Functional Regression for Analyzing Human Ratings of Ultrasound Probe Alignment

Thomas Maierhofer[1]; Markus Iseli, PhD[1]; Eric Savitsky, PhD[2]
[1]UCLA CRESST, [2] SonoSim

Introduction

Human expert raters are the current gold standard for performance assessment, but can be substantially biased. Using a functional regression model, systematic flaws can be detected and quantified in a simulated process.

Simulations are a fast, reliable, scalable, and cost efficient way of training and assessment. This study presented experts with video recordings of participants aligning a handheld ultrasound probe with a virtual probe on screen, see Figure 1 and 2. A statistical model predicting human scorings from process data gained in the simulation is used to explore the subliminal scoring principles.

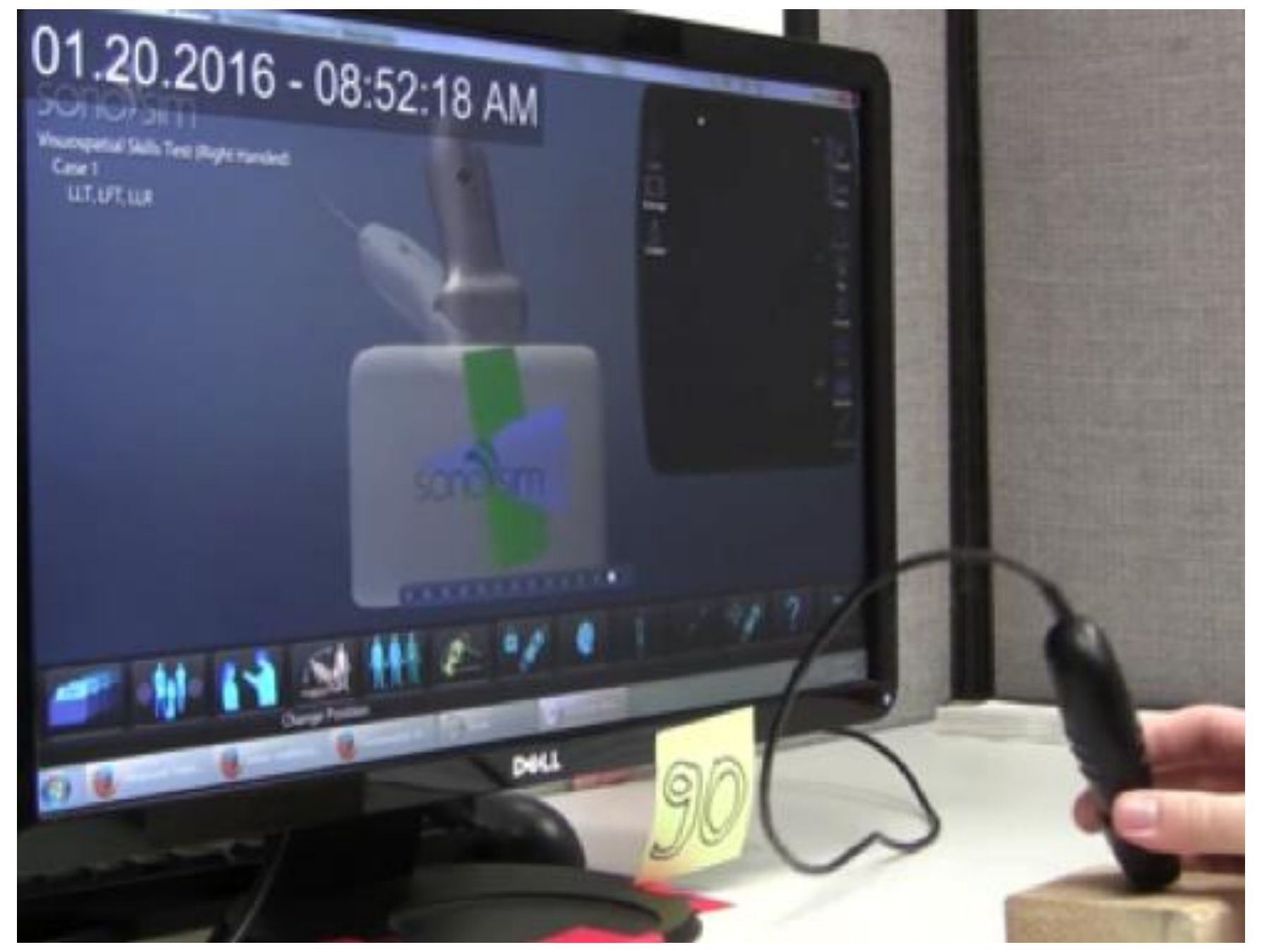


Figure 1. (left) Setup for the data collection. The SonoSimulator® was used to provide a simulation environment for probe alignment.

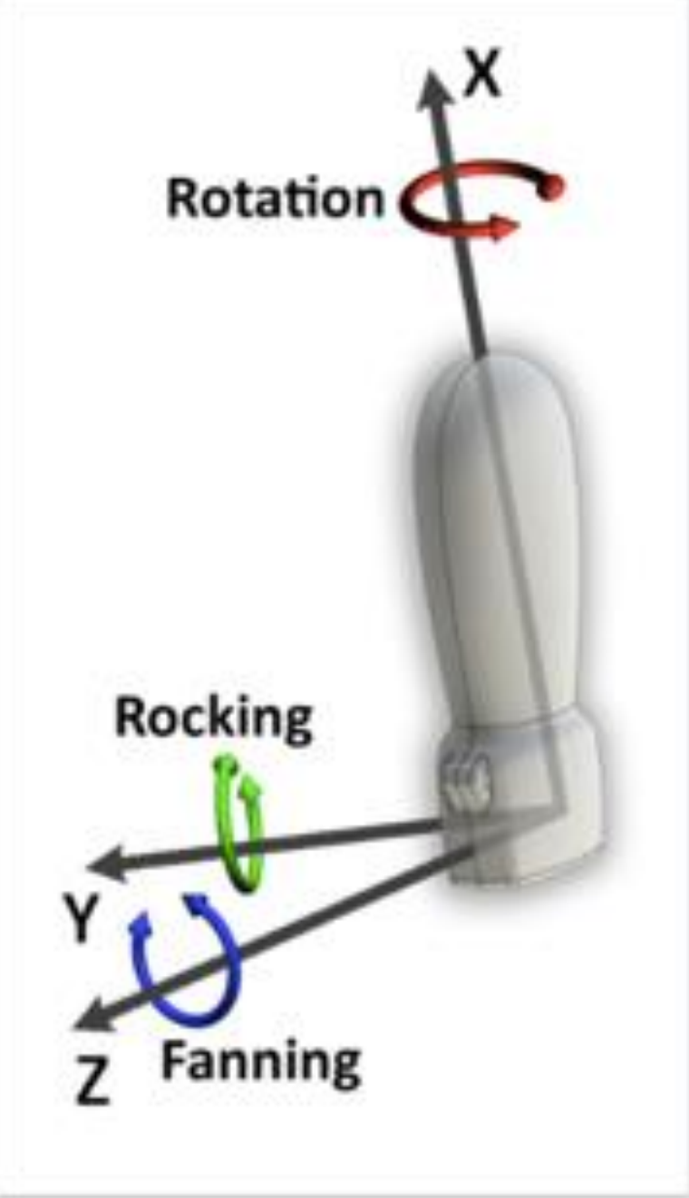


Figure 2. (right) Possible rotation axes for the ultrasound probe (Iseli, Savitsky, and Schenke, in review) .

54 study participants were asked to align a handheld ultrasound probe with a 3D probe on screen using the SonoSimulator®. All participants' hands, probes, and screens were video recorded. One task was selected and scored on a scale of 1 (poor performance) to 5 (outstanding performance) by three experts. Substantial inter-rater agreement (Krippendorff's alpha = 0.742, 95% CI = [0.682, 0.801]) allowed averaging.

Linear Regression

As a first analysis a linear regression model predicting the expert rating given the deviation in the final position and the length of the scanning path as proposed in Iseli et al., in review, is used. The resulting model formula is

$$score_i = \beta_0 + \beta_1 finalDev + \beta_2 pathLen + \varepsilon_i,$$

for participant i. The estimated model coefficients are reported in Table 1.

Functional Regression

Functional regression models are a generalization of linear regression models that allow covariates and/or the target variable to be functional (Morris, 2015). The general model formula for a scalar target variable y and functional covariates $x_1, x_2, \dots,$ is

$$y_i = \beta_0 + \int_T \beta_1(t)\, x_{1i}(t)dt + \int_T \beta_2(t) x_{2i}(t)dt + \cdots + \varepsilon_i\,,$$

for observation i, with intercept β_0, coefficient function β_j for functional covariate x_j on support T, and normally distributed error term ε_i. Non-functional covariates can naturally be included. Here, the target variable is the scalar average expert rating and the covariates are the scalar total time spent and the functional time-standardized deviation trajectories, see Figures 3 and 4. The model formula is

$$score_i = \beta_0 + f(timeSpent) + \int_0^1 \beta_1(t)\, rotDist_i(t)dt +$$
$$\int_0^1 \beta_2(t)\, rockDist_i(t) + \int_0^1 \beta_3(t) fanDist_i(t)dt \; + \varepsilon_i,$$

for participant i, with a non-linear effect for overall time spent, and rotation, rocking, and fanning distance as functional covariates. Figure 5 shows 95% confidence intervals for the model coefficients.

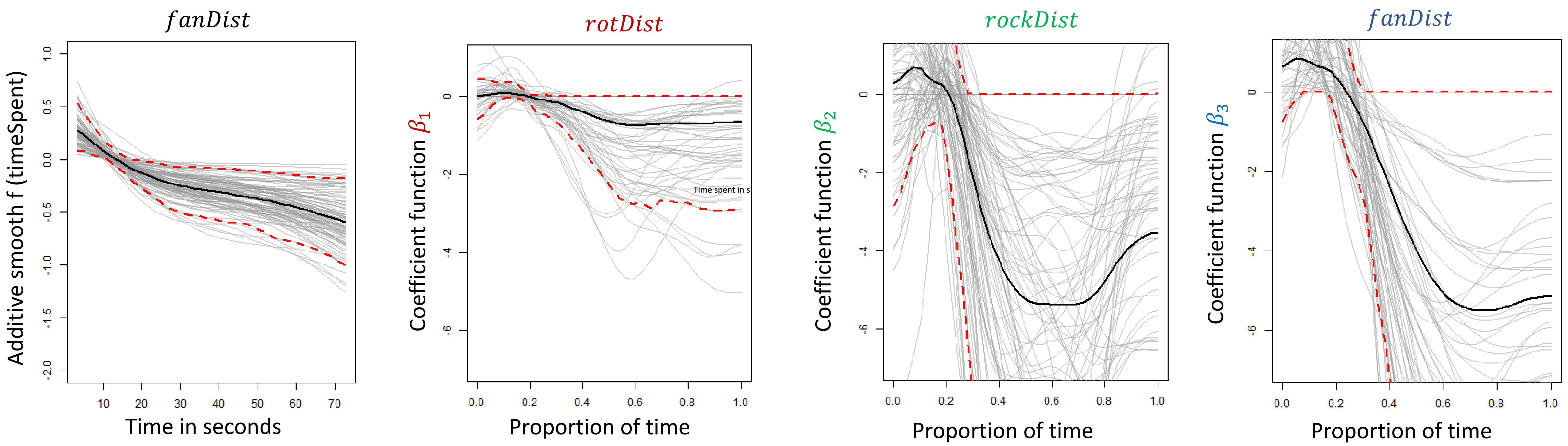


Figure 5. Nonlinear effect for time spent and coefficient functions for rotation, fanning, and rocking distance (solid black line) over relative share of time. A pointwise 95% confidence interval is estimated using boot strap (dashed red lines). The model was fitted using the R-package FDboost (Brockhaus and Rügamer, 2018).

Results

The estimated model coefficients of the linear regression model are reported in Table 1. As expected, there is a negative association of expert scores with final deviation and path length.

The functional regression model's estimated model parameters with 95% confidence intervals are depicted in Figure 5. As expected, there is a negative association of longer time spent with human ratings. The estimated functional coefficients for deviation in rotation, rocking, and fanning over time t show a similar pattern, where deviations at the beginning of the alignment process ($t < 0.2$) are hardly associated with the score, deviations in the middle of the process ($0.2 \leq t < 0.6$) have an increasingly stronger negative association, which levels out towards the end of the process ($t \geq 0.6$). Remarkably, the association of deviations in rotation with human ratings is much smaller than for fanning and rocking. This means that deviations in rotation did not lead to score reductions in the same way as did rocking and fanning deviations.

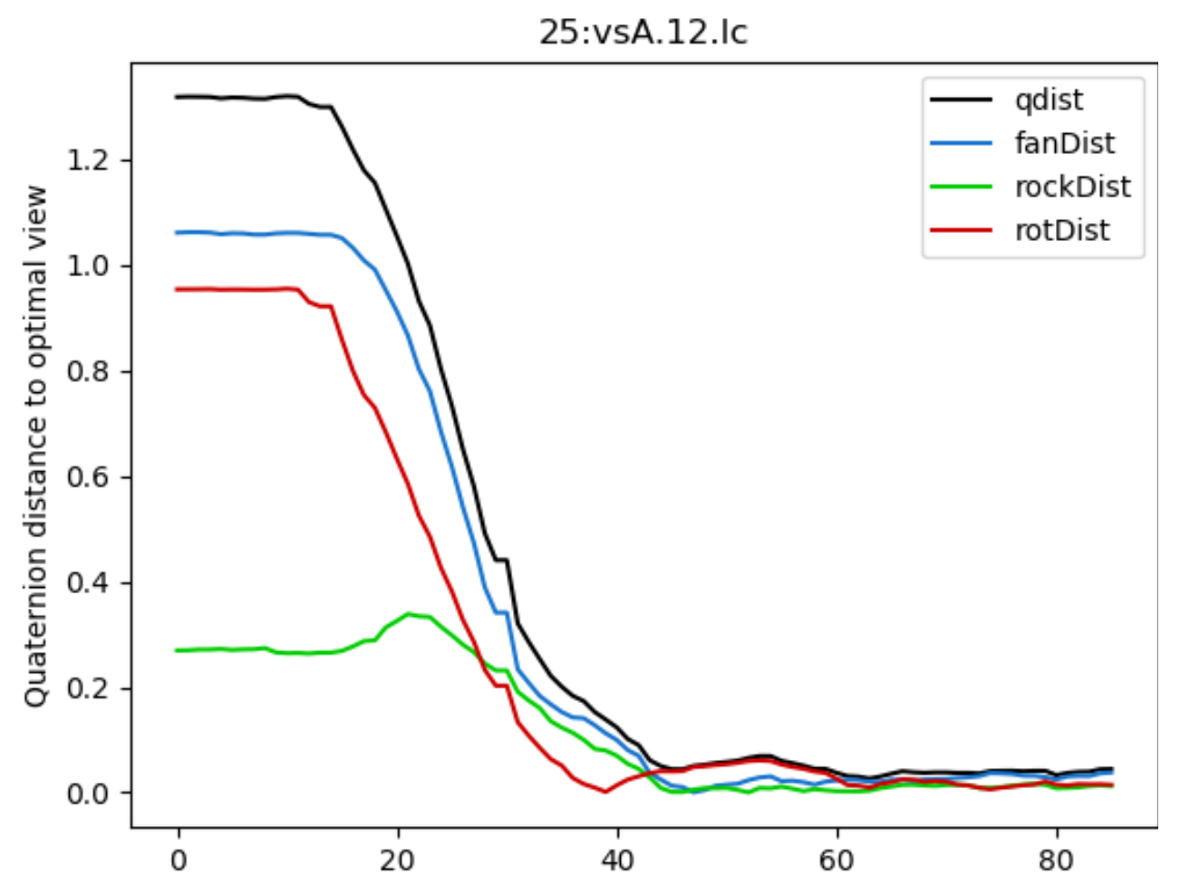
A 5 times repeated 10 fold cross validation showed a 73% decreased mean squared error between linear regression (MSE = 0.93) and functional regression (MSE = 0.25). This shows that the functional regression model is a better fit for the human ratings.

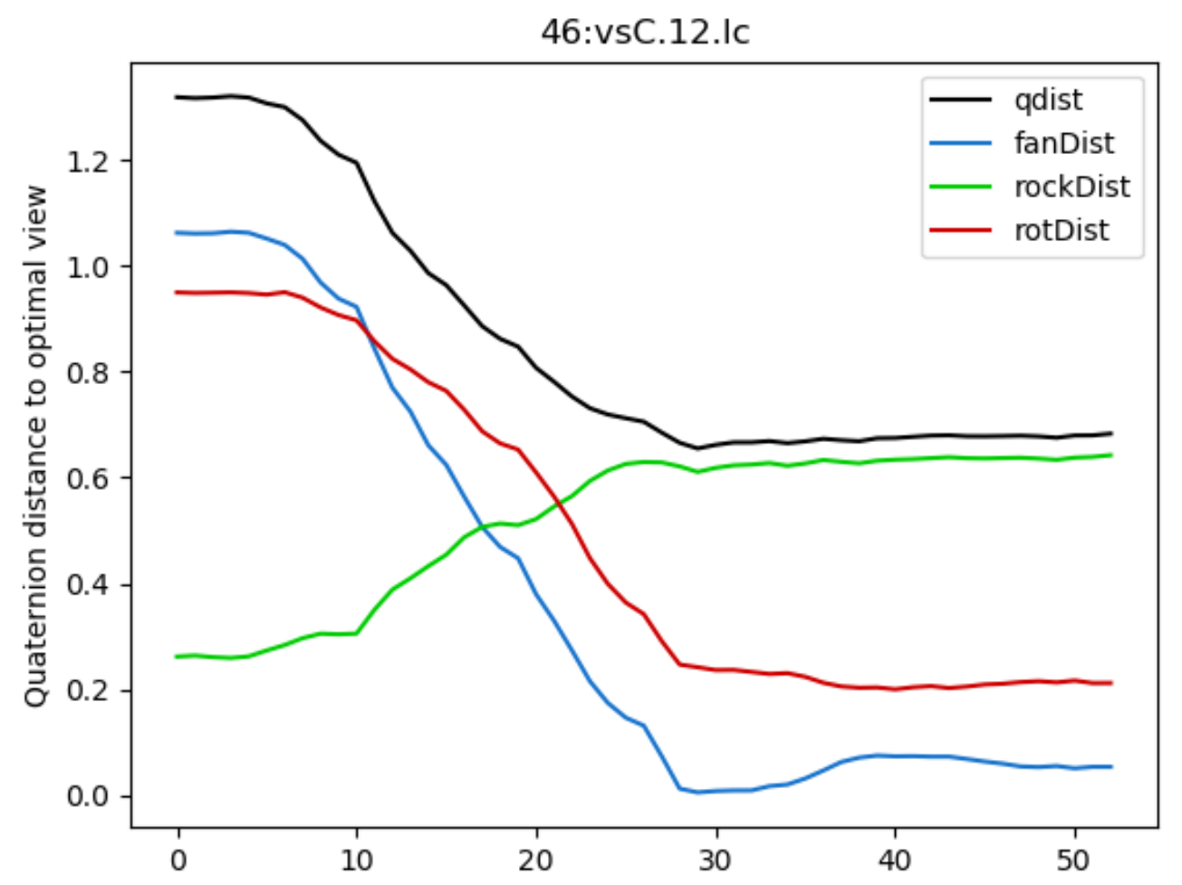
Table 1. Estimate, standard error, and p value in the linear model predicting average ratings.

	estimate	std. error	p value
Intercept	3.39	0.10	0.00
final deviation (1 sd)	-0.26	0.10	0.01
path length (1 sd)	-0.43	0.10	0.00

Figure 3 (left). High performance: Fast reduction in all rotation distances, small overall deviation (black) in final position.

Figure 4 (right). Low performance: Reduction in fanning (blue) and rotation distance (red) comes with increase in rocking distance (green).





Conclusions

Human performance ratings may serve as a baseline for performance evaluation in simulations but can be substantially biased. Functional regression is capable of detecting and describing such flaws in human judgement. As such, future work should focus on developing a scoring system that equally penalizes deviations in all directions of rotation.

Contact
Thomas Maierhofer
UCLA CRESST
290 Charles E. Young Drive N
maierhofer@cresst.org
+1-310-409-9679

References

1.Brockhaus, S. and Rügamer, D. (2018), FDboost: Boosting Functional Regression Models, R package version 0.3-1
2.Iseli, M. R., Savitsky, E., & Schenke, K. (in review). Simulation-Based Assessment of Psychomotor Skills for Ultrasound Competency Evaluation.
3.Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application*, *2*, 321-359.