Computationally Efficient Optimization Method to Quantify the Required Surgical Accuracy for a Ligament Balanced TKA

Laura Bartsoen [ID], Matthias G.R. Faes [ID], Mariska Wesseling, Roel Wirix-Speetjens [ID], David Moens [ID], Ilse Jonkers [ID], and Jos Vander Sloten [ID]

***Abstract**—Objective:* **This study proposes a computationally efficient method to quantify the effect of surgical inaccuracies on ligament strain in total knee arthroplasty (TKA). More specifically, this study describes a framework to determine the implant position and required surgical accuracy that results in a ligament balanced post-operative outcome with a probability of 90%.** *Methods:* **The response surface method is used to translate uncertainty in the implant position parameters to uncertainty in the ligament strain. The designed uncertainty quantification technique allows for an optimization with feasible computational cost towards the planned implant position and the tolerated surgical error for each of the twelve degrees of freedom of the implant position.** *Results:* **It is shown that the error does not allow for a ligament balanced TKA with a probability of 90% using preoperative planning. Six critical implant position parameters can be identified, namely AP translation, PD translation, VV rotation, IE rotation for the femoral component and PD translation, VV rotation for the tibial component.** *Conclusion:* **We introduced an optimization process that allows for the computation of the required surgical accuracy for a ligament balanced postoperative outcome using preoperative planning with feasible computational cost.** *Significance:* **Towards the research society, the proposed method allows for a computationally efficient uncertainty quantification on a complex model. Towards surgical technique developers, six critical implant position parameters were identified, which should be the focus when refining surgical accuracy of TKA, leveraging better patient satisfaction.**

***Index Terms**—Musculoskeletal model, surgical accuracy, total knee arthroplasty, uncertainty quantification.*

I. INTRODUCTION

AFTER a total knee arthroplasty (TKA), 20–30% of patients suffer from persisting pain, joint stiffness and/or inability to perform activities of daily living [1]. 52% of the patients with a TKA even indicate that they experience some degree of impairment in functional activities [2]. An aging population and increasing prevalence of TKA in young active patients results in an increase of both primary and revision TKA. The failures that result in revision are in 47.4% of the cases due to joint stiffness, joint instability or implant loosening [3]. These failure modes are often related to a sub-optimal patient-specific implant position. Preoperative planning can support the surgeon in determining the optimal patient-specific implant position. Currently, most preoperative planning processes solely account for bone geometry when determining an implant position that is consistent with a mechanically aligned TKA. A mechanically aligned knee has a knee center in line with the ankle and hip centers. Mechanical alignment however does not account for strain in the ligaments, whereas several studies [4]–[6] describe that not accounting for balanced ligament strain when determining the ideal implant position is the cause of different failure types and the high patient dissatisfaction. A TKA is considered ligament balanced when the ligaments are appropriately tensioned to provide passive stability without inducing stiffness, limited motion or pain. This study considers a TKA as balanced when the ligaments generate a force both medial and lateral throughout the squat motion and ligaments are not damaged. The studies of Provenzano *et al.* [7] and Guo *et al.* [8] show that ligament damage starts occurring from 6% strain. This zone will be further referred to as the "safe zone." As a strain larger than 6% will result in permanent deformation of the ligament without rupture, an "extended safe zone" with a maximal strain of 10% is also evaluated. This bound is still below the ultimate tensile strain of a human knee ligament [9]. However, such a ligament balance evaluation in a preoperative planning step requires complex modeling in order to predict the post-operative strain in the ligaments during movement. Furthermore, it is currently not known what surgical accuracy is needed to achieve optimal ligament balancing based on preoperative planning, nor which of the implant position parameters are most critical and would need to be achieved with most accuracy. To determine the required surgical accuracy and hence account for the associated uncertainty level, the effect

of surgical inaccuracies on the postoperative implant position and the associated risk of ligament imbalance and therefore post-operative outcome has to be quantified. Hereto, a musculoskeletal model (MSM) in combination with a computationally efficient probabilistic model (PM) is required to perform an optimization towards the maximally tolerated surgical error. The PM can then be used to quantify the uncertainty in the ligament strains caused by the error introduced during surgery. There exist three PM techniques and their variants, namely Monte Carlo simulation (MCS), response surface method (RSM) and fast probability integration (FPI) [10]. MCS involves random sampling of the complex model, in this study the MSM, where the input parameters are sampled according to their joint statistical distribution to estimate for example the cumulative distribution function (CDF) of the model output parameters. MCS is the most time consuming technique as it is simply sampling of the generally time consuming MSM with different sets of input parameters according to their joint statistical distribution. Nevertheless, it remains the golden standard as convergence to the correct uncertainty distribution is guaranteed if an infinite amount of samples are drawn from the input distribution. The FPI method computes a direct approximation of the CDF of the uncertainty in each of the outputs. With RSM, a surrogate model is trained that approximates the actual model. As the surrogate model requires a low evaluation time, the MCS on this surrogate model will require a low computation time compared to a direct MCS on the MSM itself. RSM and FPI are more efficient but result in an approximate solution. Especially FPI requires an extensive knowledge of the system in order to be able to evaluate applicability, as the method requires a monotonic system. If different combinations of parameters lead to the same output, FPI has difficulty converging.

Recently, the integration of PM in MSM workflow was studied [11]–[17]. The studies of Strickland *et al.* [14] (wear simulator), Navacchia *et al.* [16] (gait) and Smith *et al.* [15] (gait) quantify uncertainty through MCS on the MSM directly. This PM method is infeasible to be used in an optimization process due to its computational inefficiency. The studies of Strickland *et al.* [13] (wear simulator) and Arsene *et al.* [17] (stair ascent) apply the RSM method using a 2nd-order polynomial as a surrogate model. A 2nd-order polynomial model can at most model 2nd-order behavior, whereas a complex, non-linear MSM as used in this study, includes higher order behavior. The studies show that a 2nd-order polynomial is sufficient for a rough estimation of uncertainty, but if used in an optimization process it will result in inaccurate results. All mentioned studies solely evaluate the effect of specific model input parameters (i.e. implant position, ligament material properties) on kinematics and contact pressure. However none of the studies evaluated the effect on ligament strain, whereas ligament balancing has been shown to be important to TKA patient satisfaction. Furthermore none of the studies used uncertainty quantification for estimation of the implant position.

The objective of this study is to use PM in combination with MSM to determine the required surgical accuracy for TKA in order to achieve ligament balancing based on a preoperative surgical plan, with a feasible computational cost. The aim therefore is to achieve a ligament balanced post-operative outcome using preoperative planning with a probability of 90%. The as such optimized implant position maximizes the tolerated surgical error. To this end, we aim to achieve ligament balancing with 90% probability comparable to the current patient satisfaction for total hip arthroplasty which is based on literature and estimated to be about 90% [18]–[20]. Hence, uncertainty quantification in ligament strain needs to be introduced in our musculoskeletal modeling workflow, to account for uncertainties in the implant position of the femoral and tibial components due to the surgical process. Therefore, an optimization process is defined where each iteration applies a PM on a complex, non-linear MSM that simulates a squatting motion. As a PM, the RSM method in combination with MCS is used to quantify uncertainty with a feasible computational cost.

II. MATERIALS AND METHODS

Fig. 1 gives an overview of the applied workflow. The implant position parameters are 3 translational and 3 rotational components for the femoral implant and 3 translational and 3 rotational components for the tibial implant. Implant position samples are drawn according to their statistical distribution. The sampled sets of implant position parameters are evaluated by the surrogate model of the PM (Section II-B) and converted in the uncertainty of the ligament strains. The ligaments that are included are the deep medial collateral ligament (deepMCL), superficial medial collateral ligament (supMCL), lateral collateral ligament (LCL), anterolateral ligament (ALL) and popliteofibular ligament (PFL). In order to have the 90% confidence interval of the maximal ligament strains within the safe zone, the mean and standard deviation of the uncertainty distribution of the implant position parameters are adapted. This iteration continues for different means and standard deviations of the implant position parameters until the largest set of tolerated surgical errors is found. This optimization process (Section II-C) requires a computationally efficient PM to allow an optimization with feasible computational cost, since a multitude of biomechanical model evaluations is required. This is achieved by using a surrogate model to perform the MCS. The surrogate model is trained based on multiple MSM (Section II-A) evaluations. There are thirteen inputs, namely the flexion angle and 12 degrees of freedom (DOFs) of the implant position, and 5 outputs, namely the maximal strain in the deep MCL, superficial MCL, LCL, PFL and ALL. Different surrogate modeling techniques are evaluated, namely 2nd order polynomial (Poly), support vector regression (SVR), Guassian process regression (GPR) and artificial neural networks (ANN). The results of each of the surrogate modeling techniques is validated using a MCS in combination with a full non-linear model of the squatting motion.

A. Musculoskeletal Model

The MSM is based on the model described in the study of Vanheule *et al.* [21]. This study validates the MSM accounting for varus/valgus changes of the tibial implant. The MSM is implemented into the AnyBody modeling system 7.3.0 (AnyBody Technology A/S, Aalborg, Denmark). The implant system is a
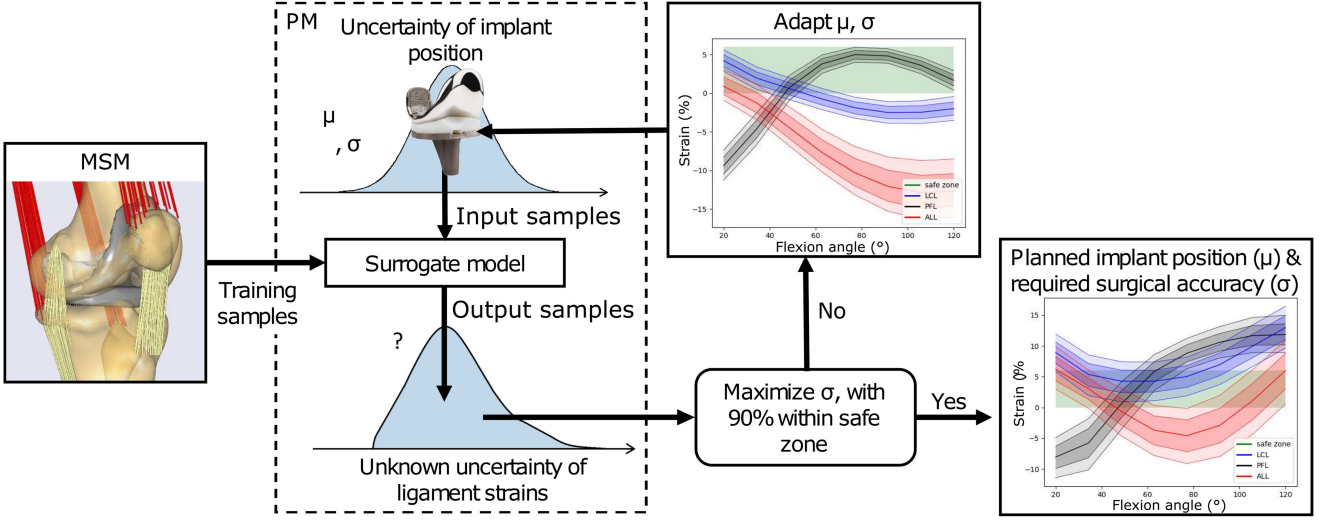
Fig. 1. Overview of the designed method for optimization towards the required surgical accuracy for a ligament balanced TKA.

posterior-stabilized system (Performance, Biomet Inc., Warsaw, IN, USA). The model simulates a squat movement from 20° to 120° of flexion. The model is made as subject specific as would currently be possible in a clinical setting for three subjects. More specifically, by segmenting the bone and cartilage geometry from MR images using Mimics 17.0 (Materialise N.V., Leuven, Belgium) as well as the ligament attachment areas. In contrast to the original model of Vanheule *et al.*, the ligaments are modeled with multiple strands. The deepMCL, supMCL, LCL, ALL, PFL, posterior capsule (PC) and patella tendon (PT) are modeled with respectively 20, 20, 10, 10, 10, 6 and 30 strands. The force/strain behavior in the ligaments is modeled using a linear elastic relationship with a quadratic slack region [22]. The force/strain behavior is given in (1), where F is the force, ϵ is the strain, k is the ligament stiffness and ϵ_l is an experimentally fitted parameter equal to 0.03. The strain ϵ can be computed through (2), where L_r and ϵ_r are respectively the strand length and strain at reference (extension) position. Further details on model implementation can be found in the study of Vanheule *et al.* [21]. The collection of the data for the study of Vanheule *et al.*, which is also used in this study, was approved by "Medical ethics committee - Faculty of Medicine - KU Leuven" with number "NH019 2015-05-01".

$$F = \begin{cases} 0 & \text{if } \epsilon < 0 \\ \frac{1}{4}\frac{k\epsilon^2}{\epsilon_l} & \text{if } 0 \leq \epsilon \leq 2\epsilon_l \\ k(\epsilon - \epsilon_l) & \text{if } \epsilon > 2\epsilon_l \end{cases} \quad (1)$$

$$\epsilon = \frac{(1+\epsilon_r)L - L_r}{L_r} \quad (2)$$

The reference strain (ϵ_r) varies between strands throughout the ligament. Four values are assigned to each ligament. The specific reference strain of a strand follows from linear interpolation between the four reference strain values dependent on the attachment position of the strand on the Femur. As the material properties (linear stiffness and reference strain) of the ligaments are generally not available in a clinical setting, these

are defined based on an optimization procedure during which the modeled tibio-femoral kinematics of the native knee are fitted to the experimentally measured ones while constraining the maximal ligament strain between 0% and 6% throughout the squat. The maximal strain in a ligament is the strain in the strand of the ligament that is the highest at the current degree of flexion. The boundaries of the constraint are motivated by the fact that literature shows that 6% is the onset of damage [7], [8] and instability in a healthy knee is improbable. The material properties are given in Table I for each of the three specimens.

B. Probabilistic Model

The PM converts the uncertainty in the implant position parameters (12 inputs) into the uncertainty in the maximal ligament strains (5 outputs). For training of the surrogate models, a total of 3000 samples are gathered using Sobol sampling with a 300 extra on the border regions of the sampling range as this is where large maximal errors were detected. The sampling bounds are taken at $+/-8$ mm or ° with respect to the implant position consistent with mechanical alignment (Kneeplanner of Materialise N.V., Leuven, Belgium). A 3-fold cross validation with 300 testing samples is performed. Model accuracy and training times are evaluated for 500, 1000, 2000 and 3000 training samples. 300 testing samples are used for each training, independent of the amount of training samples that is used. The training and testing samples are randomly selected out of all gathered samples for each cross validation. MCS is applied on this surrogate model to quantify uncertainty. Four different kinds of surrogate modeling methods are investigated, namely a 2nd order polynomial model, a Support vector regression model, a Gaussian process regression model and an Artificial neural network. The model inputs and outputs are normalized respectively to $[-4, 4]$ and $[0,1]$ in order to simplify training of the surrogate models.

1) 2nd Order Polynomial: Each of the outputs is described by a multivariate 2nd order polynomial (Poly) given in (3). In this equation, N is the amount of input parameters, x represents

TABLE I
MATERIAL PROPERTIES OF THE LIGAMENTS OF THE THREE SUBJECTS

Ligaments	Stiffness (N)			Reference strain											
				Anterosuperior			Anteroinferior			Posterosuperior			Posteroinferior		
Specimen	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
deep MCL	4600	4600	6398	0.20	0.09	0.18	0.08	0.04	0.07	-0.07	-0.06	-0.02	0.02	-0.01	0.26
superficial MCL	9400	6600	6607	0.02	0.07	-0.01	0.04	-0.05	0.12	0.05	-0.10	0.06	0.04	0.03	0.10
LCL	9400	7989	9400	0.15	0.14	0.05	-0.02	0.04	0.07	0.10	0.01	0.10	-0.01	0.16	0.08
ALL	9400	8192	8876	0.06	0.08	0.11	0.00	0.03	0.15	-0.05	-0.03	0.00	-0.27	-0.04	-0.18
PFL	9400	9400	6600	-0.01	-0.04	-0.10	-0.29	-0.33	-0.34	-0.13	-0.15	-0.13	-0.24	-0.30	-0.37
Posterior capsule	10000			0.20											
Patella tendon	80000			-0.25											

the inputs and y an output. The parameters β_0, β_i, β_{ij} and β_{ii} are found using least squares method.

$$y = \beta_0 + \sum_{i=1}^{N} \beta_i x_i + \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \beta_{ij} x_i x_j + \sum_{i=1}^{N} \beta_{ii} x_i^2 \quad (3)$$

2) Support Vector Regression: A support vector regression (SVR) model is trained by maximization of regression model flatness (first part of (4)) and prediction error (second part of (4)). By variation of a parameter C, more or less weight can be assigned to the prediction error. A threshold ϵ is used around the estimated function, so that the samples of which the absolute errors with respect to the regression function are below this threshold are not included in the prediction error. Equation (5) gives the regression function, where l is the amount of training samples, K is a kernel function, b is a constant, x are the inputs and y is an output. Equation (6) gives a radial basis function (RBF) that is used as the kernel function K. A grid search is used to select the model parameters C, ϵ and γ. The parameters α_i and α_i^* are optimized [23]. The SVR method is applied using scikit-learn v0.21.0 [24], as described by Chang and Lin [25].

$$\min_{w} \quad \frac{1}{2}||w||^2 + C\sum_{i=1}^{l} \xi_i + \xi_i^* \quad (4)$$

$$y = \sum_{i=1}^{l} (-\alpha_i + \alpha_i^*) K(x_i, x) + b \quad (5)$$

$$K(x_i, x) = exp(-\gamma||x_i - x||^2) \quad (6)$$

3) Gaussian Process Regression: A Gaussian process (GP) is a stochastic process that generates random variables in time or space, such that a finite subset of those variables follows a multivariate Gaussian distribution. In Gaussian process regression (GPR), the covariance of the GP is specified using a kernel function [26]. For the described application, a constant kernel is multiplied with an RBF kernel. A white noise kernel is added to explain the noise in the data originating from MSM inaccuracy. The function description of the kernel is given in (7). During fitting of the Gaussian Process regressor, the hyperparameters (c and l) of the kernel are optimized by minimizing the negative log-marginal-likelihood, which is typically used for GPR.

The GPR method is implemented using scikit-learn v0.21.0 [24].

$$K(x_i, x) = c \cdot exp\left(\frac{-1}{2}||\frac{x_i}{l} - \frac{x}{l}||^2\right) + noiselevel(x_i, x) \quad (7)$$

4) Artificial Neural Networks: The artificial neural network (ANN) is implemented using the open-source software Tensorflow 2.0.1 [27]. The network has architecture [13:1024:512:256:64:16:5] with activation function Softplus (8). The function given in (9) is used as the loss-function with d set to 0.1. This function computes the Huber loss and has a behavior that is similar to the mean squared error (MSE) for small errors and similar to the mean absolute error (MAE) for large errors. This way it is less affected by outliers. The Adam optimizer is used with a batch size of 32. The learning rate is halved when the loss doesn't decrease with more than 10^{-6} for more than 50 epochs and starts at 0.001. To prevent overtraining, l2 regularization is applied. The weighted average of the result of six networks that are separately trained on the same data, is used to reduce random errors (ensemble averaging).

$$a(x) = log(exp(x) + 1) \quad (8)$$

$$x = y_{true} - y_{predicted}$$

$$loss = \begin{cases} 0.5x^2 & \text{if } |x| \leq d \\ 0.5d^2 + d(|x| - d) & \text{if } |x| > d \end{cases} \quad (9)$$

5) Monte Carlo Simulation: A MCS on the complex, nonlinear MSM directly is used to validate the CDF prediction of the RSM method in combination with a MCS using the described surrogate modeling methods. A MCS can give an estimate on the CDF of the output parameters by randomly drawing samples from the input parameters while respecting the statistical distribution of their variation. In this case the input parameters are Normally distributed with a mean of 0.0 mm or ° and a standard deviation of 2.0 mm or ° from the planned position consistent with mechanical alignment. The inputs are considered uncorrelated. The drawn samples are evaluated by the MSM in order to convert the different sets of inputs to the corresponding sets of outputs. From the set of computed outputs, the CDF corresponding to their statistical distribution can be estimated. In order to compare the accuracy of the surrogate models, the relative difference of the mean and the RMSE between the CDFs estimated with MCS and the RSM method with different surrogate models are computed. The relative difference is quantified by dividing the absolute difference of the mean by the width of the 95 % uncertainty interval.

With a MCS it is important to check if convergence has occurred with the amount of samples that is used. In this study convergence is evaluated by evaluating the consistence in predicted width of the 60% and 90% confidence intervals for different sample sizes. Sample sizes are taken from 100 to 2900 with a difference of 100 samples between every evaluation. The width of the 95% confidence interval is computed on the prediction of the width of the 60% and 90% confidence intervals by drawing 500 times randomly the considered sample size from a pool of 3000 samples. Once the width of the 95% uncertainty interval on the width of the 90% and 60% intervals is smaller than 5% of the actual width of the 90% and 60% intervals, the MCS is considered converged. The actual width of the 90% and 60% intervals is estimated from the 3000 available samples. This evaluation is done at 20°, 60°, 90° and 120° of flexion. For each of the outputs and flexion angles, the amount of samples for convergence of the MCS is determined.

C. Surgical Accuracy

The tolerated surgical error to achieve ligament balancing based on preoperative planning, is identified through an optimization. To this end, the standard deviations of the implant position parameters are maximized (objective) by adapting the mean and standard deviations of the implant position parameters provided that a ligament balanced post-op outcome is realized with a probability of 90% (constraint). The optimized means relate to the implant position that should be planned to achieve a post-op ligament balanced result whereas the optimized standard deviations relate to the maximally allowed surgical error to reach a probability of 90%. The objective of the optimization problem is given in the first line of (10), where σ_i is the standard deviation of implant position DOF i, N is the amount of DOFs of the implant position (=12). The natural logarithm of the standard deviations is taken to add weight to an increase in small standard deviations compared to the same increase in large standard deviations. This to avoid that the standard deviations of the critical parameters would converge to zero. A small value (0.001) is added to the standard deviation to avoid the occurrence of infinity when σ goes to zero. The constraint of the optimization problem is given by the second line of (10), where $S_{MCS}^{90\%}$ is 90% of the samples from the MCS, θ_{FE} is the knee flexion angle ($20° - 120°$), ϵ is the maximal strain in a ligament, L is the set of all ligaments (deepMCL, supMCL, LCL, ALL and PFL), L_{lat} is the set of all lateral ligaments (LCL, ALL and PFL) and L_{med} is the set of all medial ligaments (deepMCL and supMCL). $\epsilon_{\max}$ is the upper bound on the maximal strain in the ligaments, which is 6% for the safe zone and 10% for the extended safe zone.

$$\min_{\mu, \sigma} \quad -\sum_{i=1}^{N} \ln\left(0.001 + 2\sigma_i\right)$$

$$\text{s.t.} \quad S_{MCS}^{90\%} : \forall \theta_{FE} \begin{cases} \epsilon \leq \epsilon_{\max} & \forall L \\ \epsilon \geq 0 & \exists(L_{lat} \wedge L_{med}) \end{cases} \quad (10)$$

Verification of the constraint requires an estimation of the variation of the ligament strains due to variation in the implant position parameters, which requires a MCS in each iteration of the optimization. Each MCS consists out of 5000 samples for 7 flexion angles equally divided between 20° and 120°. To avoid differences between iterations due to statistical noise on the output of the MCS, identical samples are used which are rescaled based on the current statistical distributions of the implant position parameters. This MCS is performed using the surrogate modeling method ANN, as the results (Section III-A2) show that it presents the best approximation of the MSM compared to the other evaluated surrogate modeling methods.

We opted to account in the optimization for the per-operative decision to alter the thickness of the insert. This way the optimized outcome is robust for this decision, allowing the preoperative plan still feasible without the need of additional bone cuts. To this end, each sample of the MCS that is outside of the safe zone is reevaluated with a different insert thickness by the surrogate model. Different inserts differ with 2 mm. If the sample satisfies the safe zone after change of the insert it is accepted.

The package PyMoo [28] is used to perform the constrained optimization. The optimization algorithm NSGA-II [29] is applied as this is an evolutionary algorithm. Evolutionary algorithms have the large advantage that it is unlikely to converge to a local minimum of the objective function.

The optimized standard deviations allow the classification of the implant position parameters into critical and non-critical parameters. The accuracy reachable with robot assisted surgery is used as a reference to what can be currently achieved during surgery [30]. To evaluate the optimized standard deviations with respect to robot surgery, they are divided into critical and non-critical parameters, which have a standard deviation of respectively $0-0.36$ and $0.36-3.0$ ° for rotational parameters and respectively $0-0.49$ and $0.49-3.0$ mm for translation parameters. This means that the error of the parameters that prove to be critical cannot be realized using robotic surgery or any other surgical technique with similar accuracy.

III. RESULTS

The results of the different uncertainty quantification methods and optimization are shown respectively in the Section III-A and III-B.

A. Probabilistic Model

The RSM method with four different surrogate modeling techniques is applied. The different surrogate models are evaluated using MCS.

1) Surrogate Model Training: Fig. 2(a) gives the maximum Absolute Error (AE) of 95% of the validation samples with the smallest AE for the Poly, SVR, GPR and ANN model. The maximum over the outputs is given. Training is performed with 500, 1000, 2000 and 3000 training samples on three specimens with a 3-fold cross-validation each, where an error bar indicates the variation. The figure shows that with increasing training sample size the error decreases less when adding a similar amount of samples. The decrease in error between a training set of 1000 samples to 2000 samples is considerably larger
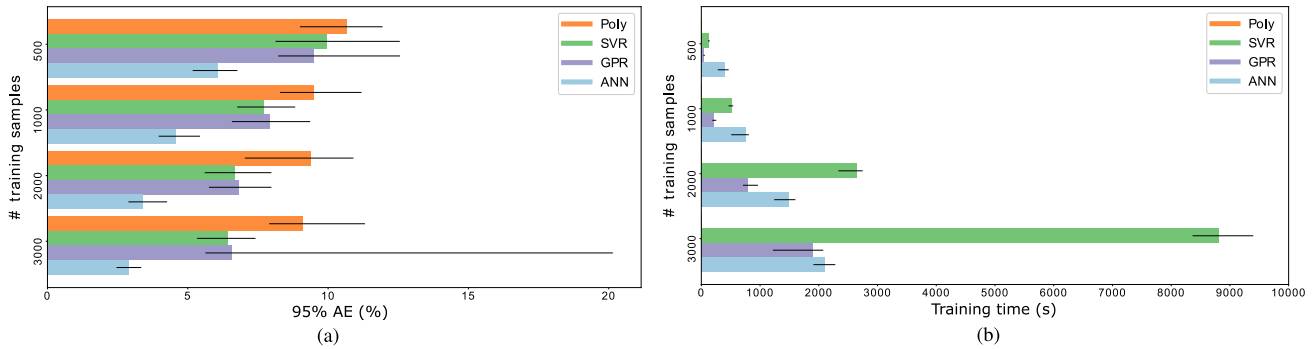
95% AE and training time for the different RSM methods for 500, 1000, 2000 and 3000 training samples for each of the three specimens. Each surrogate model is indicated with a different color given in the legend. The figure gives median and variation over the three specimens and 3-fold cross-validation. The training time of the Poly model is invisible as it is too small compared to the training time of the other surrogate models.

Fig. 2. 95% AE and training time for the RSM methods. (a) 95% AE and (b) Training time.

TABLE II
ACCURACY OF POLY, SVR, GPR AND ANN

Method	Relative Mean (%)	Relative RMSE ($\cdot 10^{-2}$)
Poly	2.19 (0.06 - 10.48)	2.54 (1.00 - 8.10)
SVR	1.77 (0.40 - 6.49)	1.97 (0.61 - 6.91)
GPR	2.70 (0.04 - 8.40)	2.68 (0.49 - 8.30)
ANN	0.50 (0.00 - 1.36)	0.70 (0.32 - 1.83)

The average difference and bounds of the relative mean, and the RMSE of the CDF generated by Poly, SVR, GPR and ANN at $20°$, $60°$, $90°$ and $120°$ of flexion compared to the MCS analysis over the ligament strains.

compared to a training set of 2000 to 3000 samples, which indicates convergence in the training with respect to the sample set size. For the Poly model, the decrease in error with increasing amount of samples is very small, as a Poly model is not capable of explaining higher order elements (>2) in the MSM. For both SVR and GPR, the errors keep decreasing but at 3000 samples they are still considerably larger compared to ANN at 1000 samples. To reach the same accuracy, more training samples would have to be used, which increases sampling and training times considerably. The training time of the Poly model (Fig. 2(b)) is negligible compared to the training time required by the other surrogate models. With increasing amount of training samples, the ANN training time increases approximately linear, whereas training times for SVR and GPR increase approximately quadratic. This would result in large training times for SVR and GPR if the amount of training samples would be further increased in order to reach similar accuracy to the ANN model. Further results are based on a training sample size of 3000.

2) Validation Through MCS: The MCS is used to validate the uncertainty quantification using the RSM technique with different surrogate modeling techniques. The MCS always converges to the correct statistical distribution if enough samples are used. Over the outputs the required amount of samples for convergence varies from 1600 to 2400 between different outputs. The 60% confidence interval for the LCL strain at $60°$ of flexion requires 1600 samples, whereas the 90% confidence intervals of the supMCL at $90°$ and $120°$ of flexion and the ALL at $120°$ of flexion require 2400 samples. Table II shows the relative difference of the mean, and the RMSE between the
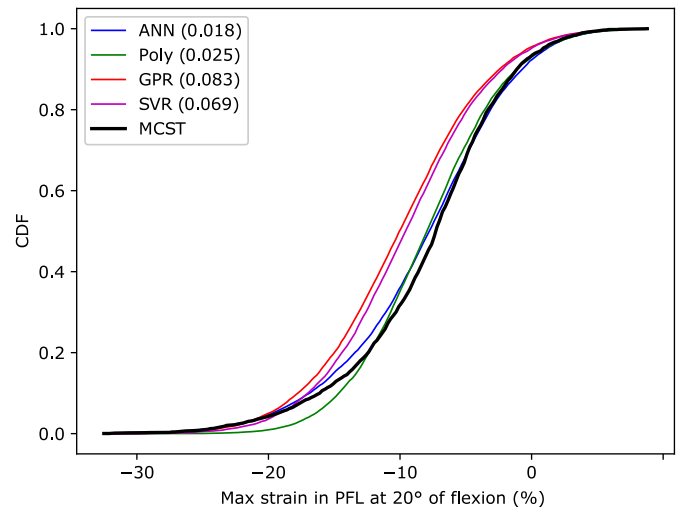

Fig. 3. CDF predicted by the surrogate models and estimated with MCS for the maximal strain in PFL at $20°$ of flexion.

CDFs estimated with MCS and the RSM method with different surrogate models. The table gives the average, minimal and maximal values of the uncertainty estimates at $20°$, $60°$, $90°$ and $120°$ of flexion. It is clear that the error of the ANN is considerably smaller compared to the other surrogate models. The Poly and GPR models perform equally well, whereas the SVR model slightly outperforms them. Fig. 3 gives the CDF estimated through a MCS on the MSM directly as well as the CDFs estimated through a MCS on the surrogate models for the strain in the PFL at $20°$ of flexion. The GPR, SVR and ANN models present the prediction with the largest RMSE on the PFL at $20°$ of flexion in comparison to other outputs at other evaluated flexion angles. This figure also illustrates a better approach of the CDF with the ANN model.

3) Computational Efficiency: MCS directly on the MSM, requires more than five-fold the amount of MSM evaluations for one set of implant position parameters in comparison with the studied techniques. For the RSM methods, 3000 MSM evaluations are required for training with another 300 for validation. The same amount of samples are needed as with a direct MCS

		Safe zone			Extended safe zone		
Specimen		1	2	3	1	2	3
Femur	Medial/lateral (mm)	1.76	1.19	0.59	3.0	3.0	2.39
	Anterior/posterior (mm)	0.23	0.17	0.28	1.3	0.45	1.04
	Proximal/distal (mm)	0.16	0.24	0.28	0.68	0.4	0.81
	Flexion/extension (°)	1.04	0.97	1.56	2.98	2.35	2.98
	Varus/valgus (°)	0.25	0.37	0.38	0.73	0.64	0.5
	Internal/external (°)	0.31	0.19	0.35	0.57	0.48	0.47
Tibia	Medial/lateral (mm)	0.65	1.36	0.62	3.0	2.57	2.83
	Anterior/posterior (mm)	0.44	0.56	0.93	1.71	1.34	1.71
	Proximal/distal (mm)	0.16	0.17	0.21	1.24	0.35	1.05
	Slope (°)	0.56	1.16	0.84	1.8	1.89	2.55
	Varus/valgus (°)	0.27	0.19	0.24	0.48	0.37	0.35
	Internal/external (°)	0.79	2.65	2.16	3.0	2.99	2.89

The ranges for the critical (red) and non-critical (green) parameters have a standard deviation that is respectively $0 - 0.36$ and $0.36 - 3.0$ ° for rotational parameters and respectively $0 - 0.49$ and $0.49 - 3.0$ mm for translation parameters. The critical ranges are not reachable with robot assisted surgery.

on the MSM. At least a 2400 samples are needed for convergence of the MCS for each of the studied outputs per flexion angle. This means if 7 flexion angles throughout the squat are evaluated, 16 800 MSM evaluations are required to perform a direct uncertainty quantification with MCS. This is considerably more compared to the 3300 MSM evaluations that are required with RSM. This proves a large increase in efficiency compared to MCS directly on the MSM.

B. Surgical Accuracy

Table III shows the optimized standard deviations of the implant position parameters for the safe zone and the extended safe zone. For the safe zone (0–6%), half of the DOFs of the implant position are critical parameters. More specifically they are Anterior/Posterior (AP) and Proximal/distal (PD) translation and Varus/Valgus (VV) and Internal/External (IE) rotation for the femoral component. For the tibial component the critical parameters are PD translation and VV rotation. In contrast, with the extended safe zone (0–10%) it can be seen that only three of the DOFs for specimen 2 and one for specimen 3 are just within the critical region.

IV. DISCUSSION

This study successfully developed a methodological framework that allows the use of uncertainty quantification in an optimization process with feasible computational cost. This study applies the technique to estimate the required surgical accuracy for a ligament balanced TKA using preoperative planning. The validation through MCS shows that overall a good approximation can be achieved with each of the studied surrogate modeling methods. This is in agreement with other studies. The studies of Arsene *et al.* [17] and Strickland *et al.* [14] used an RSM with as surrogate model a Poly model. They also find a good agreement between the validation using MCS and the used RSM technique. From our study it is however clear that with the same sampling size other (more complex) surrogate modeling techniques lead to superior accuracy.

The results of the developed PM need to be considered in the context of limitations. First, the fitting of a Poly model typically uses a design of experiment method to sample the data, e.g. Box-Behnken design or central composite design. This sampling design might lead to a better accuracy with a smaller amount of samples compared to the sobol sequence that is used in this study. Second, the training of the GPR model allows a variance based sampling in order to choose the samples based on the areas in the sampling space with the largest variance [31], [32]. This would lead to a smaller amount of samples that is required for the same accuracy. The sample choice would however be dependent on the output parameter. As this study aims to estimate a total of five output parameters, a separate variance based sampling would be needed for each output. This could in some cases eliminate the gained sampling time.

The ANN, GPR and SVR methods are generally applicable. In contrast, the Poly method will not be generally applicable. For the Poly model to be accurate, the biomechanical model has to behave in a quadratic manner, which will not be the case for most complex models. With the ANN, GPR and SVR models the estimate of the uncertainty will be accurate without the necessity of a validation through MCS, provided that training and testing samples are equally divided throughout the considered sampling space and training results in a validation error low enough for the considered application.

An extra advantage of the developed method is its potential to estimate the uncertainty caused by every possible uncertainty distribution of the input parameters that stays within the training bounds without the need for new MSM evaluations, as applied in this study to estimate the required surgical accuracy for a ligament balanced TKA using preoperative planning. Results show that using currently available surgical techniques it is impossible to reach a ligament balanced post-operative outcome using the strict safe zone with a probability of 90%. This is confirmed in all three specimens. As the strict safe zone is not successful with a large probability it is preferred to optimize towards the extended safe zone. This will result in slightly too stiff ligaments for the cases outside of the strict safe zone. This outcome is preferred over too slack ligaments, as literature [33] shows that postoperative stiffness can be mediated with good rehabilitation whereas instability due to too slack ligaments can only be corrected through revision surgery.

The optimization towards the maximal tolerated variation in the implant position parameters results in the identification of the critical parameters (Femur: AP translation, PD translation, VV rotation, IE rotation; Tibia: PD translation, VV rotation). The focus of surgical technique developers when increasing accuracy of TKA needs to be on these six critical parameters in order to achieve increased patient satisfaction.

The identification of the implant position variance has some limitations. First, the safe zone was only evaluated during a squat movement, as this type of movement addresses a large range of flexion angles. Full extension however was not evaluated. For future clinical application, alternative movements that include full extension will have to be evaluated. Second, the optimization towards the surgical accuracy was applied on a limited set of subjects. The ligament properties obtained through optimization using experimental measurement of the tibio-femoral

kinematics show a large variance between different specimens. The unavailability of these measurements in clinical practice introduces a significant amount of uncertainty in pre-operative planning. However, the results do show similar standard deviations (used to determine surgical accuracies) but highly different mean (reflecting differences in planned implant position) for each of the specimens studied. Consequently, differences in the ligament properties will mainly affect the optimized implant position and not the optimized standard deviations. The differences in ligament properties observed in the different specimens would thus not affect the conclusions of this study. Third, the optimization of the surgical accuracy is based on the proposed safe zone. This zone is based on the assumption that ligament unbalance [4], [5] is at the base of the large patient dissatisfaction in combination with studies on the relation between ligament strain and onset of damage [7], [8]. This is however not yet clinically proven. The study of Twiggs *et al.* [34] tried to identify a safe zone based on the correlation between simulated secondary kinematics and patient-reported outcome measures. The identified safe zone is however not sufficient as it can be seen that a lot of satisfied patients are outside of the identified zone, which indicates a patient-specific safe zone might be more appropriate. Future research should focus on defining a patient-specific safe zone, where this method can be applied using a different constraint to determine required surgical accuracy consistent with the patient-specific safe zone.

V. CONCLUSION

The designed optimization process allows for the computation of the required surgical accuracy for a ligament balanced post-operative outcome using preoperative planning with a feasible computational cost. Results show that, even with state of the art surgical techniques (i.e. robot assisted surgery), the error does not allow for a ligament balanced TKA with a probability of 90% using preoperative planning. A ligament balanced TKA can be achieved when allowing for slightly higher strain in the ligaments. Six critical implant position parameters can be identified, namely AP translation, PD translation, VV rotation, IE rotation for the femoral component and PD translation, VV rotation for the tibial component. The focus of surgical technique developers when increasing accuracy of TKA needs to be on these six critical parameters in order to achieve better patient satisfaction.

REFERENCES

[1] R. B. Bourne *et al.*, "Patient satisfaction after total knee arthroplasty: Who is satisfied and who is not?" *Clin. Orthopaedics Related Res.*, vol. 468, no. 1, pp. 57–63, 2010.

[2] P. C. Noble *et al.*, "Does total knee replacement restore normal knee function?" *Clin. Orthopaedics Related Res.*, vol. 431, pp. 157–165, 2005.

[3] P. F. Sharkey *et al.*, "Why are total knee arthroplasties failing today-has anything changed after 10 years?" *J. Arthroplasty*, vol. 29, no. 9, pp. 1774–1778, 2014.

[4] H. P. Delport, J. Vander Sloten, and J. Bellemans, "New possible pathways in improving outcome and patient satisfaction after TKA," *Acta. Orthop. Belgica*, vol. 79, no. 3, pp. 250–254, 2013.

[5] S. Matsuda, S. Lustig, and W. Van der Merwe, *Soft Tissue Balancing in Total Knee Arthroplasty*. Berlin, Germany: Springer, 2017.

[6] S. Babazadeh *et al.*, "The relevance of ligament balancing in total knee arthroplasty: how important is it? A systematic review of the literature," *Orthop. Rev.*, vol. 1, no. 2, pp. 70–78, 2009.

[7] P. P. Provenzano *et al.*, "Subfailure damage in ligament: A structural and cellular evaluation," *J. Appl. Physiol.*, vol. 92, no. 1, pp. 362–371, 2002.

[8] Z. Guo *et al.*, "Quantification of strain induced damage in medial collateral ligaments," *J. Biomechanical Eng.*, vol. 137, no. 7, pp. 071011-1–071011-6, 2015.

[9] K. Smeets *et al.*, "Mechanical analysis of extra-articular knee ligaments. part one: Native knee ligaments," *Knee*, vol. 24, no. 5, pp. 949–956, 2017.

[10] P. J. Laz and M. Browne, "A review of probabilistic analysis in orthopaedic biomechanics," *Proc. Inst. Mech. Engineers, Part H: J. Eng. Med.*, vol. 224, no. 8, pp. 927–943, 2010.

[11] P. J. Laz *et al.*, "Probabilistic finite element prediction of knee wear simulator mechanics," *J. Biomech.*, vol. 39, no. 12, pp. 2303–2310, 2006.

[12] S. Pal *et al.*, "Probabilistic computational modeling of total knee replacement wear," *Wear*, vol. 264, no. 7/8, pp. 701–707, 2008.

[13] M. Strickland *et al.*, "A multi-platform comparison of efficient probabilistic methods in the prediction of total knee replacement mechanics," *Comput. Methods Biomech. Biomed. Eng.*, vol. 13, no. 6, pp. 701–709, 2010.

[14] M. A. Strickland, M. Browne, and M. Taylor, "Could passive knee laxity be related to active gait mechanics? an exploratory computational biomechanical study using probabilistic methods," *Comput. Methods Biomech. Biomed. Eng.*, vol. 12, no. 6, pp. 709–720, 2009.

[15] C. R. Smith *et al.*, "The influence of component alignment and ligament properties on tibiofemoral contact forces in total knee replacement," *J. Biomech. Eng.*, vol. 138, no. 2, pp. 021017-1–021017-10, 2016.

[16] A. Navacchia *et al.*, "Prediction of in vivo knee joint loads using a global probabilistic analysis," *J. Biomech. Eng.*, vol. 138, no. 3, pp. 031002-1–031002-12, 2016.

[17] C. Arsene and B. Gabrys, "Probabilistic finite element predictions of the human lower limb model in total knee replacement," *Med. Eng. Phys.*, vol. 35, no. 8, pp. 1116–1132, 2013.

[18] R. E. Anakwe, P. J. Jenkins, and M. Moran, "Predicting dissatisfaction after total hip arthroplasty: A study of 850 patients," *J. Arthroplasty*, vol. 26, no. 2, pp. 209–213, 2011.

[19] D. Hamilton *et al.*, "What determines patient satisfaction with surgery? a prospective cohort study of 4709 patients following total joint replacement," *BMJ Open*, vol. 3, no. 4, pp. 1–7, 2013.

[20] N. Mahomed *et al.*, "The self-administered patient satisfaction scale for primary hip and knee arthroplasty," *Arthritis*, vol. 2011, 2011, Art. no. 591253.

[21] V. Vanheule *et al.*, "Evaluation of predicted knee function for component malrotation in total knee arthroplasty," *Med. Eng. Phys.*, vol. 40, pp. 56–64, 2017.

[22] L. Blankevoort and R. Huiskes, "Ligament-bone interaction in a three-dimensional model of the knee," *J. Biomech. Eng.: Trans. ASME*, vol. 113, no. 3, pp. 263–269, 1991.

[23] M. Awad and R. Khanna, *Efficient Learning Machines*. Berlin, Germany: Springer, 2015.

[24] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[25] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[26] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*. MIT pressCambridge, MA, vol. 2, no. 3, 2006.

[27] M. Abadi *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*.

[28] J. Blank and K. Deb, "PYMOO: Multi-objective optimization in python," *IEEE Access*, vol. 8, pp. 89 497–89 509, 2020.

[29] K. Deb *et al.*, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[30] A. Seidenstein *et al.*, "Better accuracy and reproducibility of a new robotically-assisted system for total knee arthroplasty compared to conventional instrumentation: A cadaveric study," *Knee Surg., Sports Traumatol., Arthroscopy*, vol. 29, pp. 1–8, 2020.

[31] B. Echard, N. Gayton, and M. Lemaire, "Ak-mcs: An active learning reliability method combining kriging and monte carlo simulation," *Struct. Saf.*, vol. 33, no. 2, pp. 145–154, 2011.

[32] M. Faes *et al.*, "On the robust estimation of small failure probabilities for strong nonlinear models," ASCE-ASME *J. Risk Uncertainty Eng. Syst., Part B: Mech. Eng.*, vol. 5, no. 4, pp. 041007-1–041007-8, 2019.

[33] M. H. Gonzalez and A. O. Mekhail, "The failed total knee arthroplasty: Evaluation and etiology," *JAAOS-J. Amer. Acad. Orthopaedic Surgeons*, vol. 12, no. 6, pp. 436–446, 2004.

[34] J. G. Twiggs *et al.*, "Patient-specific simulated dynamics after total knee arthroplasty correlate with patient-reported outcomes," *J. Arthroplasty*, vol. 33, no. 9, pp. 2843–2850, 2018.