‘A SIZE-AND-SHAPE REGRESSION MODEL FOR TRAJECTORY PLANNING OF A ROBOTIC ARM’

Michal Chojka
Student no: 25037607
Date: 02/09/2025

Supervisor: Ioanna Giorgi
Degree program: Advanced Computer Science MSc
UNIVERSITY OF KENT

UNIVERSITY OF KENT

ACCESS TO A MASTER’S DEGREE OR POSTGRADUATE DIPLOMA DISSERTATION

In accordance with the Regulations, I hereby confirm that I shall permit general access to my dissertation at the discretion of the University Librarian. I agree that copies of my dissertation may be made for Libraries and research workers on the understanding that no publication in any form is made of the contents without my permission.

Notes for Candidates: by submitting your dissertation, you agree with the following:

1 Where the examiners consider the dissertation to be of distinction standard, one copy may be deposited in the University Library and/or uploaded into Moodle as an example of good dissertation for future students.

2 If a copy is sent to the Library, it becomes the property of the University Library. The copyright in its contents remains with the candidate. A duplicated sheet is pasted into the front of every thesis or dissertation deposited in the Library. The wording on the sheet is:

> *“I undertake not to use any matter contained in this thesis for publication in any form without the prior knowledge of the author.”*

Every reader of the dissertation must sign and date this sheet.

3 The University has the right to publish the title of the dissertation and the abstract and to authorise others to do so.

SIGNATURE Michal Chojka **DATE** 24/08/2025

FULL NAME
(Please print, underlining surname, in order to assist the cataloguing of theses deposited in the Library.)

CERTIFICATE ON SUBMISSION OF DISSERTATION

I certify that:

1 I have read the University Degree Regulations under which this submission is made;

2 In so far as the dissertation involves any collaborative research, the extent of this collaboration has been clearly indicated; and that any material which has been previously presented and accepted for the award of an academic qualification at this University or elsewhere has also been clearly identified in the dissertation.

SIGNATURE Michal Chojka **DATE** 24/08/2025

This form should be completed and included in the submission of your dissertation.

1. Abstract:

This dissertation presents the development and evaluation of a novel pipeline for trajectory planning of a robotic manipulator based on statistical shape analysis. Building upon the artifacts introduced in “Regression Modeling for Size-and-Shape Data Based on a Gaussian Model for Landmarks’ (Dryden et al., 2020), the project integrated a size-and-shape regression model with polynomial regression, length-constrained alignment and a PyBullet simulation environment. Key technical contributions include the implementation of a cross-constrained alignment algorithm to ensure physical plausibility and the enhancement of predicted trajectories through polynomial smoothing and model-based landmark generation.

Results demonstrate that the proposed approach reliably predicts and simulates plausible hand and limb trajectories for various landmark configurations. The length-constrained alignment function effectively corrects physical abnormalities in regressed data, and the PyBullet simulation framework validates the feasibility of trajectories effectively dealing with errors in the data. Limitation of the regression model regarding simple landmarks sets are discussed, alongside the strengths of the polynomial smoothing and error correction modules.

Concluding, this work establishes a robust pipeline for statistical landmark regression and robotic trajectory planning, offering great value for future research in biomechanical simulation and robotics. Recommendations for further improvements include migration of the core regression model entirely to Python and enhancing interactive simulation capabilities. The findings support the practical application of this work in robotic control and trajectory planning.

2. Table of content:

Contents

3. Introduction

3.1 Technical Background

Advanced statistical shape models have emerged as powerful tools for modelling and predicting complex motions, yet their direct application in robotics remains limited. While these models excel at capturing subtle variations in shape and size, they frequently produce trajectories that violate the physical constraints essential for robotic systems—rendering their outputs impractical for real-world deployment. Bridging this gap between elegant mathematical theory and robust engineering practice is a critical challenge facing both researchers and practitioners in robotics today. This dissertation addresses this challenge by developing and validating a novel pipeline that unites statistical shape analysis with physical plausibility enforcement, culminating in a fully integrated solution for trajectory planning and robotic simulation. By combining a state-of-the-art size-and-shape regression model with a custom length-constrained alignment algorithm and leveraging the PyBullet physics environment, this work establishes a practical framework for generating physically plausible, data-driven trajectories suitable for advanced robotic applications.

3.2 Research Objectives

The primary objective of this dissertation is to implement a robust pipeline for trajectory planning in robotic manipulators with the use of statistical shape analysis techniques. Specifically, this work adapts the artefacts from Dryden et al. (2020) to perform regression modelling on size-and-shape data for landmark-based trajectory prediction, with the plan to evaluate its effectiveness in a robotic simulation.

3.3 Motivation and Context

Trajectory planning is a fundamental problem in robotics, particularly for agents requiring precise movement and manipulation. Landmark-based statistical models, originally developed for landmark analysis in biology and medicine, have recently shown promise for predicting movement in complex systems such as robotic hands. This dissertation explores the potential of those methods in robotics, focusing on size-and-shape regression, data smoothing, physical plausibility constraints, and validation through realistic simulation.

3.4 Research Questions

This work is guided by the following research questions:

- Can statistical regression model for size-and-shape landmarks data be effectively applied in trajectory planning and robotic simulation?
- How do such approaches compare to traditional regression and simulation techniques in terms of physical plausibility and accuracy?
- What challenges arise when integrating such a model with robotic simulation framework?

3.5 Methodological Overview

To address these questions, the project implements a pipeline that includes:

- Retrieving and preprocessing part for landmark trajectory data
- Application of a size-and-shape regression model (via an R-Python integration).
- Enhancement of trajectories using a polynomial for smoothing.
- Correction of landmark configurations using a length-constrained alignment algorithm.
- Validation and visualisation through the PyBullet physics simulation.

3.6 Dissertation Structure

The organisation of this dissertation follows:
- **Chapter 1** abstracts the problem and summarises the issue that the paper addresses.
- **Chapter 2** lists all the chapters in order and their subheading too.
- **Chapter 3** introduces the problem in detail, defining its relevance and associated challenges.
- **Chapter 4** presents background concepts, including core methods, technologies, and a description of method inputs and outputs.
- **Chapter 5** reviews relevant literature and technologies.
- **Chapter 6** describes the methodology, including all major pipeline components.
- **Chapter 7** details the results, testing, and verification procedures, with adequate analysis.
- **Chapter 8** offers critical evaluation and reflection on the approach and artefacts.
- **Chapter 9** concludes with a summary of contributions and recommendations for future research.
- **Chapter 10** links all the related sources and contributors to this project.

4. Problem Statement

Robotic trajectory planning requires accurate, physically plausible predictions of recorded landmarks. Traditional kinematic approaches may not fully address the complexities of shape change present in landmark-based data. This dissertation addresses the challenge of adapting statistical size-and-shape regression models - originally developed for ‘morphometric analysis’ (Dryden et al., 2020) - to predict and validate robotic manipulator trajectories.

The main problem is to determine whether landmark-based statistical regression can reliably generate trajectories that respect the mechanical and anatomical constraints of robotic systems. Key considerations include preprocessing landmark data to remove confounding factors, ensuring predicted trajectories are physically valid, and integrating results with a robotic simulation environment for verification.

This work focuses on:
- Implementing a regression pipeline for landmark trajectories prediction.
- Enforcing physical plausibility through length-constrained alignment.
- Validating trajectories in realistic robotic simulations.

By tackling these challenges, the dissertation aims to establish whether statistical shape analysis can enhance trajectory planning in robotics.

5. Background

5.1 Summary

This chapter introduces the essential concepts and technologies forming the foundation of the dissertation. The interplay between landmark-based modelling, statistical regression, physically plausible alignment and robotic simulation forms the basis for the proposed pipeline, which is detailed in subsequent chapters.

5.2 Core Concepts

5.2.1 Landmark Data in Robotics

Landmark data refers to the sets of specific, measurable points that describe the geometry and configuration of an object. In robotics, landmarks are commonly used to represent joint positions, fingertip location, or other features present in artificial manipulators and limbs. Each landmark is typically described by its spatial coordinates within a given frame of reference.

The full landmark specification of an object’s movement over time is captured from multiple landmark sets across the sequential frames, which helps researchers with modelling the dynamic trajectories. Landmark data includes four principal features: translation, rotation, scale (size), and shape. For statistical analysis, which present in size-and-shape regression, it is common practice to remove translation and rotation through techniques as Helmert transformation (Dryden & Mardia, 2016), isolating size and shape from other parameters as the primary variables of interest.

5.2.2 Size and Shape Regression

Size and shape regression is a statistical modelling technique for predicting the configuration of landmarks over time. This approach, based on a Gaussian model for landmark data (Dryden et al., 2020), is used to analyse and predict changes in object size and shape, originally for biological context, but here adapted for robotics simulations.

5.2.3 Trajectory planning

Trajectory planning in robotics is a task that involves predicting a sequence of positions and orientations for an artificial manipulator or limb such that it can accomplish a specific task while respecting its physical constraints. Landmark-based regression models aim to deliver those trajectories for simulation environments and provide an alternative to traditional kinematic approaches, potentially providing more flexible and generalizable predictions.

5.3 Technologies and Tools

5.3.1 Python and R Integration

Python have been selected for this project for its strong ecosystem in data science, machine learning and integration with robotic simulation (NumPy, Pandas, Scikit-learn, PyBullet), while R is used for specialized size and shape regression and modelling. The Rpy2 package provides tools for communication and data exchange between Python and R, enabling the use of advanced size-and-shape regression model within a Python workflow.

5.3.2 PyBullet Simulation Environment

PyBullet is an open-source physics engine, natively used for physical simulations, but also widely adapted for robotics research and prototyping. It accepts robot configurations files (URDF) and can visualize and animate manipulator trajectories, providing a practical platform for validating the outputs of regression models.

5.3.3 Data Analysis and Visualisation

Packages such as NumPy, Pandas, SciPy, and Matplotlib are integral for data management, statistical computation, optimization and visual representation of results throughout the pipeline.

5.4 Input and Output Overview

The following table summarises the main inputs and outputs of each component in the pipeline:

Method/Function	Input(s)	Output(s)
Size & Shape Regression	Raw landmark array (K x D x F)	Predicted landmarks (K x D x F)
Conversion Pipeline	Python data, R regression model	Python-compatible array of predicted landmarks
Polynomial Regression	Predicted landmarks, Polynomial degree	Smoothed/model-like landmarks (K x D x F or K x D)
Length-Constrained Alignment	Predicted landmarks, Reference, Connections set, Lengths set	Aligned, physically plausible landmarks (K x Dx F)
PyBullet Simulation	Landmarks per frame, Robot configuration (URDF)	Simulated movement & visualisation, Error metrics,

6. Literature and Technology Review

6.1.1 Theoretical Foundations

Statistical shape analysis is a standing field in mathematics and statistics, focusing on the study of the geometry of objects through for example, landmark data (Dryden & Mardia, 2016). Landmark-based methods have been widely adopted in ‘biological morphometrics, medical diagnostics, and computer vision and robotics. The central interest of this approach is the representation of object shape via a set of specific points in space, allowing for the rigorous quantification and comparison of change in shape.

Regression modelling for shape and size data, as developed by Dryden et al. (2020), improve these concepts by creating a way for predicting landmark configuration as functions of explanatory variables (such as time or different cases). The use of Gaussian models provides a framework for handling uncertainty and variation in the data and landmark position, supporting robust inference and synthesis of new configurations.

6.1.2 Landmark Regression in Robotics

The statistical shape analysis is a growing area, especially in robotics, and we are looking for more method for its application. Robotic manipulation and trajectory planning traditionally rely on kinematic models and optimisation algorithms (Siciliano & Khatib, 2016), which may struggle to generalise across diverse configurations and noisy real-world data. Landmark-based regression offers a data-driven alternative, potentially providing better adaptability and robustness in dynamic environments.

6.2 Review of Software Tools and Technologies

6.2.1 Python and R

Python is known for its versatility in computing, machine learning, and robotics environments, supported by libraries such as NumPy, Pandas, and Scikit-learn (Oliphant, 2006; McKinney, 2010). R, meanwhile, excels in statistical modelling and data analysis, with specialised packages for shape analysis and regression (Dryden & Mardia, 2016). The Rpy2 library enables integration of R methods into Python processes, facilitating cross-language data exchange and computation.

6.2.2 PyBullet

PyBullet is a widely used physics simulation engine, providing real-time dynamics, collision detection, and incorporates robotic control within a Python-accessible framework (Coumans & Bai, 2016). It supports the Unified Robot Description Format (URDF), allowing for flexible robot implementation and offers visualization tools for trajectory analysis and debugging.

6.2.3 Data Analysis and Visualisation

NumPy and Pandas are core packages used for efficient computation and data management in Python (Oliphant, 2006; McKinney, 2010). SciPy provides advanced algorithms for optimisation and minimisation, integral to the length-constrained alignment method in this project. Matplotlib is used for the visualisation of landmark data and regression results (Hunter, 2007).

6.3 Review of Literature Review

6.3.1 Size and Shape Regression Model

The size and shape regression model is a statistical approach for analysis and prediction the configuration of landmark data following geometric shape (shape) and scaling differences (size). Dryden et al. (2020) introduced regression modelling for size-and-shape data using a Gaussian model to deal with uncertainties, enabling the prediction of landmark position as a function of explanatory variables such as time, physical conditions and more.

In this model, a set of landmarks is represented as a matrix of coordinates, and variation among observations is divided into translation, rotation, scale and shape. After removing translation and rotation by Helmert transformation or Procrustes alignment, the centroid size and aligned shape coordinates are used as core descriptors of the data. The Gaussian regression framework allows for simultaneous modelling of centroid size and shape, providing a flexible means to deal with uncertainties in landmark configurations.

The main advantage of the size and shape regression model is its ability to capture changes in size and shape spectrum (morphological changes), while maintaining a rigorous statistical basis for uncertainty estimation. Applications in biology have included the analysis of growth patterns, disease progression, and treatment effects (Dryden & Mardia, 2016; Paine et al., 2020). In robotics, it is less common, but it offers great potential for improving trajectory predictions and alternative implementations.

Several limitations have been identified, too. The model’s accuracy may depend on the quality and density of the landmark data and Gaussian assumptions may not always hold in highly constrained systems as robotic manipulators. Furthermore, the regression model does not enforce any physical constraints (e.g., fixe link lengths), which is realistic assumptions in medical field, it would not hold in real-world robots unless additional correction algorithms have been introduced.

This dissertation adapts and extends the size and shape regression model for robotic trajectory planning, implementing post-processing steps as length-constrained alignment to address physical plausibility. By critically evaluating its performance in simulation, this work contributes to the understanding of statistical morphometric methods in robotics.

6.4 Summary of Literature Gaps and Technological Opportunities

While statistical shape regression models have proven successful in biological and medical fields, their application in robotics remains limited. Key gaps include:
- The need for physical plausibility in trajectory generation, respecting robotic constraints.
- Efficient integration of statistical modelling with simulation environments.
- Handling of landmark data with varying levels of complexity and noise.

This dissertation aims to address these gaps by combining advanced statistical modelling methods with practical simulations tools, evaluating the utility of landmark-based regression for robotic trajectory planning.

7. Methodology

7.1 Overview

This chapter details the methodology for this project, as well as adapting statistical shape analysis and regression models to the problem of robotic trajectory planning. The proposed pipeline integrates landmark data acquisition, regression modelling, prediction improvement and smoothing, physical plausibility enforcement, and simulation-based validation. This approach is motivated by practical challenges encountered during this project, with detailed explanations, justification for each method, and explicit documentation of data sources, assumptions, and pipeline modularity. Design emphasises modularity, user control, and extensibility, ensuring applicability across both biometrics and robotics.

7.2 Comparative analysis of Approaches

7.2.1 Simulation Platform Selection

Two major platforms were considered for creating a robotic simulation to test the artefacts:

- ROS (Robot Operating System): While feature-rich and widely used, ROS’s complexity and Linux-only support made it impractical for rapid prototyping and cross-platform use within this project.
- PyBullet: PyBullet is primarily a physics simulator but offers robust rigid body dynamics, Python integration, and strong documentation. A dedicated testing script was created to validate PyBullet’s capabilities for animating and optimising articulated limbs. Results confirmed PyBullet as suitable and flexible for trajectory planning, justifying its selection.

7.2.2 Programming Language and Ecosystem

Python was chosen for its rich ecosystem (NumPy, Pandas, Matplotlib, Scikit-learn) and compatibility with PyBullet. Its flexibility enabled efficient prototyping and integration of various modules.

7.2.3 Regression Model Implementation

The size-and-shape regression model, a core part of the pipeline, was supplied in R by the original professor of Dryden et al. (2020). Direct re-implementation in Python was attempted but found impractical due to the lack of specific libraries and complexity, but it would be possible to implement if more time were provided for the project. Instead, the Rpy2 package was adopted for Python-to-R communication, allowing use of validated peer-reviewed code and focusing development on pipeline integration and optimisation, by directly implementing the R size-and-shape regression model to Python processes.

7.2.4 Addressing Output Discrepancies

Physical implausibility (e.g., segment length violations) in regression output was identified as a major challenge, caused by statistical alignment (Helmert) and lack of kinematic constraints. Modifying the R model to respect physical plausibility was not feasible due to the lack of direct re-implementation of the model in Python and experience in R; instead, a dedicated post-processing step (Length-Constrained Alignment) was developed, tailored for robotic simulation needs, making simulation possible.

7.2.5 Data Acquisition and Validation

Professor’s Dataset: 3D hand landmark configurations (21 landmarks, 10 frames) were acquired directly the original authors, serving as a benchmark and external validation reference. Those landmarks resemble growth of a hand.

Dummy Hand & 3DoF Robot: Custom landmarks datasets were created for the purpose of the validation. Dummy Hand (21 landmarks, 15 frames) resemble landmarks of artificial simplified hand and was used for grasping animation tests. 3DoF robot (4 landmarks, 15 frames) is configuration of landmarks of the 3 degrees of freedom robot that was implemented for simple trajectory and regression model testing.

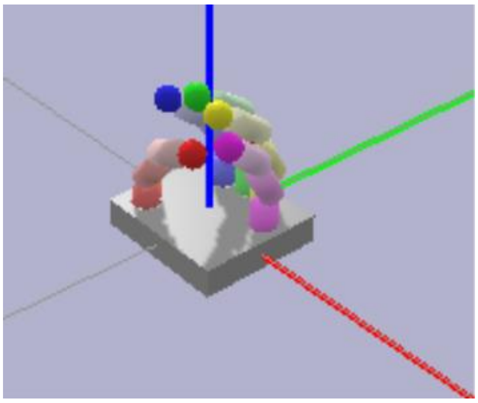


Figure 7.1: Dummy hand robot in PyBullet simulation

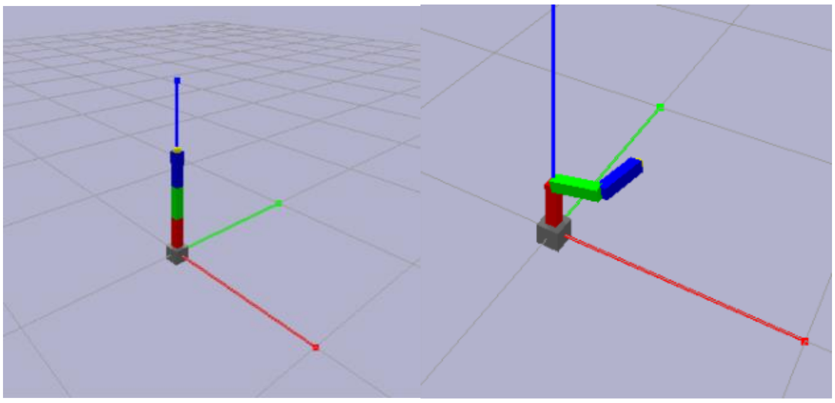


Figure 7.2: 3DoF Robot in PyBullet simulation

Validation Process: All datasets were inspected visually and numerically for consistency. The professor’s data was compared to published results, while custom datasets were tested iteratively to ensure suitability for regression, alignment and simulation.

7.2.6 Size-and-Shape Regression Model (in R)

Summary & Purpose:
Statistical backbone of the pipeline, enabling prediction of landmark configuration in size and shape spectrum as a function of explanatory variables (e.g., time). Robust for high-dimensional, multi-landmark datasets; best suited for biometrics applications, and this project aims to find its applicability in robotic trajectory planning.

Why This Method:
This method was an initial inspiration for this project, providing a real alternative to standard kinematic models that focus on joint positions and ignore subtle shape changes. It handles both geometric shape and overall size changes, capturing complex motion patterns with statistical rigour, being the statistical core of the function. Using this function, we aim to capture a full-size change of the landmarks over time and use it for trajectory planning for a set of joints.

Implementation Details:

- Alignment using Helmert transformation.
- Least squares regression under Gaussian assumptions.
- EM algorithm for parameter optimisation.
- Prediction of new configurations using fitted coefficients.

Assumptions:
Requires high-dimensional, well-sampled landmarks data for accurate predictions.

Limitations:
Does not enforce physical/kinematic constrains; best fit for biometrics but less so for robotics without tailored post-processing methods.

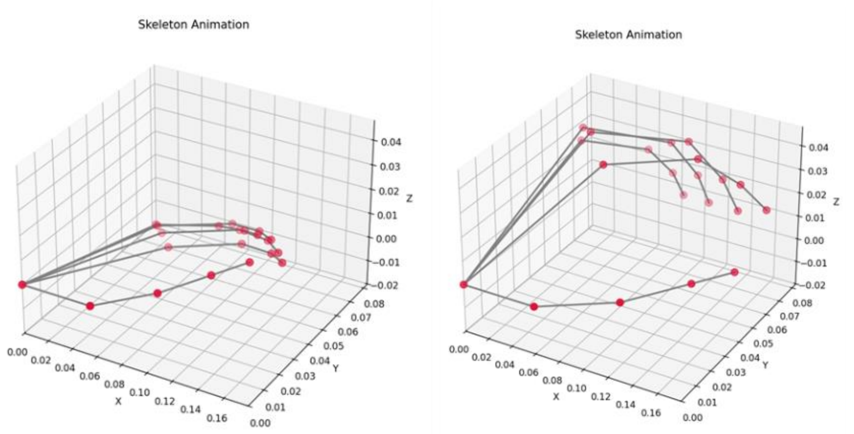


Figure 7.3: Size-and-shape regression landmarks (professor’s hand)

7.2.7 Conversion Pipeline (Python to R)

Summary & Purpose:
Facilitates integration of the R-based regression model into Python. Converts data structures

between NumPy and R routines and retrieves results for further processing. The alternative to implementing a size and shape regression model in Python.

Why This Method:
Direct Python re-implementation was infeasible; Rpy2 enabled rapid, validated integration.

Implementation Details:

- Conversion of multidimensional arrays.
- Error handling for data type and shape mismatches.

7.2.8 Polynomial Regression (Optional)

Summary & Purpose:
Polynomial regression smooths and regularizes landmark trajectories and generates model-like predictions; continuous landmark sequences for animation and predictive planning. This model is particularly effective for biometrics and simple robotic data for smooth underlying motions but also works as post-processing method for size-and-shape regression model.

Why This Model:
Provides interpretable, computationally efficient smoothing and the ability to create predictions beyond original frames. The method wasn’t initially planned, and it was inspired during the development of the size-and-shape method. Polynomial model proved to outperform the size-and-shape regression model in simple robot trajectory planning, and when working together demonstrated ability in smoothing the data due to the polynomial curve.

Implementation Details:

- Uses `PolynomialFeatures` and `LinearRegression` for each landmark.
- `poly_smoothing` parameter enables smoothing for noisy data.
- `poly_model` parameter enables generation of experimental continuous model-like predictions at arbitrary time points, producing a denser motion trajectory.

Model-like Predictions:

- Model-like predictions are created by evaluating the polynomial fit at many time points (e.g., 2-10x original frames), yielding smooth, continuous landmark trajectories for animation, planning or simulation.
- Output is an array of predicted positions: [landmarks, model frames].

Assumptions:
Underlying trajectories are smooth and can be meaningfully approximated by a polynomial (feasible as it usually deals with robot landmarks provided from size-and-shape regression)

Limitations:
Over-smoothing may obscure sharp transitions in the landmarks, distorting the real transition, therefore it is optional method. Additionally, it is most applicable to simple robot motions.

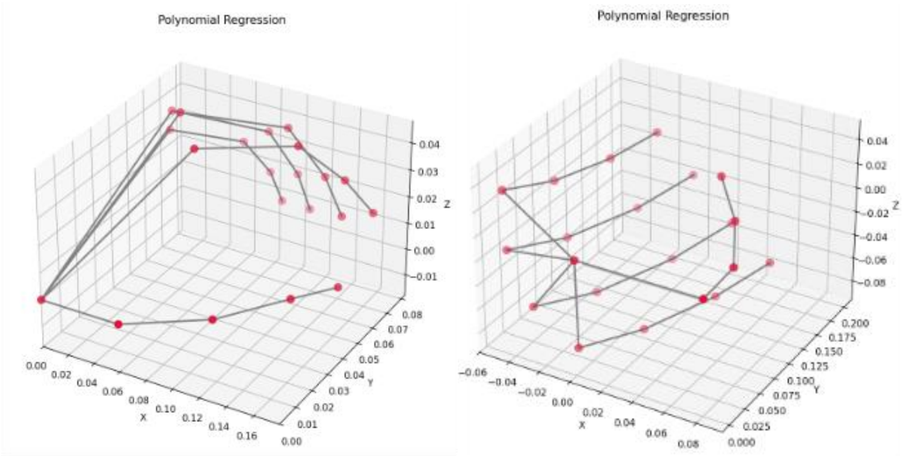


Figure 7.4: Polynomial Regression landmarks (Professor’s landmarks, Dummy Hand)

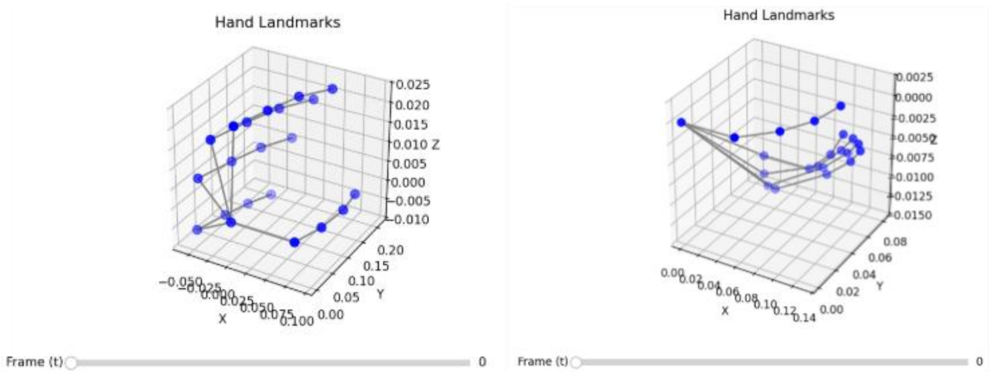


Figure 7.5: Polynomial model-like landmarks (Dummy Hand, Professor’s landmarks)

7.2.9 Length-Constrained Alignment

Summary & Purpose:
Custom post-processing algorithm designed to restore physical plausibility to regressed landmark trajectories. Ensures segment lengths and anchor positions remain consistent with the robot’s kinematic structure.

Why This Method:
Regression models (especially after Helmert alignment) ignore robot-specific constraints, resulting in misaligned or physically impossible poses for robotic agents. This method is tailored for robotics and physical simulations, being a direct response to size-and-shape regression misalignment and lack of physical plausibility in data, addressing both issues directly, preparing data for physical simulation.

Implementation Details:

- Anchoring: Fixes the landmark/s position in the original position (base), to prevent global drifting.
- Alignment (Kabsch Algorithm): Minimises RMSD between predicted and reference landmarks, correcting for overall rotation/translation.

- Length Constrained Optimization: Nonlinear optimization enforces true segment lengths while maintaining anchoring and alignment.

Logical Step Reasoning:
Anchoring establishes a stable frame, anchoring shape to its base; global alignment follows, then local length and alignment adjustments ensure kinematic feasibility. Minor errors may persist but are corrected downstream in the simulation.

Assumptions:
User provides reference landmark configuration, segment connectivity, and true lengths. Anchoring is essential for robotics, less so for biometrics.

Limitations:
Requires accurate robot topology and connection lengths, tailored for robotic and simulation tasks.

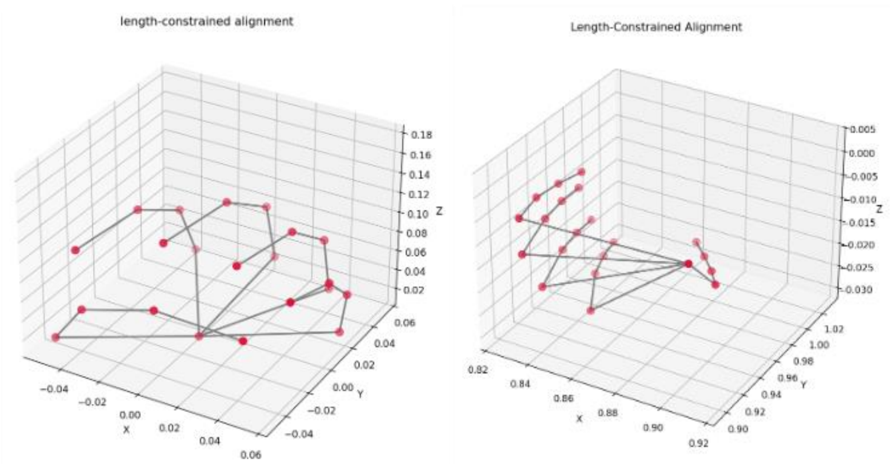


Figure 7.6: Length-constrained alignment (Dummy hand, professor’s landmarks)

7.2.10 PyBullet Simulation Framework

Summary & Purpose:
Validates physical plausibility and utility of predicted landmark trajectories by animating the robot in a physics-based simulation. Enables visual, quantitative, and practical assessment of pipeline outputs.

Why This Method:
Simulation is essential for confirming that statistical predictions are feasible and useful in real robotics systems, being a milestone in proving that this model is applicable in real-life scenarios.

Implementation Details:

- Loads URDF robot models
- Applies landmark trajectories as motion guides
- Optimises joint angles for each frame, subject to physics constraints.
- Reports quantitative error and provides real-time visualisation.

Assumptions:
Input landmark data is physically plausible (as ensured by previous steps); minor errors can be corrected by preprocessing algorithm.

Limitations:
Simulation accuracy depends on the quality of prior alignment and constraint enforcement.

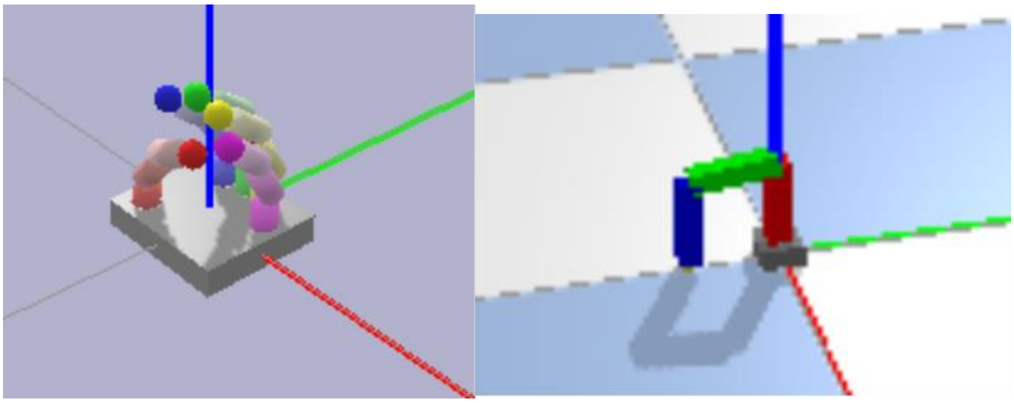


Figure 7.7: PyBullet simulation (Dummy hand, 3Dof robot)

7.3 Pipeline Modularity and Optionality

Each most of the modules (regression, polynomial smoothing/model generation, alignment, simulation) can be enabled or disable depending on user needs and data characteristics (either with parameters or direct editions to the code).
Size-and-shape regression and polynomial regression are well suited for biometrics and morphometrics, but does not prevent regression from generating physically unplausible outputs; length-constrained alignment and simulation are essential for robotics and physical plausibility, but distort the original size-and-shape and polynomial regression and possibly should not be tested for morphometrics or biometrics.
Model-like outputs from polynomial regression are intended for future predictive planning or high-fidelity animation, which may be implemented in future studies using more advanced methods.

7.4 Collaboration and Sourcing

Professor’s dataset and original regression model code were sourced via direct communication with the model author/s, ensuring external validation and reproducibility and tested manual by the author of this paper.
Custom datasets (Dummy Hand, 3DoF robot) were designed and validated by the author for use in testing and development.

7.5 Summary

The methodology combines advanced statistical modelling, tailored post-processing for physical plausibility, and simulation-based validation. All components are justified theoretically and practically, with explicit documentation of data sources, assumptions, and pipeline modularity to ensure transparency, reproducibility, and extensibility.

8. Results, Testing and Verification

8.1 Overview

This section details the testing and evaluation process for the trajectory planning pipeline. Each test was motivated by a specific hypothesis or validation goal: verifying model correctness, assessing applicability to different dataset complexities, and ensuring physical plausibility for simulation. Visual inspection, quantitative error analysis, and simulation feedback were all used to interpret results and guide improvements.

8.2 Visual Inspection and Figure-Based Validation

For every regression and alignment step, results were validated both numerically and visually.

- Visual Inspection: Initial and regressed landmark configurations were compared side-by-side for plausibility and anatomical consistency.
- Figures: Before-and-after images for each dataset (professor’s hand, Dummy Hand, 3DoF robot) are included to illustrate method effects and help document orientation changes, misfits, and trajectory smoothness.

8.2.1 Size-and-Shape Regression Model

(Benchmark Test with Professor’s Dataset)

Reason for Test:
To validate the regression model using externally sourced, benchmark data and to establish a reference point for integration accuracy, by comparing the results of the conversion framework with the results that were run in the R program shared by researchers.

Method:
Regression performed on the professor’s 21-landmark, 10-frame hand dataset, compared against the original regression run fully in R

Results:

- Stable, reproducible trajectories, visually and quantitatively consistent with published results.
- Minor misfits and orientation changes noted, and outlining first-frame landmarks.
- Figure 8.1: Comparison of initial and regressed landmarks.
- Results in R proved to be identical to those gathered in Python.

Importance:
Confirmed model integration; benchmarked pipeline accuracy.

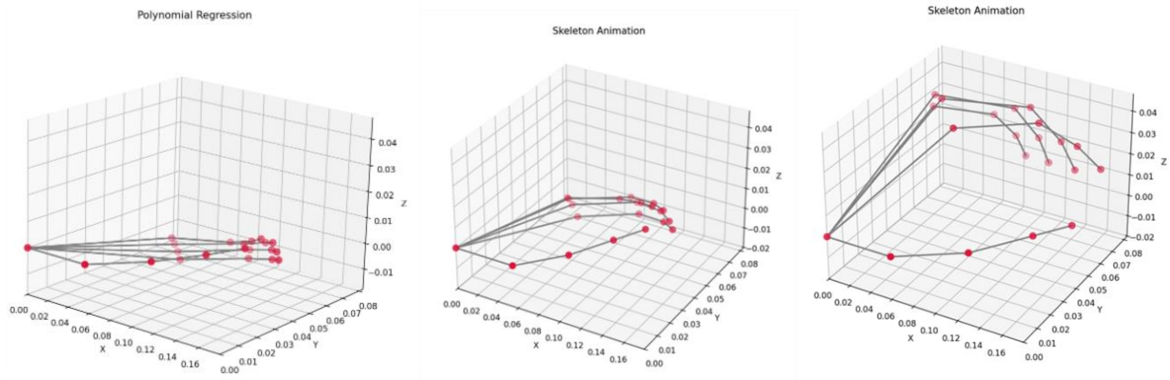


Figure 8.1: Plot for size-and-shape regression of professor’s landmarks (including initial misfit)

8.2.2 Dummy Hand Test (Complex Custom Data)

Reason to Test:
To evaluate model generalisation to custom, complex hand trajectories created for this project.

Method:
Regression applied to the Dummy Hand dataset (21 landmarks, 15 frames) and results were inspected with as a plotted data.

Results:

- Produced plausible, continuous grasping motion.
- Orientation flips and first-frame misfits observed (see Figure 8.2)
- Length-constrained alignment partially corrected errors (not part of the test).

Importance:
Demonstrated pipeline utility for novel data, highlighted the need for post-processing.

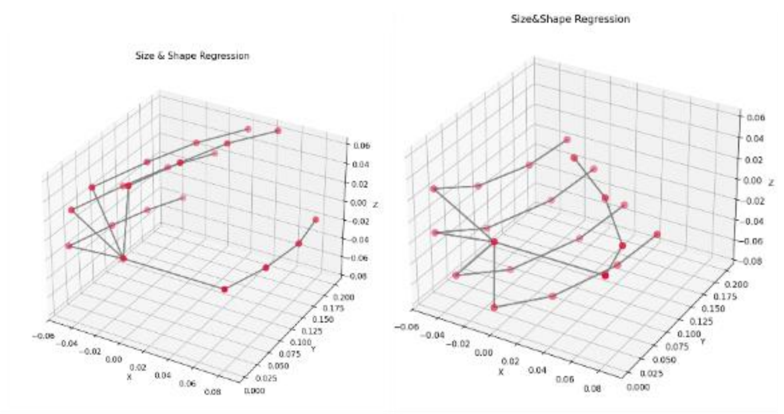


Figure 8.2: Plot for size-and-shape regression of dummy hand (including initial misfit)

8.2.3 3DoF Robot Test (Simple Custom Data)

Reason for Test:
To test the regression model robustness on minimal, low-dimensional robotic data from a robotic arm.

Method:
Regression was performed on the 3DoF robot dataset (4 landmarks, 15 frames), and the results were inspected as plotted data.

Results:

- Regression failed: outputs were erratic, physically implausible, and did not reflect any real trajectory.
- Visual inspection confirmed a lack of meaningful trajectory (see Figure 8.3).

Importance:
The revealed model overfitting and limitations for sparse datasets suggest the need for alternative methods to address this issue.

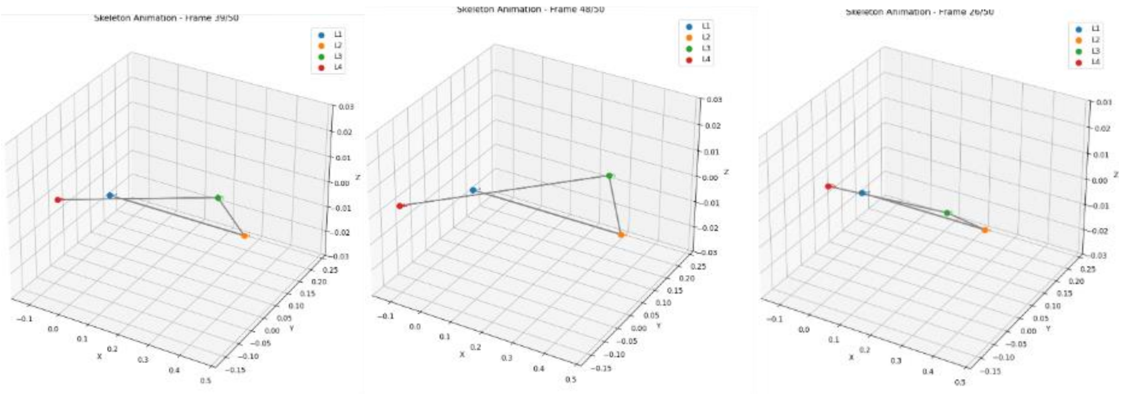


Figure 8.3: Plot for size-and-shape regression of 3DoF robot

8.2.4 Polynomial Regression

As a Stand-Alone Model with PyBullet simulation (3DoF Robot)

Reason for Tests:
To determine if polynomial regression operating on Cartesian coordinates could effectively model simple robotic trajectories where statistical shape models fail, and be verified with PyBullet simulation.

Method:
Fitted polynomial curves to each landmark in the 3DoF robot dataset and applied them to an experimental PyBullet simulation after a length optimisation process using a prototyped algorithm.

Results:

- Produced smooth, physically plausible robot trajectories.

- Outperformed size-and-shape regression for this scenario.
- PyBullet simulation proved to be consistent with the researcher’s expectations (Figure 8.4).

Importance:
Established polynomial regression as preferable for simple robotic configurations and further discredit the size-and-shapemodel performance on low-dimensional landmarks.

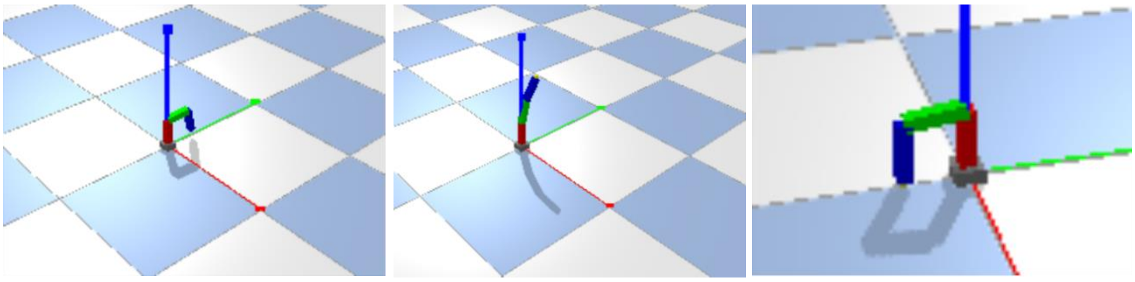


Figure 8.4: 3DoF robot simulated in PyBullet simulation with regressed landmarks by a Polynomial model

8.2.5 Smoothing Post-Processing Method (Hand Datasets)

Reason for Test:
To evaluate whether a polynomial model can improve and regularise abrupt or irregular transitions in regressed hand landmarks.

Method:
Polynomial smoothing applied to outputs from size-and-shape regression (professor’s hand, Dummy Hand).

Results:

- Smoother, more regular animations; abrupt transitions stayed within the data, further distorting it. (Figure 8.5).
- By smoothing the data and spreading it in time, there is potential loss of time-sensitive or sharp motion features.
- Optionality emphasised: module can be toggled based on user need.

Importance:
Validated polynomial regression as a valuable post-processing tool; clarified trade-offs for user control, and realised the importance of the toggled option.

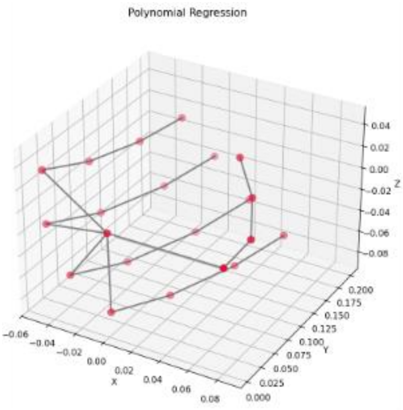


Figure 8.5: dummy hand landmarks smoothed with polynomial regression

8.2.6 Length-Constrained Alignment

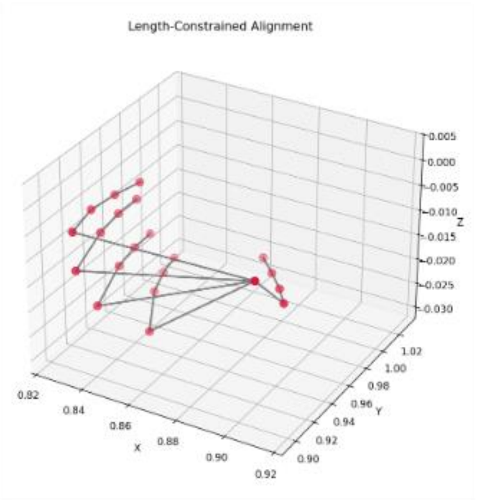
Reason for Test:
To restore physical plausibility to regressed trajectories, enabling meaningful robotic simulation in the future.

Method:
Length-constrained alignment was applied to regressed landmarks, with and without anchoring, and the results were inspected with plots.

Results:

- Without anchoring: overfitting and ‘spinning’ result, but otherwise physically plausible poses.
- With anchoring: well-aligned, plausible trajectories; systematic first-frame misfits and orientation flips largely corrected (Figures 8.6 and 8.7).
- Quantitative errors reduced to ~0.005-0.01 meters (tested with PyBullet simulation) but not fixing it entirely.
- Suspected that bug in initial frames in size-and-shape regression may actually present correct results but flip the entire plot in space.

Importance:
Confirmed anchoring and length constraints are essential for simulation realism; documented residual bugs and limitations.



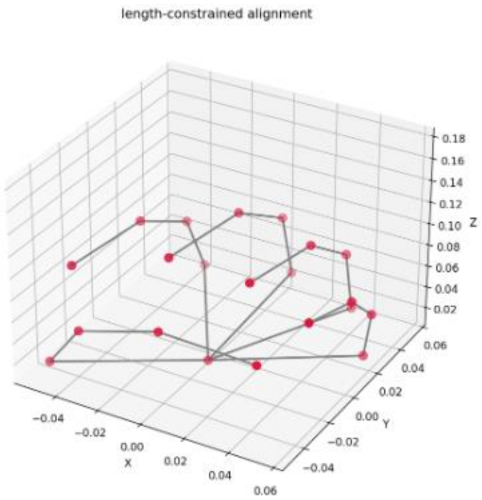


Figure 8.6: Professor’s landmarks after length-constrained alignment
Figure 8.7: Dummy Hand after length-constrained alignment

8.2.7 PyBullet Simulation

Reason for Test:
To verify that pipeline outputs are animatable and physically plausible in a physics-based simulation environment is a milestone for trajectory planning for artificial limbs.

Method:
Dummy Hand configuration loaded into PyBullet; regressed and aligned landmarks used for grasping animation (Professor’s landmarks did not have a robot, and the 3DoF robot was never successfully regressed using size and shape model).

Results:

- PyBullet dynamically adjusted joint angles, correcting minor errors in real time left after length-constrained alignment.
- Mean simulation error: ~0.007 ± 0.002 meters.
- Simulation limited to the Dummy Hand dataset due to a lack of a physical robot for the professor’s landmarks.

Importance:
Demonstrated end-to-end feasibility and robustness of the method and entire pipeline; confirmed simulation constraints and limitations, proved applicability of the method in real-life problems.

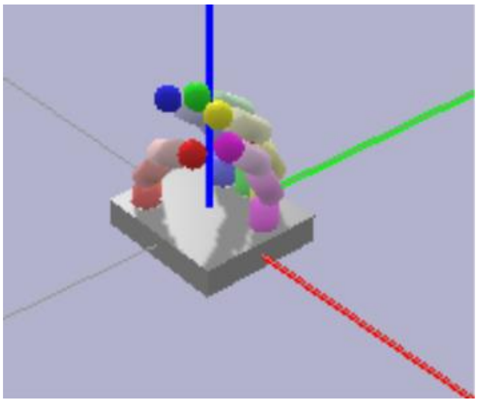
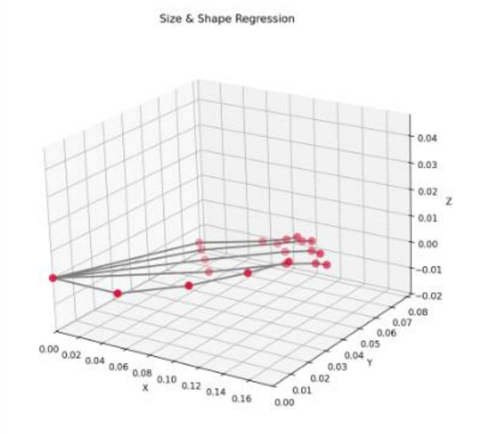


Figure 8.8: PyBullet simulation of Dummy Hand using regressed landmarks

8.3 Bugs, Limitations, and observed Issues

Size-and-Shape Regression Model:

First-frame misfits, orientation flips, and overfitting for sparse datasets were observed. It is suspected that the alignment method (Helmert, Procrustes) may rotate the initial result, but nonetheless provide a valid output.



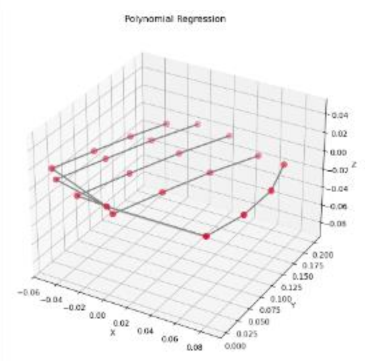


Figure 8.9: Size-and-shape initial misfits of the professor’s landmarks

Figure 8.10: Size-and-shape initial misfits of dummy hand

Length-Constrained Alignment:

Anchoring proved to be essential in the method; without it, overfitting persists within the data.

Polynomial Regression:

Smoothing can obscure sharp transitions and limit time-precise features by spreading input in time; optionality may be toggled on if the user needs. Furthermore, alternating the size-and-shape regression possibly distorts the original regression, especially harmful due to an initial frame bug present in size-and-shape regression, which alternates a solution of the whole polynomial regression. Especially visible in model-like landmarks is a whole, precise transition is present.

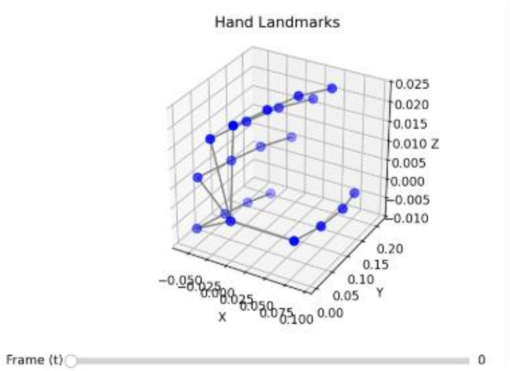


Figure 8.11: Missfits present in model-like landmarks

Simulation:
PyBullet corrects minor errors but requires plausible inputs for accurate simulation; only the Dummy Hand was simulated due to hardware constraints.

8.4 Future Use Cases and Research Extensions

Model-Like Predictions:
Polynomial regression enables continuous, model-like landmark generation for future trajectory planning or precise and predictive simulation. This approach would require improving the length-constrained alignment method and providing a more sophisticated simulation pipeline.

Pipeline Modularity:
Flexible module toggling (e.g., polynomial smoothing, alignment) allows adaption to new datasets or research needs, like simple robots (for example, 3DoF robot).

Precise Trajectory planning:
The ability to choose the point desired to get to by the user in PyBullet simulation, would further prove applicability of the model in real-life scenarios.

8.5 Summary

The pipeline is robust for complex biometrics data, requires post-processing for robotics, and is modular for diverse applications.
Visual and quantitative evaluation at each stage ensures transparency and reproducibility of the artefacts.
Figures and error feedback provide concrete documentation of method effects and limitations.

9. Evaluation and Critique

9.1 Methodological Strengths

Modular and Adaptable Design:
The pipeline’s modular structure enables tailored execution for both biometrics and robotics. Optional modules (polynomial smoothing, length-constrained alignment with manual change) allow the user to choose the most appropriate configuration for their data and applications, increasing flexibility and extensibility.

Comprehensive Validation:
Multiple datasets (benchmark, custom hand, robot) were used for testing. Both visual inspection and quantitative measures (e.g., error metrics, frame-by-frame comparisons) confirmed the accuracy and plausibility of the results.

Physical Plausibility and Simulation Integration:
The length-constrained alignment and PyBullet simulation steps ensure that output is not only statistically plausible but also physically valid, which is essential for robotic deployment and for verifying the utility of the regression models beyond theoretical analysis.

9.2 Module-Specific Evaluation and Critique

9.2.1 Size-and-Shape Regression & Conversion Pipeline

Strengths:
Successfully integrates advanced regression methods from R into Python, handling complex landmark trajectories and enabling cross-language interoperability.

Limitations:

- Direct Python implementation was not achieved due to the time constraints and complexity, limiting modifiability and accessibility for researchers unfamiliar with R.
- Regression often rotates initial landmarks, creating downstream issues for polynomial smoothing and physical plausibility.
- Model struggles with simple/sparse configurations (e.g., 3DoF robot), leading to erratic or implausible outputs.

User Input & Automation:
Requires careful data preparation and validation; the pipeline is flexible but not fully automated in handling edge cases.

9.2.2 Polynomial Regression

Strengths:
Excels at modelling simple landmark trajectories and provides effective smoothing for more

complex data. Enables model-like predictions and trajectory generation. May provide an alternative for size-and-shape regression for simple robot trajectory planning.

Limitations:

- Tends to propagate defects from input data (e.g., first-frame misfit), especially harmful in full-size model-like predictions that try to rationalise the trajectory.
- May overfit or over-smooth landmarks, masking sharp transitions and distorting original data; thus, smoothing is made optional.
- When smoothing is applied, it can shift or evenly spread landmark transitions over time, which may be undesirable for time-sensitive applications.
- Distorting input with present defects (e.g., first-frame misfit) may further harm prediction, creating changes to the data that may not be possible to fix by length-constrained alignment.

9.2.3 Length-Constrained Alignment

Strengths:
Restores physical plausibility, corrects misalignments caused by size-and-shape and polynomial regression, and ensures segment lengths and anchor position match robotic requirements, preparing data for further robotic simulation.

Limitations:

- Computationally intensive for complex datasets due to the optimisation process.
- Requires the user to supply detailed topology and connection lengths; not fully automated.
- Anchoring, while currently optional, was shown to be essential – future versions should either implement an alternative alignment function (to fix alignment error) or make anchoring mandatory for robotics.
- Length constraints interact with Kabsch alignment in ways that may not fully preserve properties, sometimes requiring additional correction in simulation.

9.2.4 PyBullet Simulation Framework

Strengths:
Robust and highly effective for validating physical plausibility and debugging regression outputs.

Limitations:

- Limited interface – user cannot interact with the simulation or specify custom target positions, restricting use for real-world control or feedback.
- Only datasets with robot models can be simulated; benchmark biometrics data lacking robot configurations cannot be physically validated in this way.

9.3 Bugs, Issues and Their Impact

Regression Model Bugs:
Systematic first-frame misfits and orientation flips in output; partially resolved by post-processing with length-constrained alignment.

Polynomial Regression Bugs:
Smoothing propagates defects, especially visible for full model-like predictions. It is likely to amplify errors if present in input.

Length-Constrained Alignment Bugs:
Without proper anchoring, misalignment is likely to happen. Slow computation for large datasets.

Simulation Bugs:
PyBullet corrects minor errors but cannot compensate for major physical implausibility in input data. Regression model and alignment process has been proposed for optimizing the trajectories.

9.4 Alternative Approaches and Future Directions

Direct Python Implementation of Regression:
Reimplementing the size-and-shape regression model in Python would improve modifiability, accessibility, and cross-platform support, allowing more researchers to contribute and adapt the method.

Alternative Alignment Methods:
Exploring other physical alignment or constraint-enforcement techniques, such as learning-based or optimisation-driven approaches, could improve efficiency and reduce user intervention, possibly making anchoring fully optional.

Automated Topology and Length Extraction:
Automating segment length and topology extraction from data would reduce user workload and make the pipeline more approachable for users. Unfortunately, automated topology extraction might prove complex or impossible for complex structures like hand landmarks.

Robust Smoothing and Defect Detection:
Integrating automated defect detection and smarter smoothing (e.g. robust regression, outlier rejection) would prevent the propagation of errors and improve output quality. This approach wasn’t considered due to the fear of future distortion of the size-and-shape regression output.

Enhanced Simulation Interface:
Developing a more interactive PyBullet interface, allowing specification of target positions and real-time feedback, would expand the framework’s applicability to practical robotics problems.

Alternative Regression Models for Simple Data:
For sparse or simple configurations, consider regression methods with better generalisation (e.g., splines, Gaussian processes, or domain-specific models).

9.5 Recommendations

Automate and Simplify User Input:
Reduce reliance on manual topology and segment length specification; integrate automated data analysis and error detection.

Mandate anchoring for Robotics:
Make landmark anchoring a default or required parameter for robotic use cases, given its proven necessity.

Improve Defect Correction:
Address root causes of regression defects (e.g., initial frame bug) before smoothing or alignment.

Expand Simulation Capabilities & Utilise Model-Like Landmarks:
Build interactive simulation tools, enabling real-world feedback and precise control, utilising model-like landmarks produced by the Polynomial model.

The pipeline’s modularity and extensibility make it a valuable tool for both biometrics and robotics research. Its integration of statistical, physical, and simulation methods sets a precedent for future work in trajectory planning and morphometrics analysis. With further automation and enhanced interfaces, the approach could underpin next-generation prosthetics, robotic control systems, and biomechanical modelling.

9.6 Summary

This evaluation provides a balanced view of the pipeline’s strengths, limitations, and areas for improvement. By critically analysing each module, discussing alternative approaches and outlining future directions, it offers a roadmap for continued development and practical application in computational research.

10. Conclusion

10.1 Summary of Work

The development of new methods for optimal trajectory planning is a critical challenge in robotics and data analytics. In this thesis, a modular pipeline was created to address this need, integrating four key components:

1. **Size and Shape Regression** for core landmark analysis, enabling prediction of plausible trajectories based on landmark geometry.
2. **Polynomial Model** for trajectory smoothing and the creation of model-like landmarks, supporting more natural motion and advanced trajectory prediction.
3. **Length-Constrained Alignment**, a novel approach for re-aligning and physically correcting landmark data, ensuring suitability for robotic simulation.
4. **PyBullet Simulation Framework**, use to validate and demonstrate the full trajectory in a physics-based environment, dynamically correcting data for physical plausibility.

This approach was successfully applied to both benchmark (real hand) and custom (artificial dummy hand, 3DoF robot) datasets. The pipeline produced optimal and physically plausible trajectories, with robust validation through numerical study and simulation. In particular, PyBullet simulation confirmed the method’s utility for landmark analysis and trajectory planning in robotic agents.

10.2 Key Findings and Achievements

10.2.1 Findings

Modularity and Flexibility:
The pipeline’s design accommodates diverse datasets and research needs, allowing users to select and combine methods best suited for their application – whether in biometrics or robotics.

Method Selection Matters:
Statistical regression models excel for complex, high-dimensional data, while polynomial regression is preferable for simple landmark configurations or as a smoothing tool.

Physical Plausibility is Critical:
Length-constrained alignment and simulation steps are essential for translating statistical predictions into feasible robotic actions.

Validation and Transparency:
Comprehensive testing, including visual and quantitative evaluation, ensures transparency and supports reproducibility for future research.

10.2.2 Achievements

- Demonstrate the pipeline’s effectiveness for both biometrics and robotics contexts.
- Achieved reliable regression, smoothing, alignment and simulation of hand trajectories.
- Validated robustness with real and artificial data, proving the value of the method for practical trajectory planning.
- Establishing a flexible and extensible framework that future research can build upon.

10.3 Limitations

Despite its strengths, the pipeline faces several limitations:

- Dependency on R for shape regression complicates modification to the regression model and cross-platform compatibility.
- Manual specification of robot topology and segment lengths, which may hinder usability.
- Occasional bugs (first-frame misfit, orientation flips) in regression output that require further refinement.
- Simulation constrained to datasets with robot models, limiting evaluation of pure biometrics data.

10.4 Contributions

Developed an extensible pipeline combining statistical and physical modelling for trajectory planning.
Provided a framework suitable for both academic research and practical robotics applications.
Identified and addressed key challenges in regression, alignment and simulation.
Offered clear documentation, evaluation and recommendation to guide future development.

10.5 Future Work

To build upon this foundation, future research should focus on:

- Migration of Size and Shape Regression to Python: This is a high priority to enable easier modification, reduce technical barriers, and facilitate broader adoption.
- Advancing Trajectory Planning: Implementing methods that consider start/end points, enabling optimal pathfinding, and improving model-like landmark generation to avoid overfitting.
- Refining Length-Constrained Alignment: Enhancing this module to better handle model-like data and automate topology extraction.
- Interactive Simulation Framework: Developing a user interface for the simulation, allowing researchers to input target positions and study optimal trajectory responses.

- Applicability for Real-Life Robotics: With these improvements, the pipeline stands to be directly applicable to artificial limb control and broader robotic motion planning problems

10.6 Closing Thoughts

This dissertation confronted a critical challenge at the intersection of statistical data modelling and robotics. In contrast, advanced statistical methods offer powerful tools for predicting motion, but their theoretical output often fails to respect the physical laws governing a robotic system. The central achievements of this thesis are the design, implementation, and validation of a modular pipeline that successfully bridges this gap between abstract statistical prediction and tangible robotic application.

Using this pipeline, we were able to successfully use captured landmarks of the robot in a size-and-shape regression model – a powerful tool for finding an optimal trajectory in the size-and-shape spectrum following the Gaussian assumption – use them with a Polynomial model to create a more natural trajectory and apply them with length-constrained alignment to restore physical plausibility to the data. Using those results, we were also able to run a physical simulation using PyBullet and prove the real-life application of this method.

The proposed pipeline offers a versatile solution for trajectory planning in both research and industry. It can be applied to the design and control of robotic manipulators, prosthetic limbs, and biomechanical systems where accurate and physically plausible motion is essential. Beyond robotics, the method is also well-suited for morphometric analysis in biological research, medical diagnostics, and any domain requiring robust landmark-based modelling and predictive simulation. Its modular design and adaptability enable researchers and engineers to tailor the approach for diverse tasks, from automated motion planning to real-time simulation and data-driven analysis.

The pipeline developed herein demonstrates how a thoughtful combination of statistical, physical, and computational tools can advance the state of trajectory planning for artificial limbs. By critically evaluating each component and remaining open to future innovations, this work lays the groundwork for the next generation of research in computational morphometrics and robotic motion planning.

11. Bibliography

Coumans, E. and Bai, Y. (2016) PyBullet, a Python module for physics simulation for games, robotics and machine learning. Available at: https://pybullet.org/ (Accessed: 31 August 2025).

Dryden, I.L., Koloydenko, A. and Zhou, D. (2020) ‘Size-and-shape regression for landmark data’, Biometrika, 107(4), pp. 825–842. doi:10.1093/biomet/asaa038.

Dryden, I.L. and Mardia, K.V. (2016) Statistical Shape Analysis: With Applications in R, 2nd edition. Chichester: Wiley.

Hunter, J.D. (2007) ‘Matplotlib: A 2D graphics environment’, Computing in Science & Engineering, 9(3), pp. 90–95. doi:10.1109/MCSE.2007.55.

McKinney, W. (2010) ‘Data structures for statistical computing in python’, Proceedings of the 9th Python in Science Conference, pp. 51–56. Available at: https://pandas.pydata.org/.

Oliphant, T.E. (2006) ‘A guide to NumPy’. USA: Trelgol Publishing.

Paine, C.E.T., Marthews, T.R., Purves, D., Rees, M., Smith, S.W. and Lewis, S.L. (2012) ‘How to fit nonlinear plant growth models and calculate growth rates: An update for ecologists’, Methods in Ecology and Evolution, 3(2), pp. 245–256. doi:10.1111/j.2041-210X.2011.00155.x.

Rpy2 Contributors (2025) Rpy2 Documentation. Available at: https://rpy2.github.io/doc/latest/html/index.html (Accessed: 31 August 2025).

Scikit-learn Developers (2025) Scikit-learn: Machine Learning in Python. Available at: https://scikit-learn.org/stable/documentation.html (Accessed: 31 August 2025).

NumPy Developers (2025) NumPy Documentation. Available at: https://numpy.org/doc/ (Accessed: 31 August 2025).

Matplotlib Developers (2025) Matplotlib Documentation. Available at: https://matplotlib.org/stable/users/index.html (Accessed: 31 August 2025).

Professor’s Hand Landmark Dataset (authors of ‘Size-and-shape regression for landmark data’).

Dummy Hand and 3DoF Robot Landmark Datasets (Created and validated by dissertation author).

Siciliano, B. and Khatib, O. (2016) Springer Handbook of Robotics, 2nd edition. Cham: Springer.