the response measurements can be depicted by certain functions. Two case studies will be introduced to validate the generality of this method.

3.2 Case Study 1: Kidney Viability Forecasting

To validate the proposed method, fresh porcine kidneys were procured and perfused in MPS during preservation. Pigs have many similar characteristics to humans in terms of heart rate, blood pressure, function and size, making use of the results from measurements of their kidneys readily translatable [46]. Six porcine kidneys from young female pigs were collected from local abattoir and were flushed with physiologic saline to remove residual blood. Next, the kidneys were transported back to the laboratory in a portable cooler, and they were immediately connected to a MPS [47] via renal artery and vein as below the Figure 1(a) after arrival. Once the kidney was connected, more phosphate-buffered physiologic saline was infused, and a closed-loop perfusion system was established to mimic renal blood perfusion. The perfusion time for each kidney was 12 hours in order to allow ample deterioration of the kidney over time, and a 12-hour time span is a similar comparison to the time transported from donor to recipient in reality. To accurately assess the viability of kidneys, each kidney was divided into four locations and biopsied as Figure 1(b). We pre-defined each location as one task because of the vascular structure of the kidney. The locations closest to the artery and vein were perfused differently than the outer regions. Therefore, two locations, say, Locations 1 and 4 in Figure 1(b) were at the extremities because they were at far-end of the artery and vein. In contrast, Locations 2 and 3 are more well-perfused regions.



(a) Machine perfusion system [47] (b) Kidney biopsy locations

Figure 1: Kidney viability experiment

The first biopsy was taken as soon as the kidney was connected to the machine, and then three more biopsies were procured with 4-hour interval at 4th hour, 8th hour, and 12th hour. We intended to verify the viability loss during perfusion time but limit the number of biopsies, because repeating biopsies too many times at one location causes permanent damage to the kidney [23]. The time and location of biopsies for each kidney were randomized as Table 1 in order to reduce bias. Each biopsy was taken with a punch hole, and then was placed in 10% neutral-buffered formalin to preserve the samples until evaluated histologically. The biopsy defect on the kidney was filled with surgical Gelfoam (Pfizer, Groton, CT) in order to prevent the damaging region expanding to other locations. All the biopsy samples were blindly graded by an experienced pathologist and ranked on a scale of 1, 2, and 3 representing bad, good, and excellent. Good and excellent biopsy scores signified the kidney was considered transplantable, otherwise they will be discarded. The biopsy scores were summarized in Table 1.

Table 1. Kidney location and biopsy score summary

	Categories	**0 Hour**	**4 Hour**	**8 Hour**	**12 Hour**
Kidney 1	Locations	*4th*	*1st*	*3rd*	*2nd*
	Biopsy Scores	3	2	2	1
Kidney 2	Locations	*1st*	*2nd*	*3rd*	*4th*
	Biopsy Scores	2	2	1	1
Kidney 3	Locations	*1st*	*4th*	*2nd*	*3rd*
	Biopsy Scores	2	2	1	2
Kidney 4	Locations	*4th*	*1st*	*2nd*	*3rd*
	Biopsy Scores	2	1	2	1
Kidney 5	Locations	*1st*	*4th*	*3rd*	*2nd*
	Biopsy Scores	2	1	2	1
Kidney 6	Locations	*2nd*	*4th*	*1st*	*3rd*
	Biopsy Scores	2	3	2	2

Based on the information from above table, a 4×4 spatial-temporal biopsy score response surface representing biopsy scores of each kidney was generated. A monotonic Gaussian process model [48] was used to fit the data and generate the response surfaces because the deteriorating process is irreversible [49]. Figure 2 shows the six response surfaces with x, y, z-axis representing location, time and biopsy score for the six kidneys. Moreover, we assumed that the biopsy scores had spatial-temporal correlation. However, the sample size was limited. Therefore, interpolation was employed to increase the sample frequency from every 4 hours

to every 30 minutes based on the response surfaces. As a result, six spatial-temporal biopsy score response surfaces were extended to 4×24 representing biopsies scores of each location every 30 minutes.

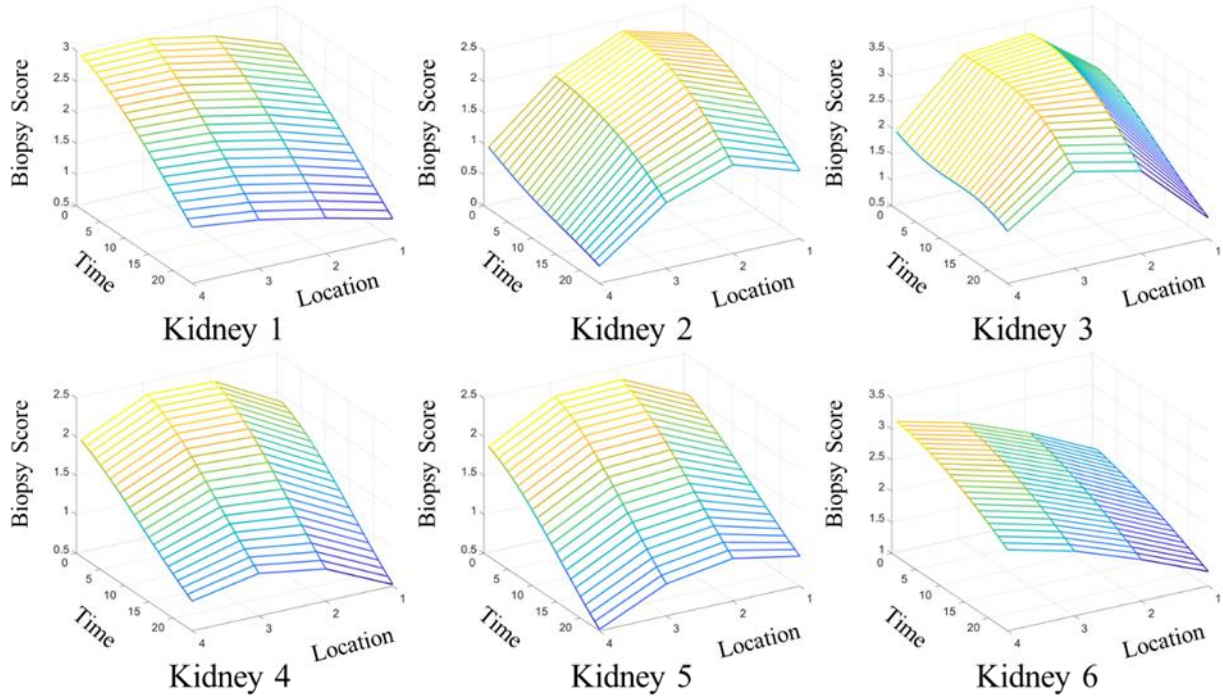
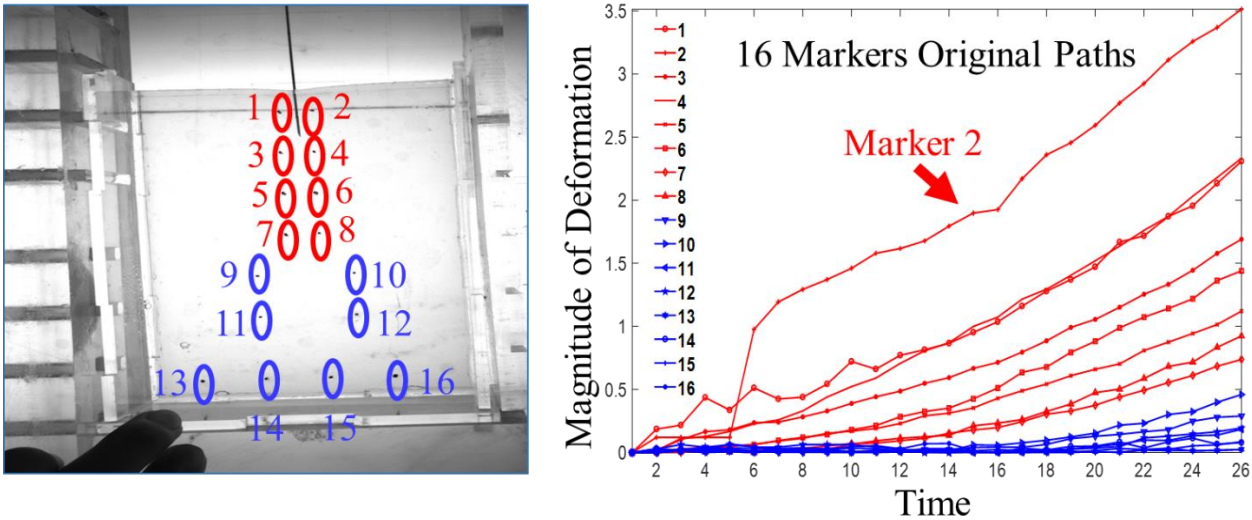


Figure 2: Spatial-temporal biopsy score response surfaces

3.3 Case Study 2: Needle Biopsy Insertion Forecasting

The proposed model is generalized to apply in the other kidney assessment process within a certain scope, in which a certain function form can roughly describe the tissue deformation during biopsy needle insertion as response measurements. Needle biopsy, as a common medical tool, is applied as the assessment in transplantation to verify the viability of the kidney [18]. Some researchers investigated needle-tissue interaction mechanics to develop a theoretical model that quantifies the deformation of tissues and needles [50]. Since the magnitude of phantom deformation shows a path, the proposed method can be applied. To forecast the magnitude deformation path, the biopsy accuracy can be improved with online feedback guiding needle insertion process to obtain tissues from the target region. Moreover, this application can be potentially used in medical training and biopsy education by providing online feedback on tissue deformation during the needle insertion process.

The experiment setup shows as Figure 3(a). The needle is placed above the phantom surface at 10 cm, and then inserted into the phantom. A polyvinyl alcohol hydrogel (PVA-H)-based phantom was used to simulate the human tissue because its transparency made it easy to record the deformation of the tissue. The PVA-H phantoms were made of dimethyl sulfoxide and polyvinyl alcohol. The Young's modulus of phantom was between 10Kpa and 50Kpa (similar to liver) and the Poisson ration is 0.49. A CCD camera was fixed in front of the phantom to capture markers inside the tissue, and 16 markers were made to record the displacements as Figure 3(b). The photo sample frequency was two frames per second. In total, 26 photos were collected for analysis, and the displacements of x-axis and y-axis of each maker were calculated by imaging processing. The magnitude of marker deformation was used as the response in this case study. We treated each marker as one task, assuming each location has similar pattern, and four replicate measurements were made in each task.



(a) Experimental setup (b) Magnitude of marker deformation

Figure 3: Needle insertion experiment

4. RESULTS AND DISCUSSIONS

In Case Study 1, we applied the proposed method to analyze kidney viability data and compared forecasting accuracy with benchmark models including MTL, GPM, and GLM by root mean square error (RMSE), which quantifies the difference between forecasted biopsy scores and ground truth. Because MTL, GMP, and GLM have benefits for regularization from cross-task effects, within-task effects, and general random effects, respectively. MT-GPM method decomposes common and individual effects representing cross-task and within-task effects in one modeling approach to improve forecasting accuracy.

The "leave-one-kidney-out" cross validation method is conducted to imitate the kidney viability forecasting by iteratively using five kidneys as training data and the remaining one as testing data. Therefore, in total, six scenarios are generated. MT-GPM can be applied without knowing any information of the new kidney by only using common effects from existing kidneys and forecast the new kidney viability from the "cold start" stage. In contrast, the benchmark models have to rely on retrieving initial data of the testing kidney to conduct forecasting, such as Bayesian approach degradation models [51]. In addition, the proposed MT-GPM limits the required sample size to accurately forecast the kidney viability (sample size ≤ 3). Each biopsy score is added to the training data following time sequence, and the model is updated by transferring information from new added biopsy scores. As more information about the new kidney is obtained, the model will perform better by considering individual effects. Therefore, the model forecasting will be performing more accurately as the RMSE continuously decreased. In Figure 4, we conclude that the proposed MT-GPM has the smallest overall prediction error compared to other benchmark models, and all six leave-one-kidney-out cross validation scenarios support the same conclusion. The RMSE are rapidly decreased to <0.3 in most of the six scenarios, because the common effects are so effective in model prediction that only few biopsy scores are needed to accurately estimate the model with individual effects. In Leave-1st –out (L1O), L4O and L5O scenarios, the increasing RMSEs of the first two or three timestamps show in Figure 4 is due to not having enough degrees of freedom for the quadratic model. Therefore, fluctuation exists when adding the first several biopsy scores.

By applying MT-GPM model to kidney viability forecasting, the kidney matching decision can be improved by the estimation of kidney viability loss in preservation. Moreover, the best biopsy region of the kidney can be recommended for the pre-transplantation biopsy. For instance, the lowest ranked biopsy score region can be recommended to make sure the lowest biopsy score region is viable. Therefore, the entire kidney is transplantable. In contrast, the highest ranked biopsy score region can be recommended for biopsy if the overall viability is forecast to be bad. Thus, the

practitioner can assure that the kidney should be discarded if the highest scored region is confirmed to be inviable after biopsy. As a result, the number of biopsies can be minimized to reduce the potential damaging effects of repeating biopsies.

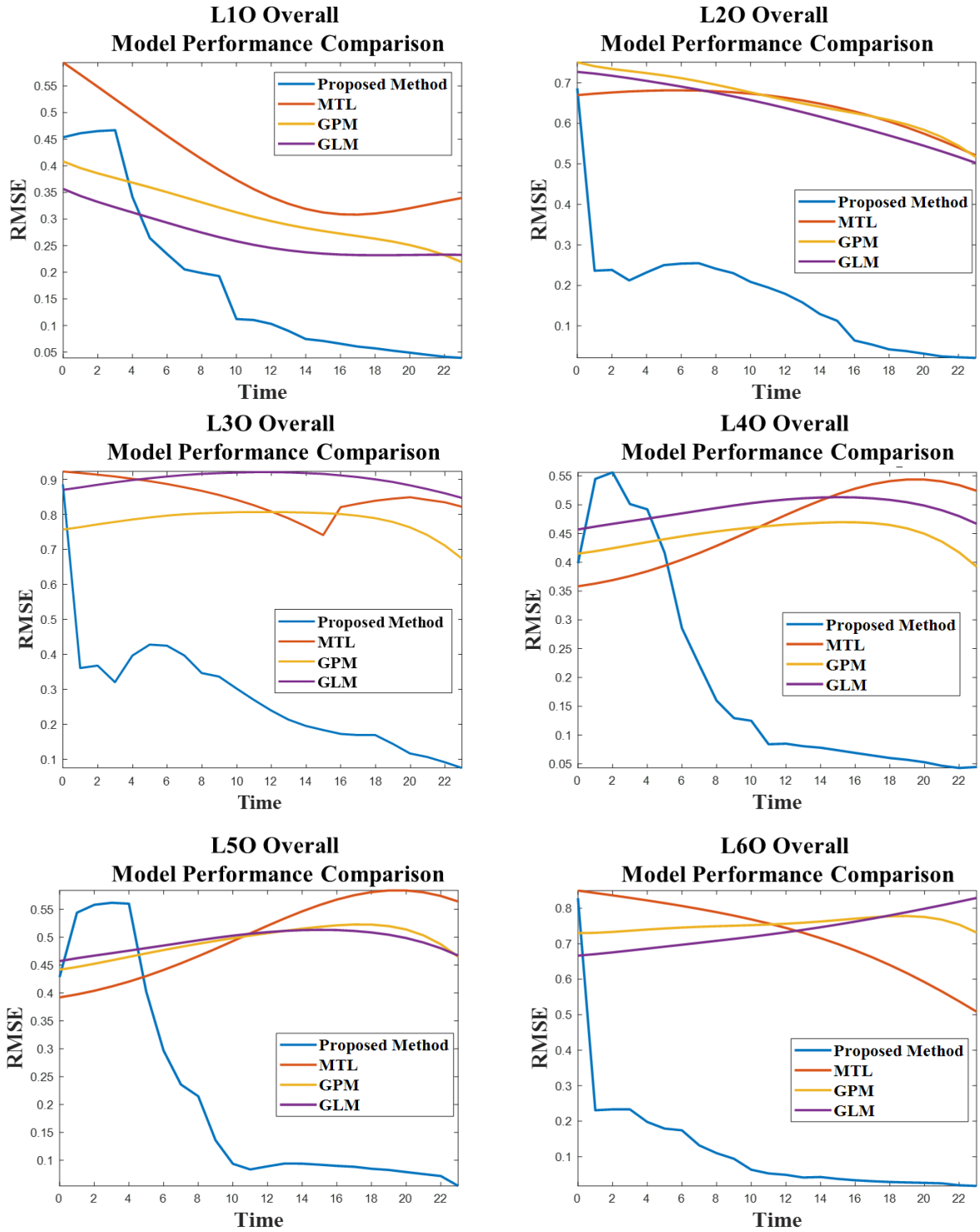


Figure 4: Forecasting accuracy (RMSE) of kidney viability case

MT-GPM has the best forecasting accuracy by distinguishing common effects and individual effects in one-step modeling approach, and it can potentially improve the effectiveness of kidney matching decision by forecasting the viability. In Appendix A, we conclude that the model accuracy has increased by updating the model coefficients as more biopsy scores from the testing kidney are added to the training data. In addition, the proposed model is generalized to not only benefit the kidney matching decision but also support other decision-making scenarios when the scope is such that a certain function form can roughly describe the response measurements, such as the needle biopsy process to improve the overall kidney assessment accuracy.

In Case Study 2, Figure 3(a) shows two groups of markers: red markers at top rows are close to the needle path having larger deformation and blue markers at bottom rows are further from the needle path with small deformation. Figure 3(b) shows the true magnitude of maker deformation paths. Therefore, only mark 1 to 8 will be used for analytics since the magnitudes of deformation of mark 9 to 16 are too small. In addition, they cannot be represented by quadratic model. The model forecasting RMSEs are summarized in Figure 5. Similar to Case Study 1, the proposed MT-GPM outperforms the benchmark models in terms of forecasting accuracy. As the needle is repetitiously inserted into the phantom, more deformation path information was added to training data, so the model accuracy was significantly increased. Figure 5 shows that the proposed method in Leave-2nd -marker-out scenario has a different trajectory of deformation path compared to the rest, because the deformation path of marker 2 is significantly different from others showing in Figure 3(b). However, MT-GPM can still address the heterogeneity issue and outperform the benchmark models. The proposed method can support needle biopsy decision-making, because it can learn the tissue deformation at different locations and interactive forces between the needle and tissue in order to provide meaningful and potentially real-time feedback for guiding needle insertion. The accuracy of pre-transplantation biopsy can be potentially increased to improve the overall transplantation performance. Model diagnostic and discussion of both case studies are in Appendix B.

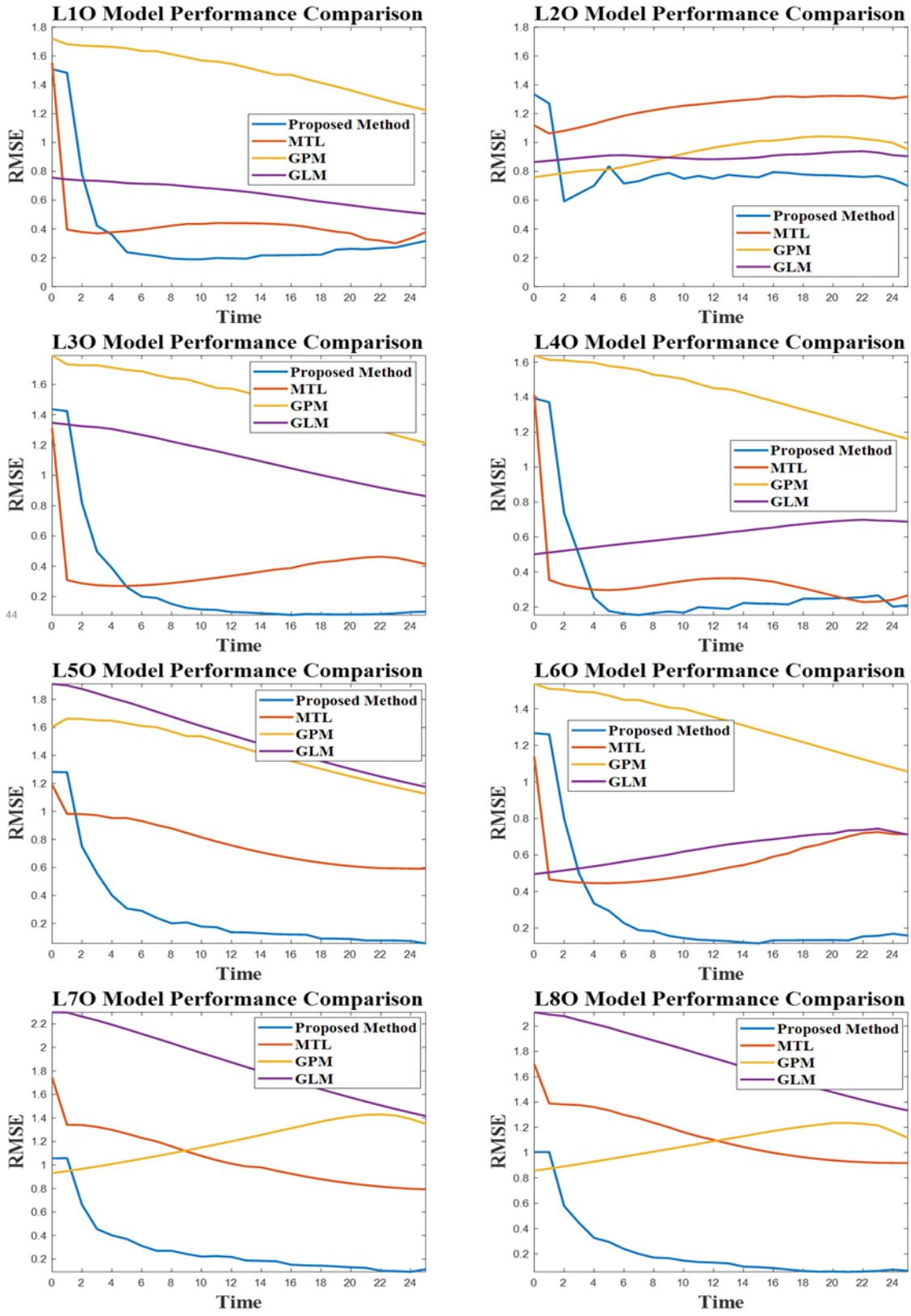


Figure 5: Forecasting accuracy (RMSE) of needle biopsy case

5. CONCLUSIONS

In this research, we proposed the MT-GPM to forecast kidney viability during preservation in order to support kidney matching decision-making. The case study shows that the proposed method has satisfactory accuracy in forecasting kidney viability, and it outperforms the benchmark models including MTL, GPM, and GLM. The proposed method integrates MTL and GPM in a one-step modeling approach by learning the commonality from all tasks and capturing the heterogeneity by extrapolating each viability loss path. As a result, the forecasting performance can be improved. To learn the viability loss path of each kidney, it helps the kidney matching decision to be more efficient in providing customized kidney transportation services because the matching can be extended to facilitate greater distance but recognize the needs and potential success for recipients that are more urgent. In addition, this method can suggest the optimized location for conducting pre-transplantation biopsies and potentially saves more currently discarded kidneys for transplantation. The generality of the proposed method has been validated by the second case study to improve overall kidney assessment performance in transplantation by providing online decision-making support in the pre-transplantation needle biopsy process.

Since MT-GPM is a new technique, follow-up studies need to be investigated in the future. For example, the kidney can be replaced with other organs, such as liver and lung, to test the universality of the proposed method. In addition, the "cold start" modeling performance is similar to the benchmark models. To address this issue, profile information, such as donor biological differences and medical records will be considered as covariates adding to the MT-GPM to further improve prediction performance in "cold start" scenario.

ACKNOWLEDGMENTS

The authors acknowledge Joanna Appugliese, Brandon Brown, Ben Codiroli, and Patrick Johns at the Grado Department of Industrial and Systems Engineering at Virginia Tech for their help in data collection and pre-processing for kidney experiments, and Chip Aardema for experimental support. The authors also acknowledge Murong Li and Pengyue Hu at Zhejiang University for their help in data collection and pre-processing for needle biopsy experiments.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

QL and XC are co-first authors. All authors contributed to the design, experiments, and writing of the manuscript.

REFERENCES

[1] Rebecca Ireland. 2012. Making kidneys count. *Nature Reviews Nephrology* 8, no. 6 (2012): 311-311.

[2] National Kidney Foundation, Organ Donation and Transplantation Statistics. 2019. Retrieve from https://www.kidney.org/news/newsroom/factsheets/Organ-Donation-and-Transplantation-Stats (Accessed 20 September 2019).

[3] Jesse D. Schold and Dorry L. Segev. 2012. Increasing the pool of deceased donor organs for kidney transplantation. *Nature Reviews Nephrology* 8, no. 6 (2012): 325-331.

[4] Oscar Bronsther, John J. Fung, Andreas Izakis, David Van Thiel, and Thomas E. Starzl. 1994. Prioritization and organ distribution for liver transplantation." *Jama* 271, no. 2 (1994): 140-143.

[5] Shruti Mittal, Anna Adamusiak, Catherine Horsfield, Ioannis Loukopoulos, Nikolaos Karydis, Nicos Kessaris, Martin Drage, Jonathon Olsburgh, Christopher JE Watson, and Chris J. Callaghan. 2017. A re-evaluation of discarded deceased donor kidneys in the UK: are usable organs still being discarded?. *Transplantation* 101, no. 7 (2017): 1698-1703.

[6] Peter P. Reese, Meera N. Harhay, Peter L. Abt, Matthew H. Levine, and Scott D. Halpern. 2016. New solutions to reduce discard of kidneys donated for transplantation. *Journal of the American Society of Nephrology* 27, no. 4 (2016): 973-980.

[7] Beat Moeckli, Pamela Sun, François Lazeyras, Philippe Morel, Solange Moll, Manuel Pascual, and Léo H. Bühler. 2019. Evaluation of donor kidneys prior to transplantation: an update of current and emerging methods. *Transplant International* 32, no. 5 (2019): 459-469.

[8] Qing Lan, Ran Jin, and John L. Robertson. Quantitative and qualitative evaluation for organ preservation in transplant. In *IIE Annual Conference. Proceedings*, p. 2229. Institute of Industrial and Systems Engineers (IISE), 2015.

[9] Oguzhan Alagoz, Andrew J. Schaefer, and Mark S. Roberts. 2009. Optimizing organ allocation and acceptance. In *Handbook of optimization in medicine*, pp. 1-24. Springer, Boston, MA.

[10] Kota Takahashi, Hiroshi Kawaguchi, Takashi Yagisawa, Kazunari Tanabe, Hayakazu Nakazawa, Motoshi Hattori, Hiroshi Toma et al. 1993. Partial kidney transplantation: a successful kidney transplantation in a child with severe cardiac failure by surgical mass reduction of an adult donor kidney. *Transplant International* 6, no. 3 (1993): 173-175.

[11] Rich Caruana. 1997. Multitask learning. *Machine learning* 28, no. 1 (1997): 41-75.

[12] C. Joseph Lu and William O. Meeker. 1993. Using degradation measures to estimate a time-to-failure distribution. *Technometrics* 35, no. 2 (1993): 161-174.

[13] John A. Nelder and Robert W.M. Wedderburn. 1972. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)* 135, no. 3 (1972): 370-384.

[14] Ronghai Deng, Guangxiang Gu, Dongping Wang, Qiang Tai, Linwei Wu, Weiqiang Ju, Xiaofeng Zhu, Zhiyong Guo, and Xiaoshun He. 2013. Machine perfusion versus cold storage of kidneys derived from donation after cardiac death: a meta-analysis. *PLoS One* 8, no. 3 (2013): e56368.

[15] Mark-Hugo J. Maathuis, Steffen Manekeller, Arjan van der Plaats, Henri G.D. Leuvenink, Nils A't Hart, A. Bastiaan Lier, Gerhard Rakhorst, Rutger J. Ploeg, and Thomas Minor. 2007. Improved kidney graft function after preservation using a novel hypothermic machine perfusion device. *Annals of surgery* 246, no. 6 (2007): 982-991.

[16] A. Kwiatkowski, M. Wszola, M. Kosieradzki, R. Danielewicz, K. Ostrowski, P. Domagala, W. Lisik et al. 2007. Machine perfusion preservation improves renal allograft survival. *American journal of transplantation* 7, no. 8 (2007): 1942-1947.

[17] Cyril Moers, Jacqueline M. Smits, Mark-Hugo J. Maathuis, Jürgen Treckmann, Frank van Gelder, Bogdan P. Napieralski, Margitta van Kasterop-Kutz et al. 2009. Machine perfusion or cold storage in deceased-donor kidney transplantation. *New England Journal of Medicine* 360, no. 1 (2009): 7-19.

[18] Shawn D. St Peter, Charles J. Imber, and Peter J. Friend. 2002. Liver and kidney preservation by perfusion. *The Lancet* 359, no. 9306 (2002): 604-613.

[19] Don C. Rockey, Stephen H. Caldwell, Zachary D. Goodman, Rendon C. Nelson, and Alastair D. Smith. 2009. Liver biopsy. *Hepatology* 49, no. 3 (2009): 1017-1044.

[20] Marco Fiorentino, Davide Bolignano, Vladimir Tesar, Anna Pisano, Wim Van Biesen, Giovanni Tripepi, Loreto Gesualdo, and ERA-EDTA Immunonephrology Working Group. 2016. Renal biopsy in 2015-from epidemiology

to evidence-based indications. *American Journal of Nephrology* 43, no. 1 (2016): 1-19.

[21] Sonal Kothari, John H. Phan, Todd H. Stokes, and May D. Wang. 2013. Pathology imaging informatics for quantitative analysis of whole-slide images. *Journal of the American Medical Informatics Association* 20, no. 6 (2013): 1099-1108.

[22] Anna J. Dare, Gavin J. Pettigrew, and Kourosh Saeb-Parsy. 2014. Preoperative assessment of the deceased-donor kidney: from macroscopic appearance to molecular biomarkers. *Transplantation* 97, no. 8 (2014): 797-807.

[23] Harald Mischak. 2015. Pro: urine proteomics as a liquid kidney biopsy: no more kidney punctures!. *Nephrology Dialysis Transplantation* 30, no. 4 (2015): 532-537.

[24] Miles N. Wernick, Yongyi Yang, Jovan G. Brankov, Grigori Yourganov, and Stephen C. Strother. 2010. Machine learning in medical imaging. *IEEE signal processing magazine* 27, no. 4 (2010): 25-38.

[25] Julia Lasserre, Steffen Arnold, Martin Vingron, Petra Reinke, and Carl Hinrichs. 2012. Predicting the outcome of renal transplantation. *Journal of the American Medical Informatics Association* 19, no. 2 (2012): 255-262.

[26] June-Goo Lee, Sanghoon Jun, Young-Won Cho, Hyunna Lee, Guk Bae Kim, Joon Beom Seo, and Namkug Kim. 2017. Deep learning in medical imaging: general overview. *Korean journal of radiology* 18, no. 4 (2017): 570-584.

[27] Randolph C. Barrows Jr and Paul D. Clayton. 1996. Privacy, confidentiality, and electronic medical records. *Journal of the American Medical Informatics Association* 3, no. 2 (1996): 139-148.

[28] Yaorong Ge, David K. Ahn, Bhagyashree Unde, H. Donald Gage, and J. Jeffrey Carr. 2013. Patient-controlled sharing of medical imaging data across unaffiliated healthcare organizations. *Journal of the American Medical Informatics Association* 20, no. 1 (2013): 157-163.

[29] Margaret S. Pepe. 2003. *The statistical evaluation of medical tests for classification and prediction*. Medicine.

[30] William Q. Meeker, Luis A. Escobar, and C. Joseph Lu. 1998. Accelerated degradation tests: modeling and analysis. *Technometrics* 40, no. 2 (1998): 89-99.

[31] Suk-Joo Bae and Paul H. Kvam. 2004. A nonlinear random-coefficients model for degradation testing. *Technometrics* 46, no. 4 (2004): 460-469.

[32] Jamie Coble and J. Wesley Hines. 2011. Applying the general path model to estimation of remaining useful life. *International Journal of Prognostics and Health Management* 2, no. 1 (2011): 71-82.

[33] Rensheng R. Zhou, Nicoleta Serban, and Nagi Gebraeel. 2011. Degradation modeling applied to residual lifetime prediction using functional data analysis. *The Annals of Applied Statistics* (2011): 1586-1610.

[34] Vilijandas Bagdonavičius, Algimantas Bikelis, and Vytautas Kazakevičius. 2004. Statistical analysis of linear degradation and failure time data with multiple failure modes. *Lifetime data analysis* 10, no. 1 (2004): 65-81.

[35] Yili Hong, Yuanyuan Duan, William Q. Meeker, Deborah L. Stanley, and Xiaohong Gu. 2015. Statistical methods for degradation data with dynamic covariates information and an application to outdoor weathering data. *Technometrics* 57, no. 2 (2015): 180-193.

[36] Zhibing Xu, Yili Hong, and Ran Jin. 2016. Nonlinear general path models for degradation data with dynamic covariates. *Applied Stochastic Models in Business and Industry* 32, no. 2 (2016): 153-167.

[37] Xiao Liu, Kyongmin Yeo, and Jayant Kalagnanam. 2016. Statistical Modeling for Spatio-Temporal Degradation Data. *arXiv preprint arXiv:1609.07217* (2016).

[38] Ashlee N.F. Versypt, Paul D. Arendt, Daniel W. Pack, and Richard D. Braatz. 2015. Derivation of an analytical solution to a reaction-diffusion model for autocatalytic degradation and erosion in polymer microspheres. *PloS one* 10, no. 8 (2015): e0135506.

[39] Qing Lan, Hongyue Sun, John Robertson, Xinwei Deng, and Ran Jin. 2018. Non-invasive assessment of liver quality in transplantation based on thermal imaging analysis. *Computer methods and programs in biomedicine* 164 (2018): 31-47.

[40] Samuel M. Gross and Robert Tibshirani. 2016. Data Shared Lasso: A novel tool to discover uplift. *Computational statistics & data analysis* 101 (2016): 226-235.

[41] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, no. 1 (1996): 267-288.

[42] Hui Zou. 2006. The adaptive lasso and its oracle properties. *Journal of the American statistical association* 101, no. 476 (2006): 1418-1429.

[43] Geoff Gordon and Ryan Tibshirani. 2012. Karush-kuhn-tucker conditions. *Optimization* 10, no. 725/36 (2012): 725.

[44] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. 2008. Convex multi-task feature learning. *Machine learning* 73, no. 3 (2008): 243-272.

[45] Yurri E. Nesterov. 1983. A method for solving the convex programming problem with convergence rate O (1/k^ 2). In *Dokl. akad. nauk Sssr*, vol. 269, pp. 543-547. 1983.

[46] Marco A. Pereira-Sampaio, Luciano A. Favorito, and Francisco JB Sampaio. 2004. Pig kidney: anatomical relationships between the intrarenal arteries and the kidney collecting system. Applied study for urological research and surgical training. *The Journal of urology* 172, no. 5 (2004): 2077-2081.

[47] Qing Lan, Yifu Li, John L. Robertson, and Ran Jin. 2020. Modeling of pre-transplantation liver viability with spatial-temporal smooth variable selection; manuscript submitted for publication.

[48] Jaakko Riihimäki and Aki Vehtari. 2010. Gaussian processes with monotonicity information. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 645-652. 2010.

[49] M. Zakliczynski, J. Nozynski, H. Zakliczynska, P. Rozentryt, and M. Zembala. 2003. Deterioration of renal function after replacement of cyclosporine with sirolimus in five patients with severe renal impairment late after heart transplantation. In *Transplantation proceedings*, vol. 35, no. 6, pp. 2331-2332. Elsevier, 2003.

[50] Dedong Gao, Yong Lei, Bin Lian, and Bin Yao. 2016. Modeling and simulation of flexible needle insertion into soft tissue using modified local constraints. *Journal of Manufacturing Science and Engineering* 138, no. 12 (2016).

[51] Nagi Z. Gebraeel, Mark A. Lawley, Rong Li, and Jennifer K. Ryan. 2005. Residual-life distributions from component degradation signals: A Bayesian approach. *IIE Transactions* 37, no. 6 (2005): 543-557.

A APPENDICES

A.1 Comparison of Predicted Biopsy Scores and Actual Biopsy Scores

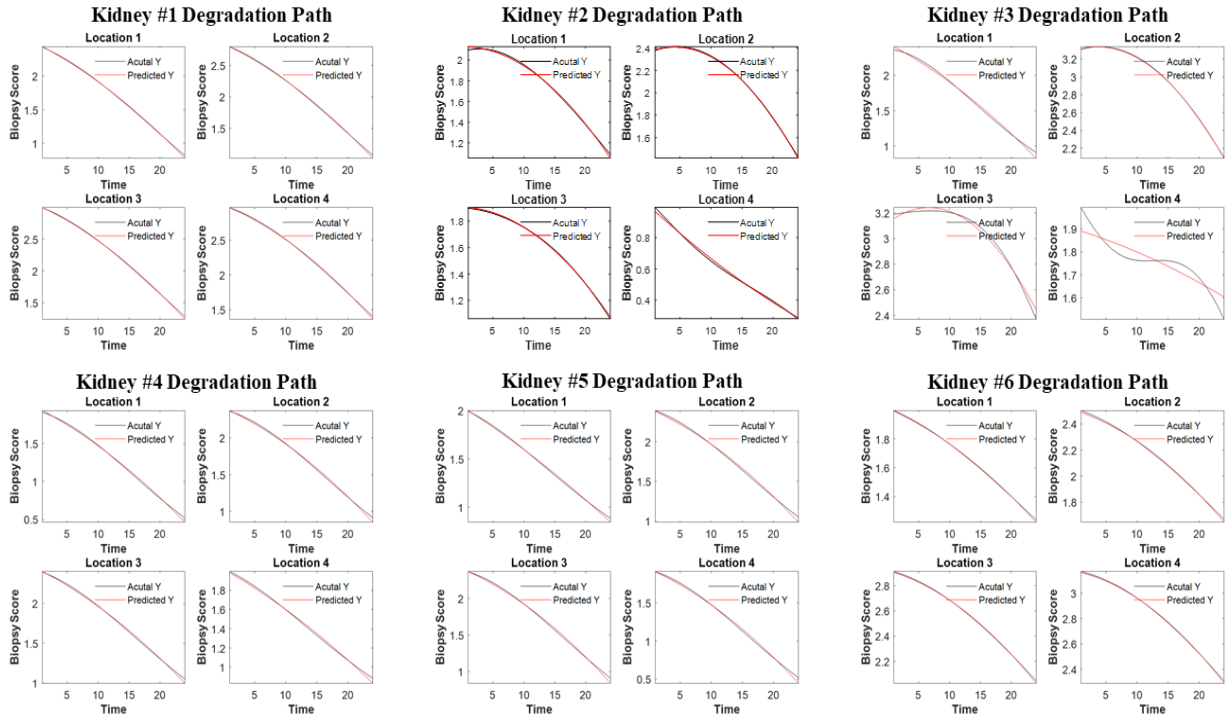


Figure A.1: Predicted biopsy score and actual biopsy score comparison at the last timestamp

Figure A.1 shows the comparisons of predicted biopsy scores and the actual biopsy scores at the last timestamp. We conclude that the modeling accuracy becomes very accurate in forecasting as we learn more information by adding more biopsy scores of the testing kidney to the training data set. As more information is known about the testing kidney, the model coefficients are updated to get more accurate forecasting results.

B.1 Model Diagnostic and Discussions

Figure B.1, each 6×6 matrix shows pairwise dissimilarities among model coefficients estimated from six kidneys, where each row/each column represents model coefficients estimated from one of the six kidneys. The off-diagonal entry at the i-th row and the j-th column represents the model coefficient similarity between i-th and the j-th kidneys, $i = 1, \ldots, 6, j = 1, \ldots, 6$. In Figure B.1, darker blue representing greater dissimilarity between kidneys. In contrast, lighter blue means more similarity in model coefficients. By visualizing this connectivity matrix, the dissimilarities among different kidneys can be determined based on color intensity. Therefore, it can easily conclude that the third kidney is significantly different from the rest of the kidneys.

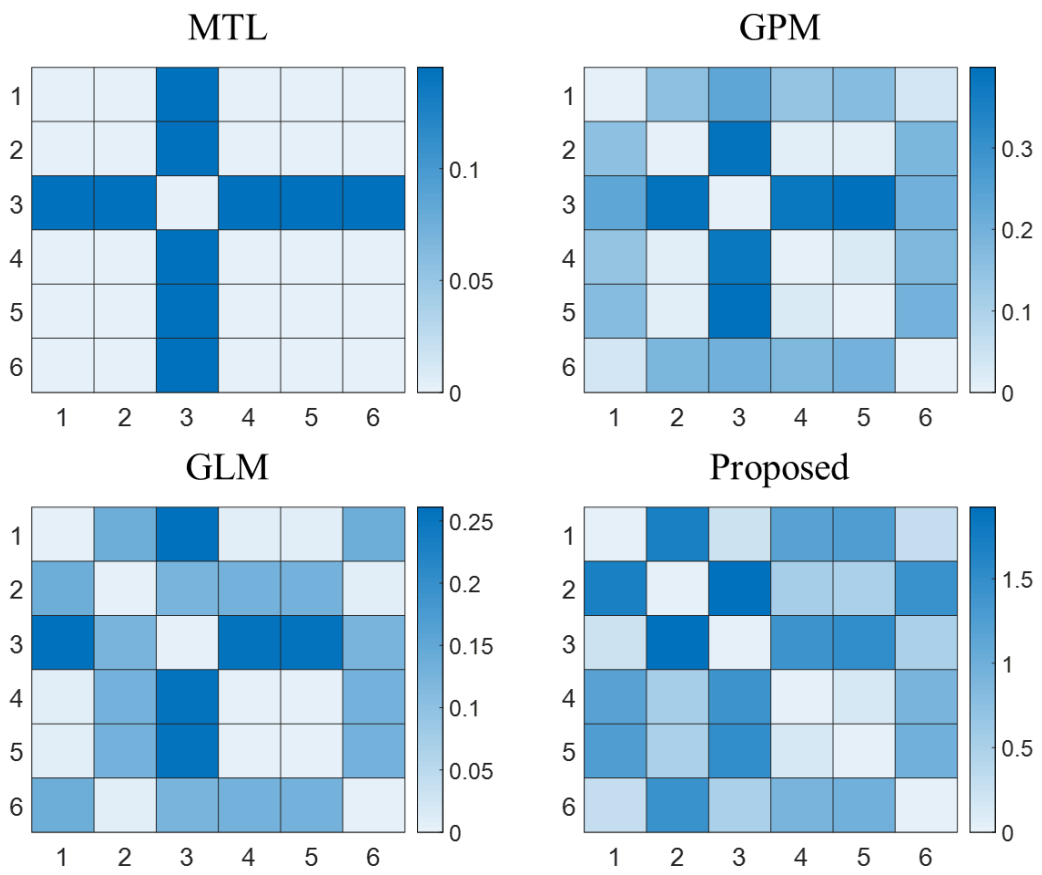


Figure B.1: Pairwise distance between model coefficients in Case Study 1

In each leave-one-kidney-out scenario, each 4×4 matrix in Figure B.2 to Figure B.7 shows pairwise dissimilarities among model coefficients estimated from four tasks. The heterogeneity of each location can be indicated by the model coefficients from the heat map.

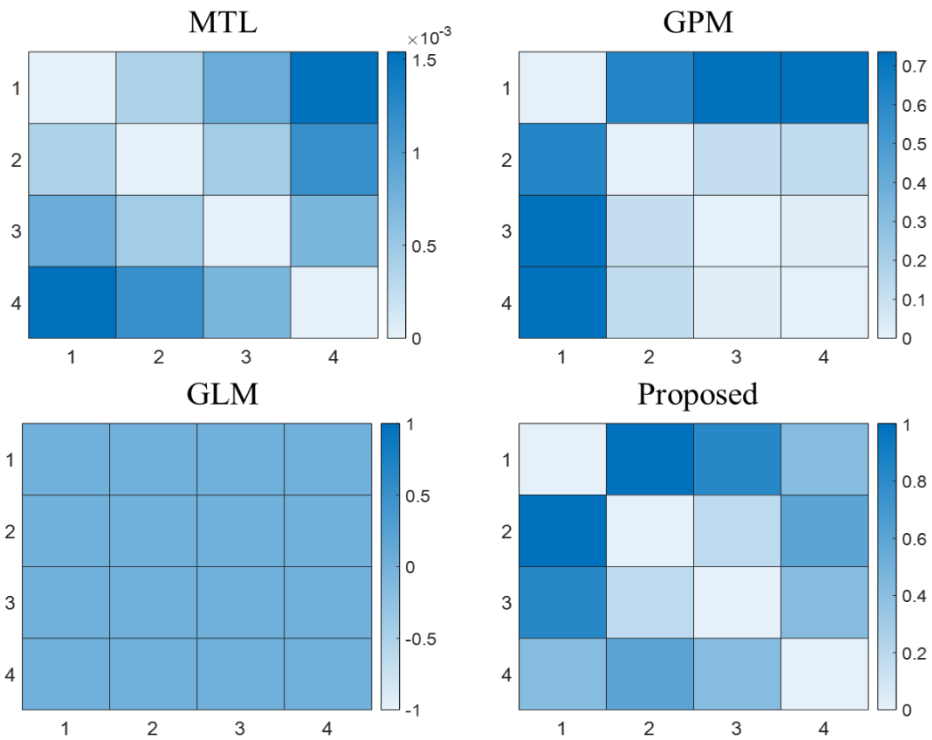


Figure B.2: L1O pairwise distance between model coefficients in Case Study 1

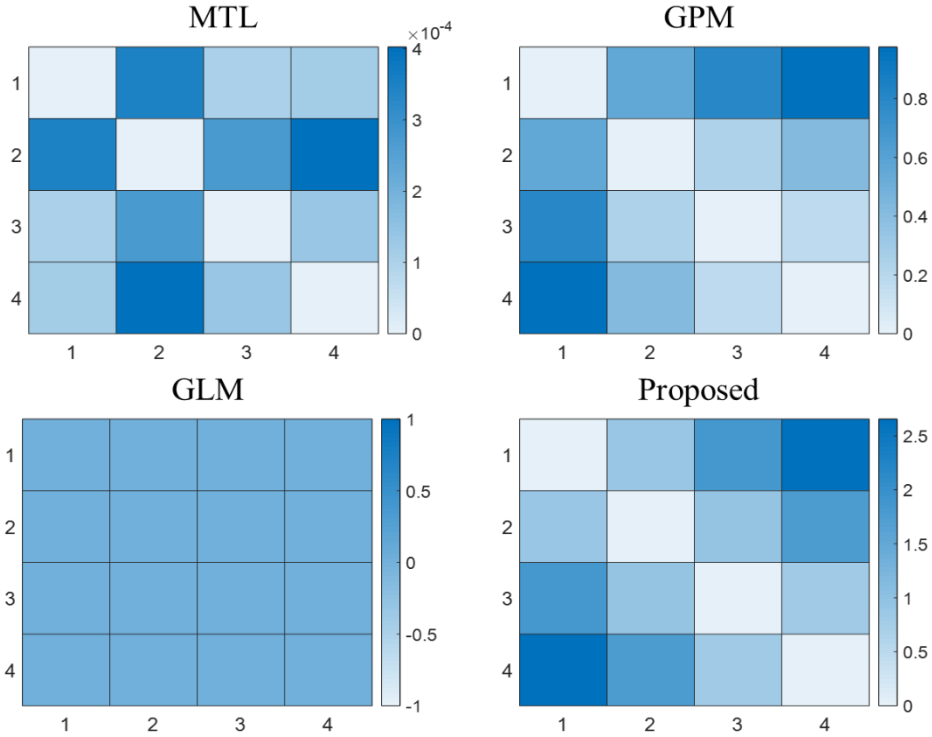


Figure B.3: L2O pairwise distance between model coefficients in Case Study 1

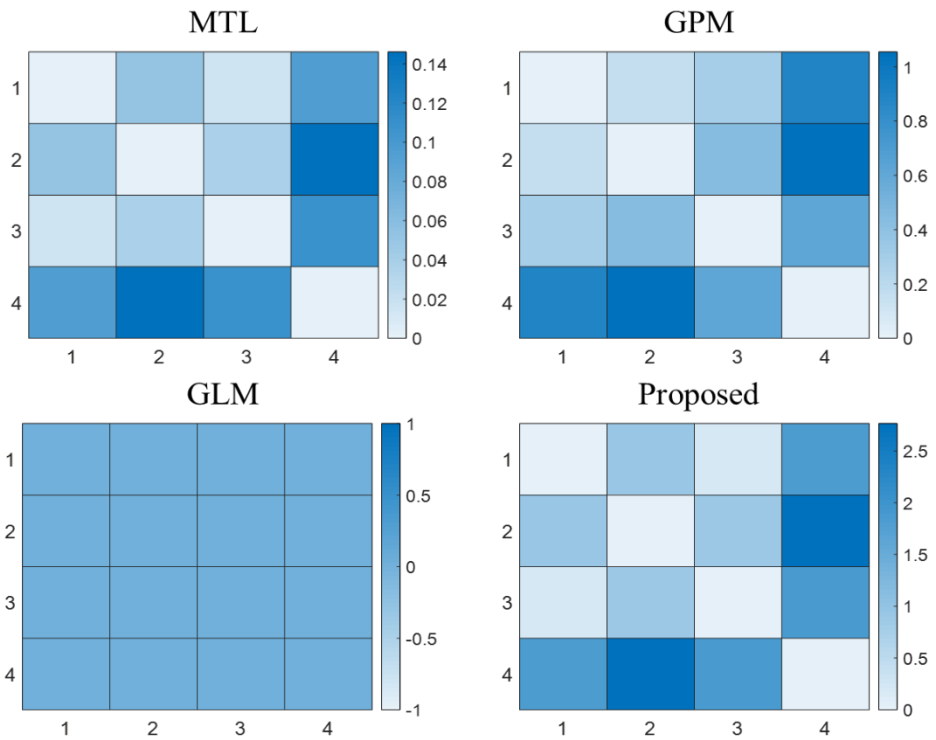


Figure B.4: L3O pairwise distance between model coefficients in Case Study 1

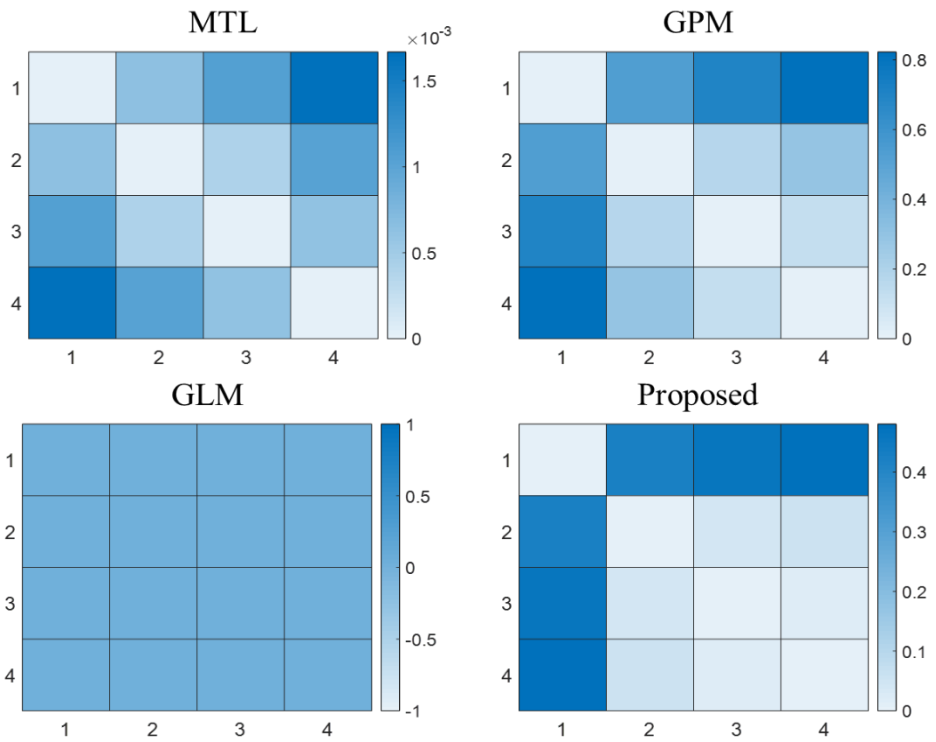


Figure B.5: L4O pairwise distance between model coefficients in Case Study 1

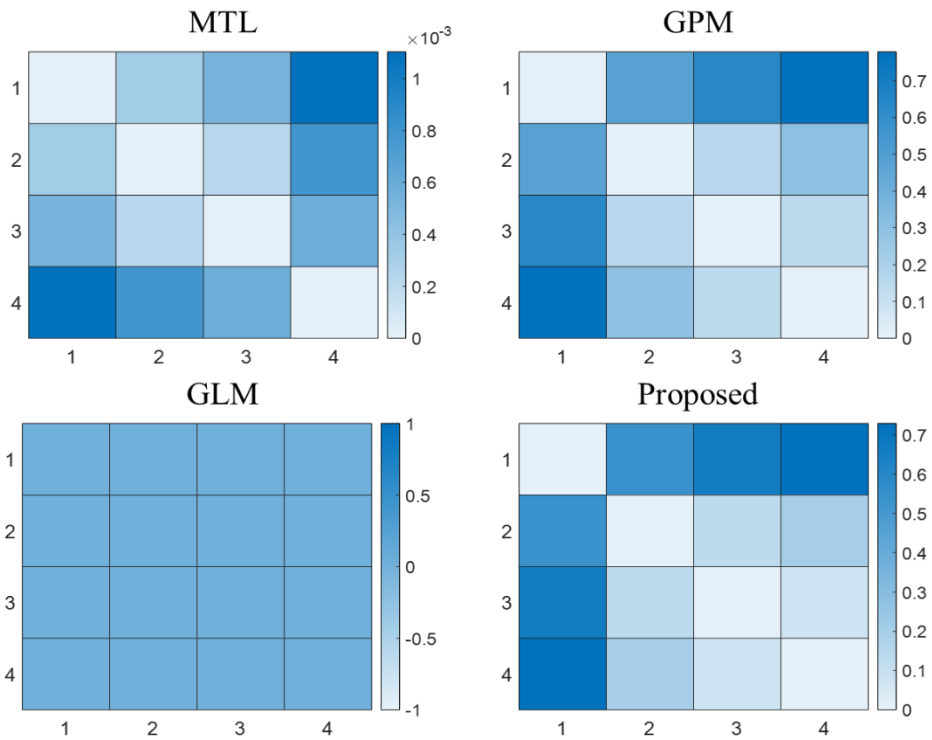


Figure B.6: L5O pairwise distance between model coefficients in Case Study 1

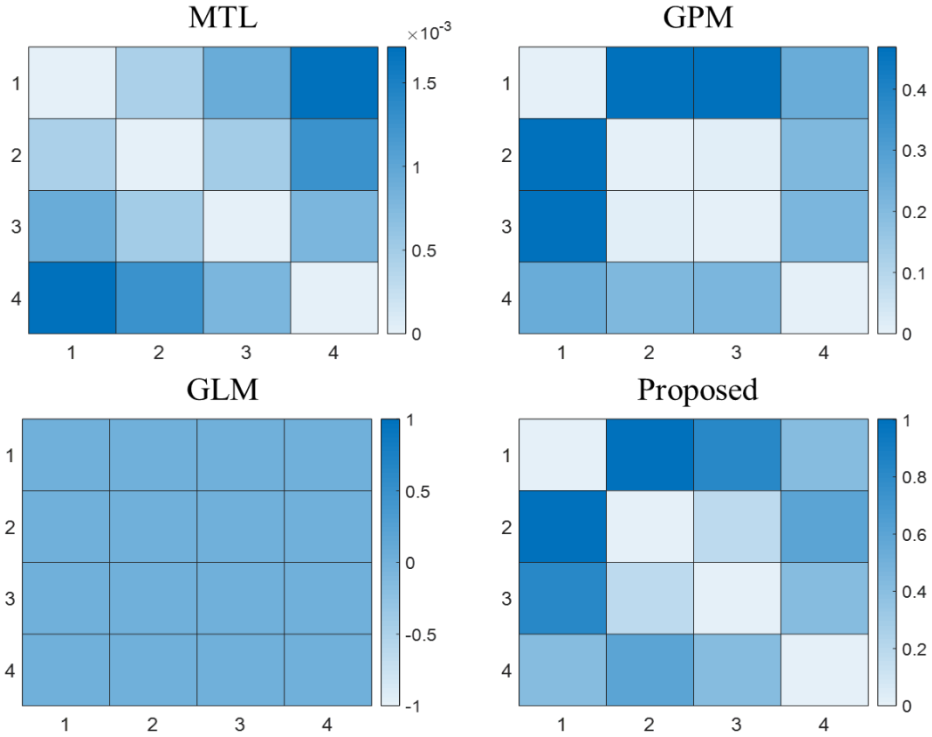


Figure B.7: L6O pairwise distance between model coefficients in Case Study 1

Similar to Case Study 1, each 8×8 matrix in Figure B.8 shows pairwise dissimilarity among model coefficients estimated from eight markers in Case Study 2. It can conclude that the model of the second marker shares the least commonality in model coefficients with the rest of markers, which is in coincidence with L2O model performance comparison in Figure 5. As discussed in previous section, the trajectory of the second marker deformation path is significantly different from the other markers.

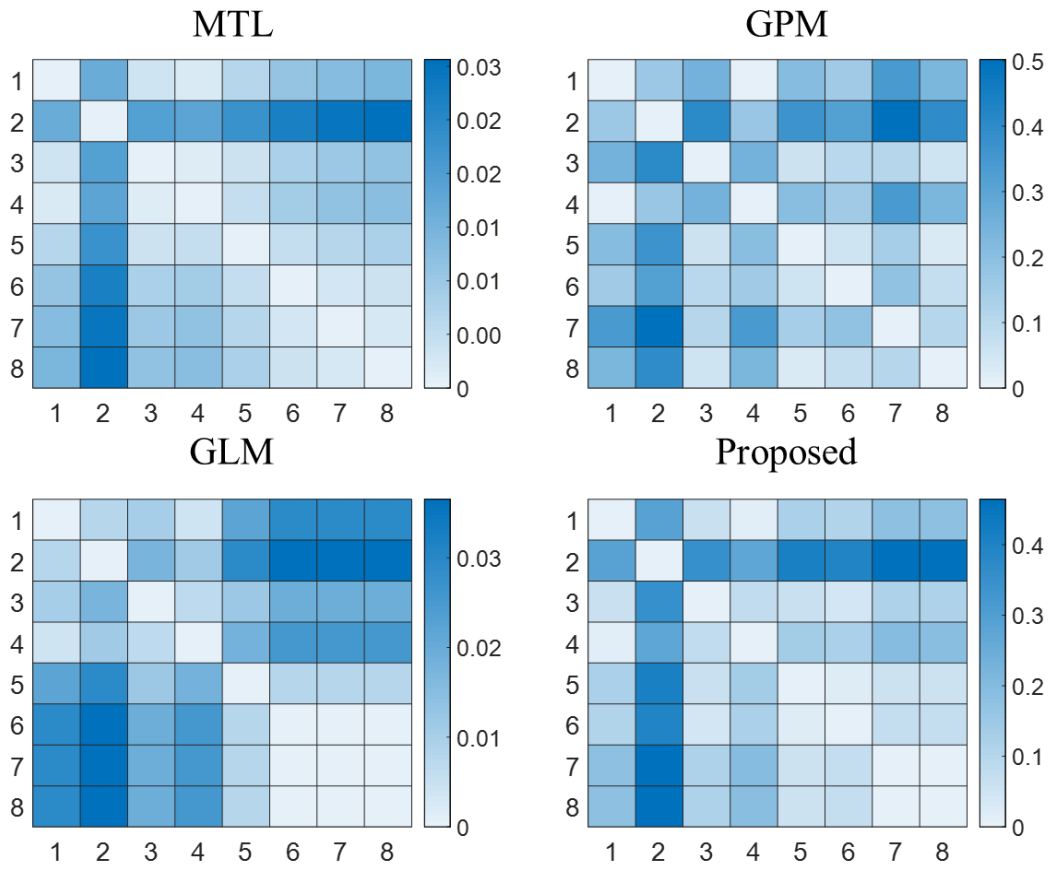


Figure B.8: Pairwise distance between model coefficients in Case Study 2

As discussed in the main body of this paper, only mark 1 to 8 will be used for analytics since the magnitudes of deformation of mark 9 to 16 are too small. In addition, they cannot be represented by quadratic model. Figure B.9 demonstrates the difference of two groups by their model coefficients.

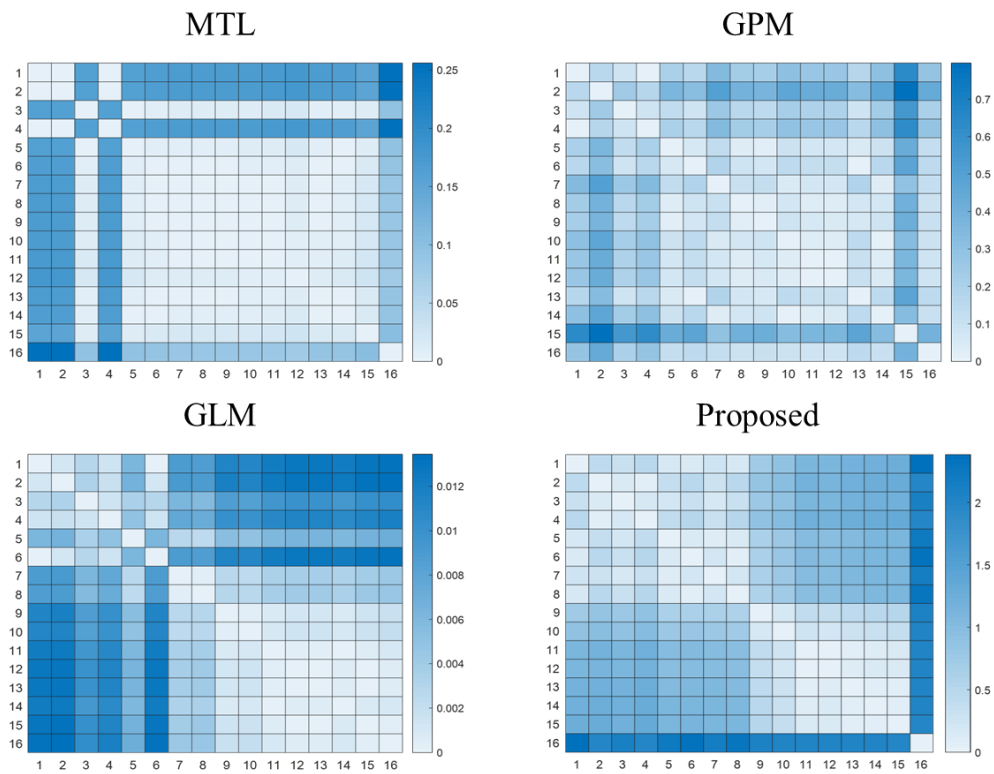


Figure B.9: Pairwise distance between model coefficients in Case Study 2 (16 markers)