Analysing data for the Bee Project: overview

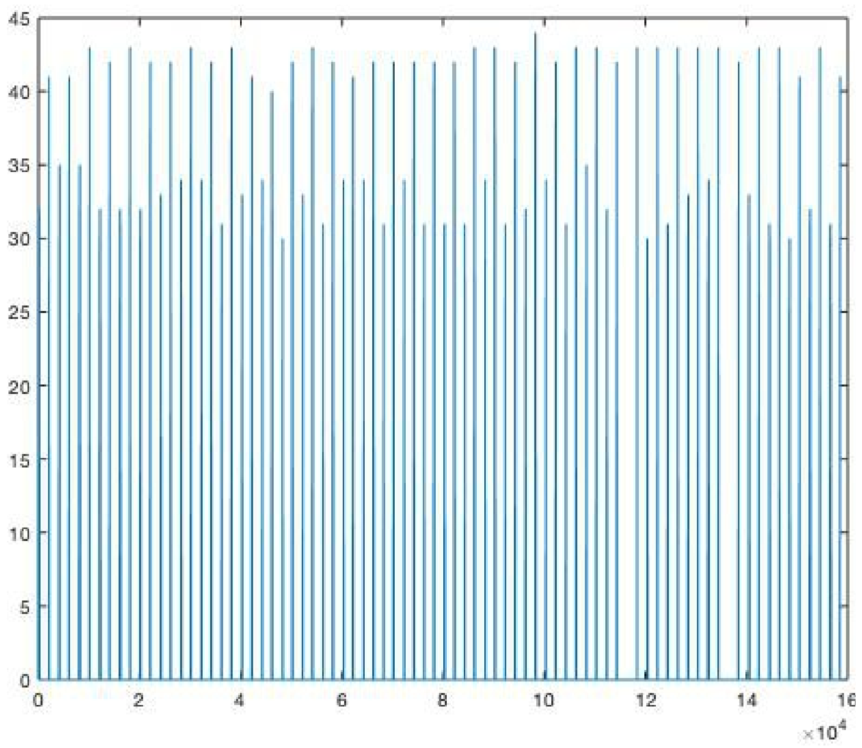
- What we've done previously: processing, summarising, visualising data
- Modelling: lookup model and Gaussian model
- Cross-modelling between experiments
- Machine learning

Processing experiment data

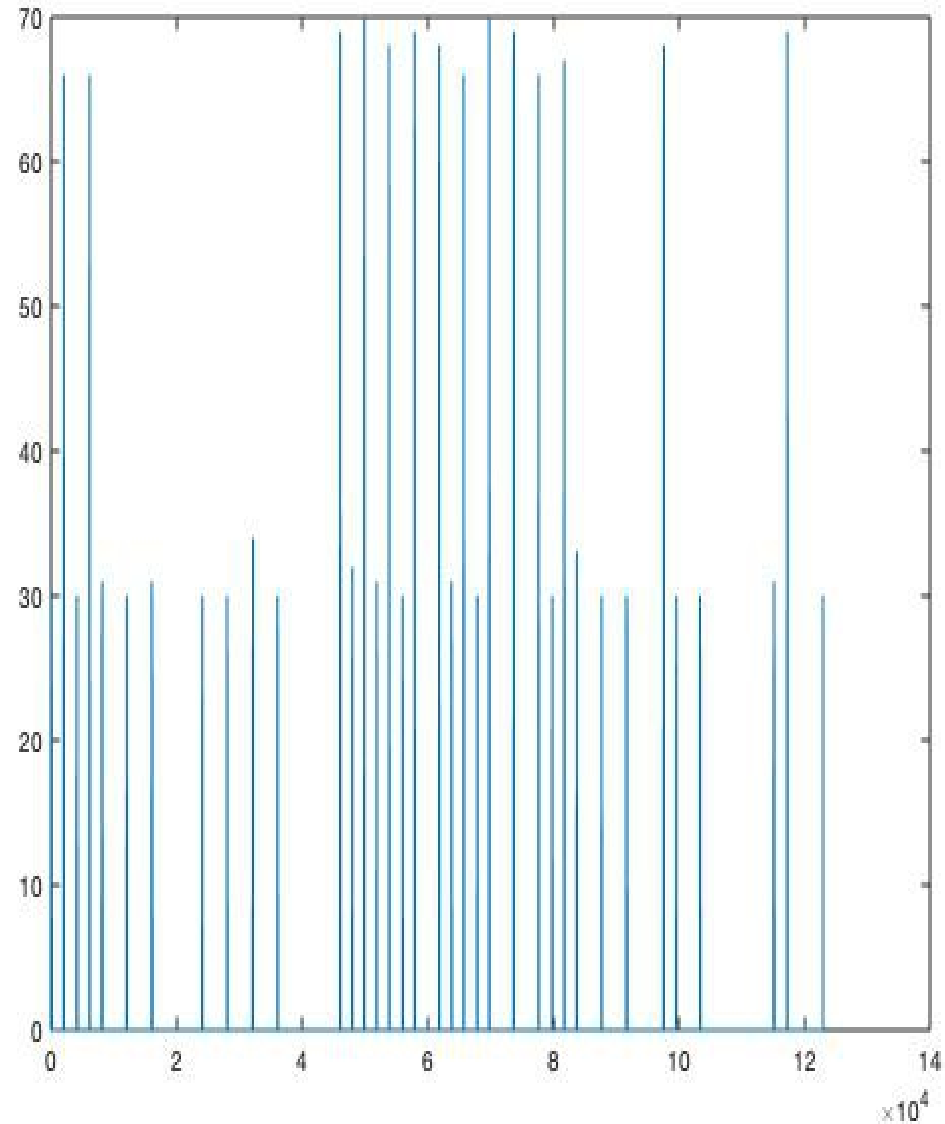
- Removing pulse envelop offsets
- Concatenating pulse data from different files for one sample
- Finding and retaining peaks
- Averaging across a sample

Processing experiment data – rotating experiments

- Identifying interferences
- Retaining 'bee' signal



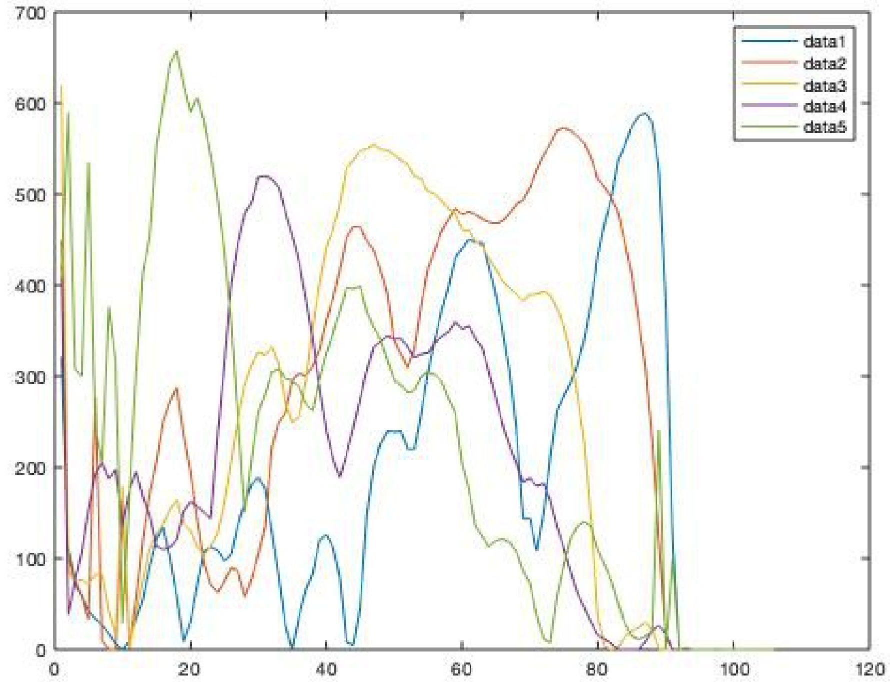
Exp 25



Exp 26

Summarising data

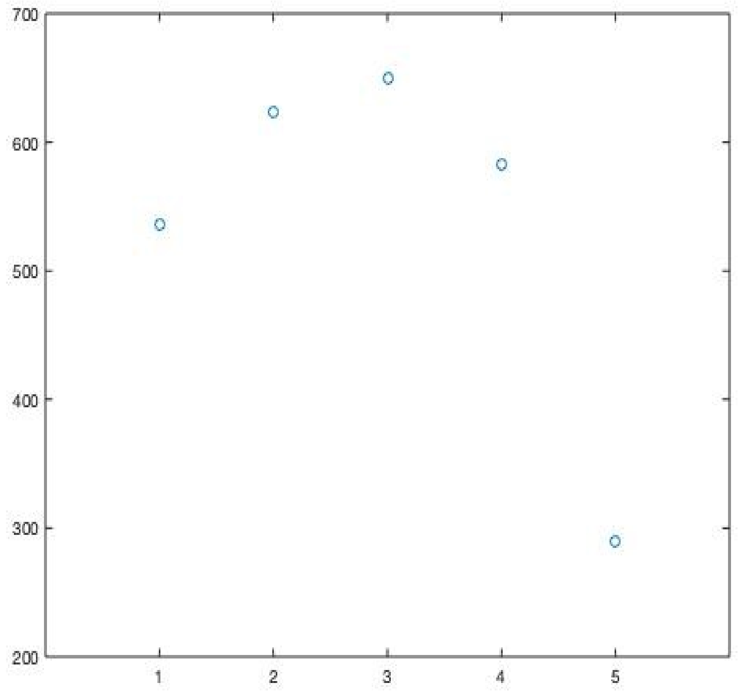
- integral



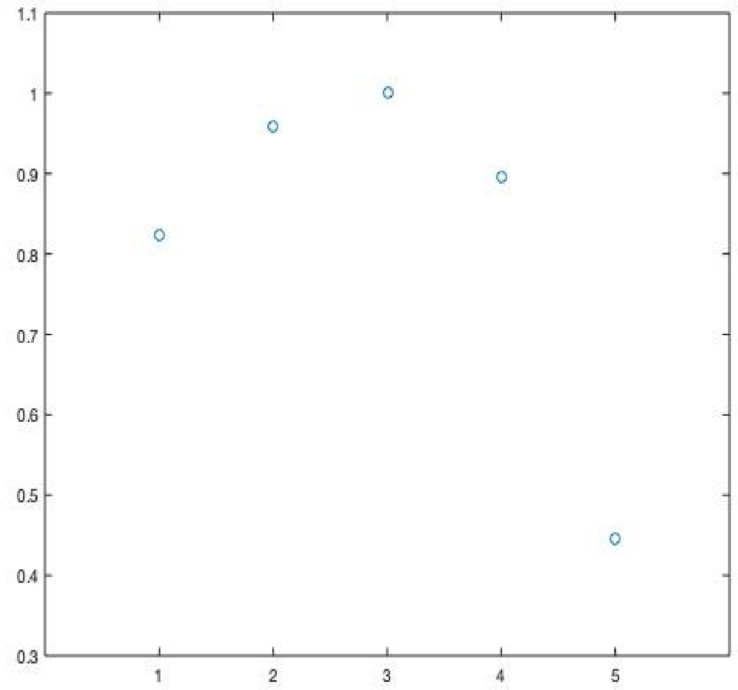
Normalising

Divide each quantuple by its maximum's absolute value

It keeps the ratios between values, normalising all values to the range [0,1]

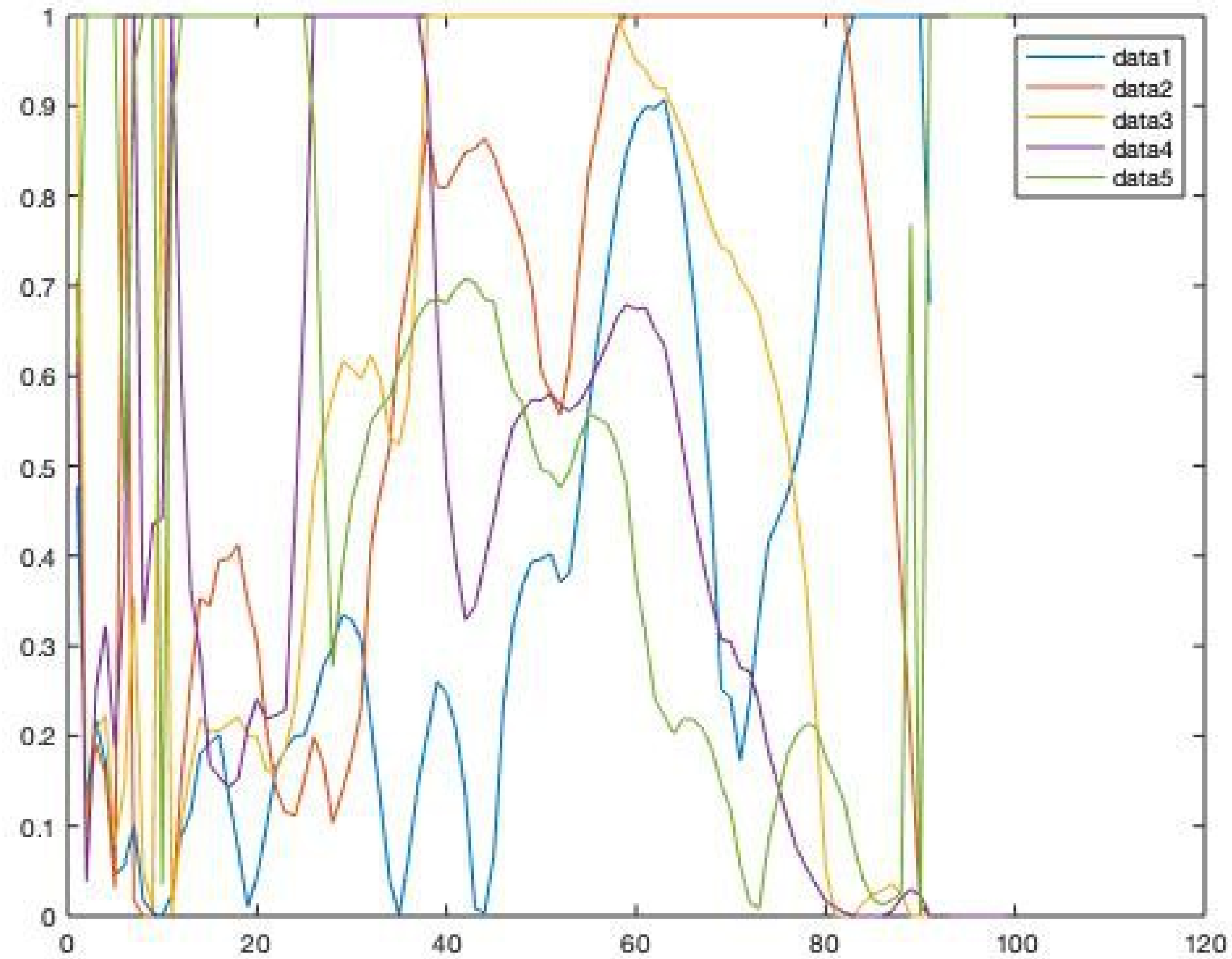


original

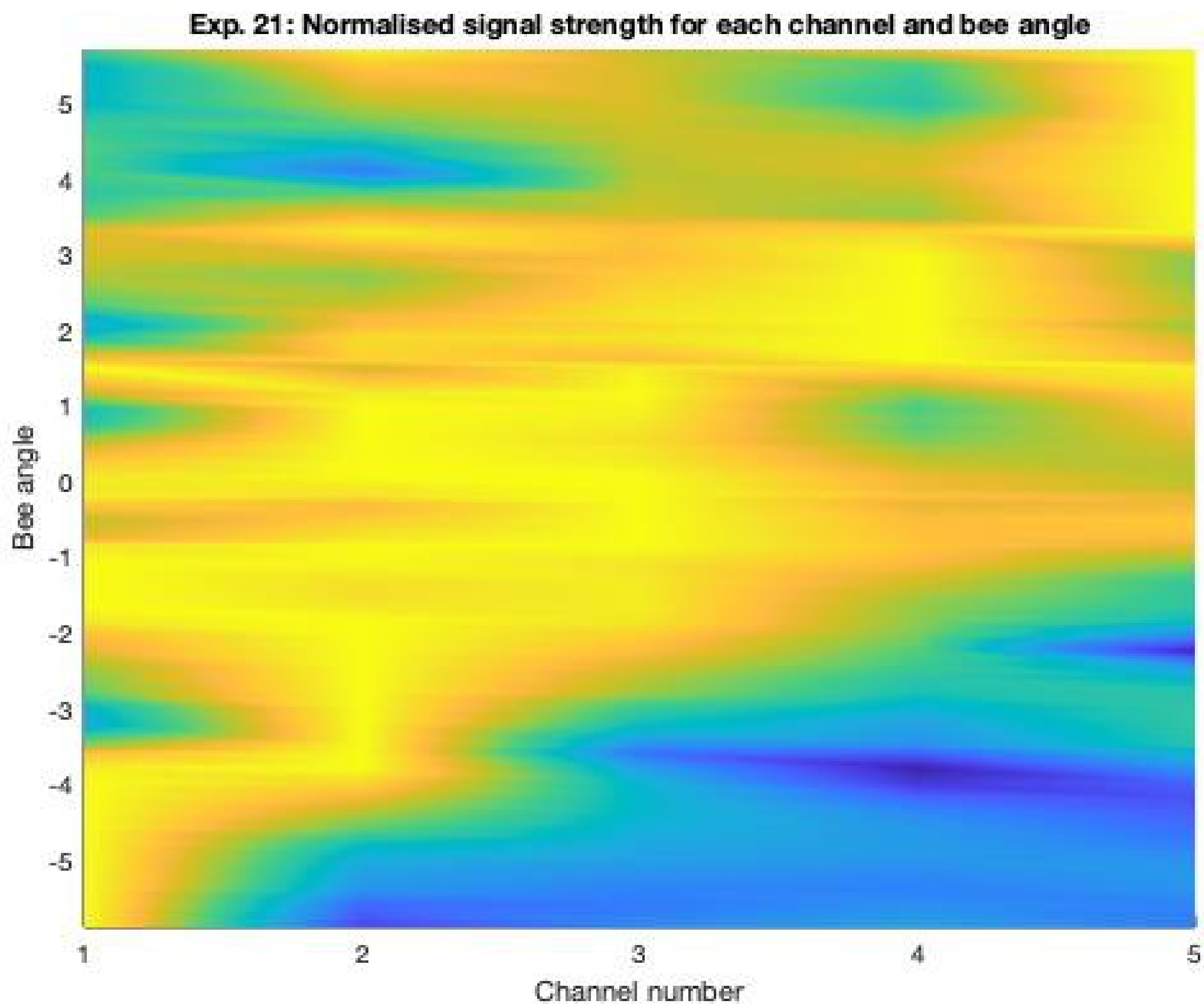


normalised

Normalised data



Normalised data



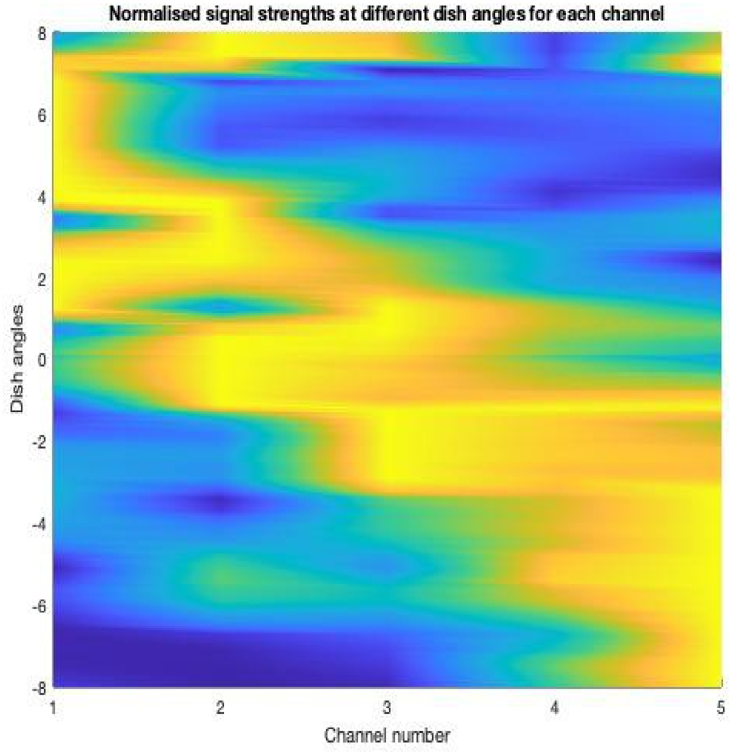
Modelling

Use existing data to predict elevation for new measurements

Measuring and predicting bee angle instead of elevation is independent of the distance

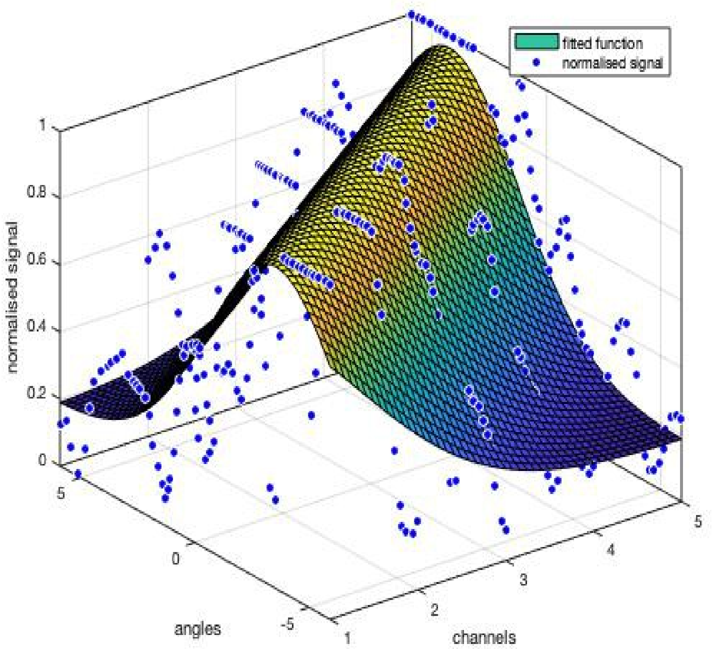
Lookup model

- Lookup model is a table of all quintuples seen in the experiment; modelling a new quintuple involves finding the one with the least square difference

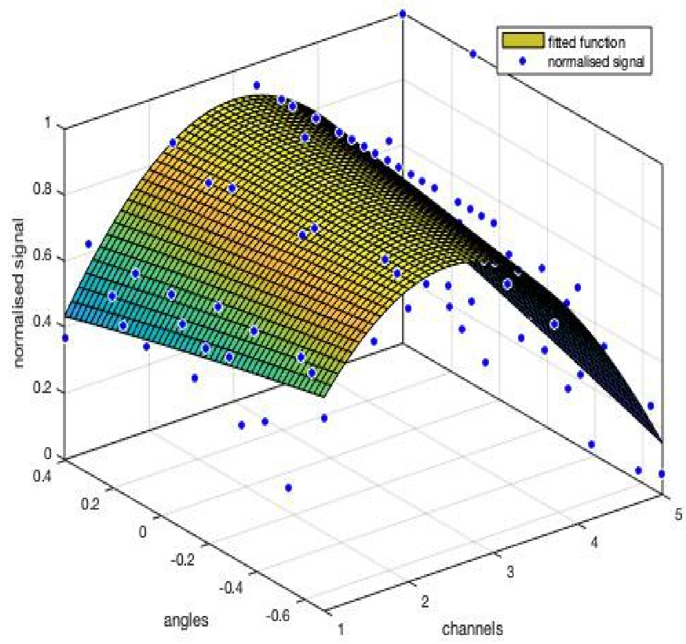


Gaussian model

Gaussian model is a Gaussian function with constraints that best fits the experiment data; modelling a new quintuple involves fitting it to the Gaussian surface

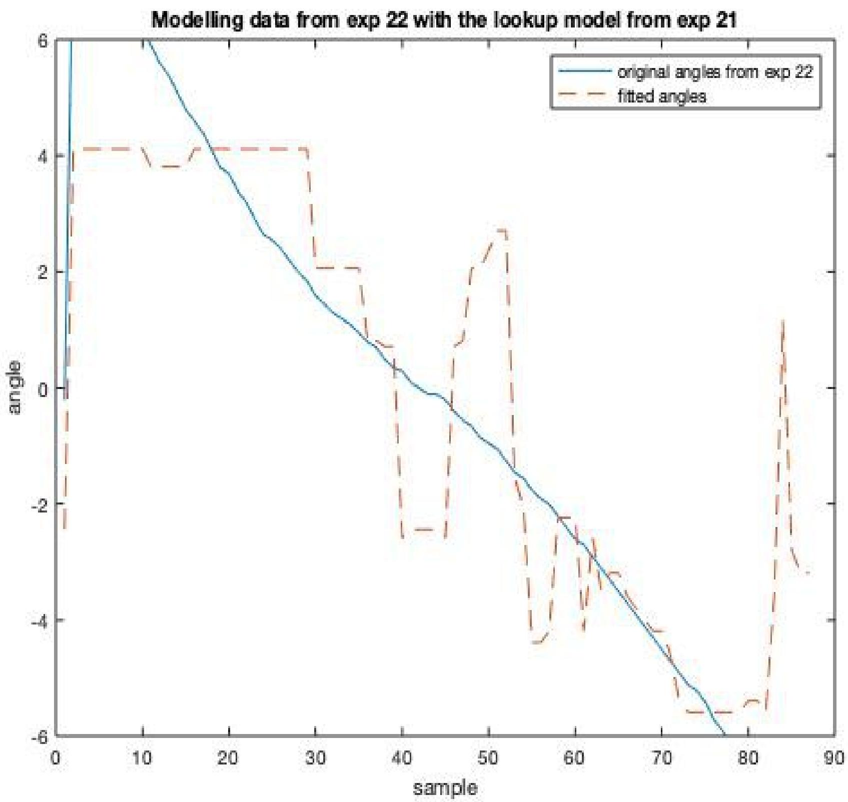


Exp 22 static

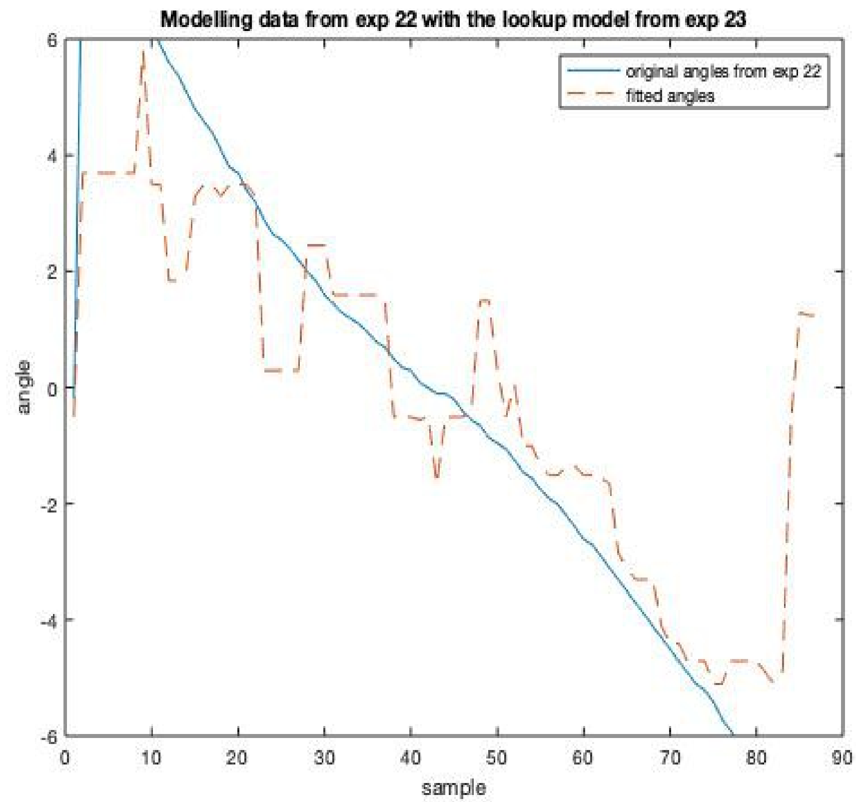


Exp 26 rotating

Cross-modelling static experiments
Lookup model
modelling exp 22

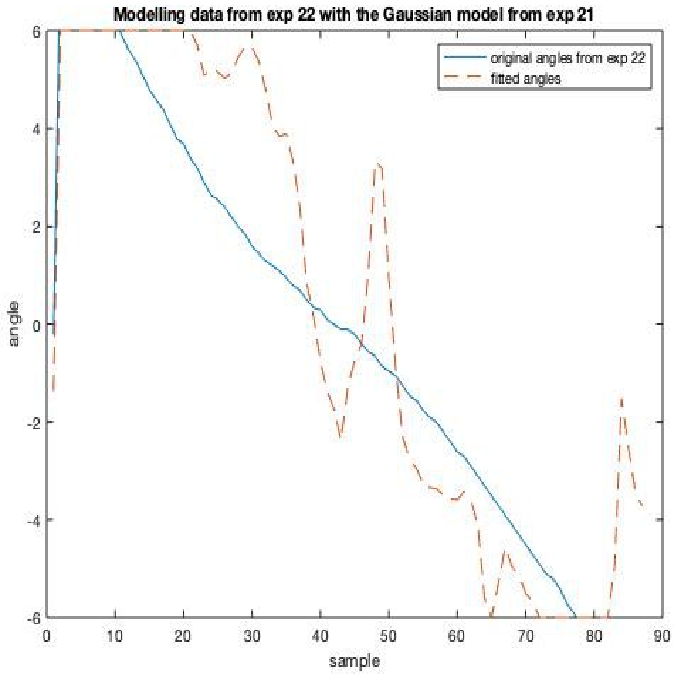


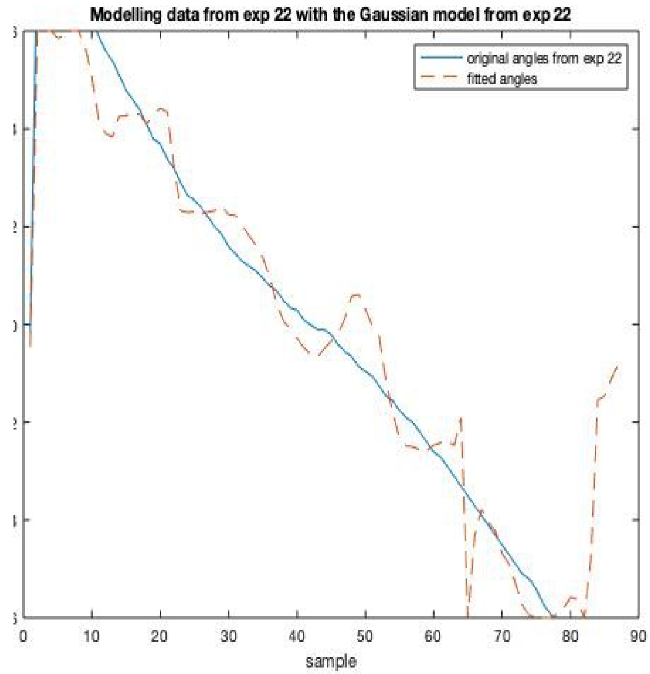
Model from exp. 21

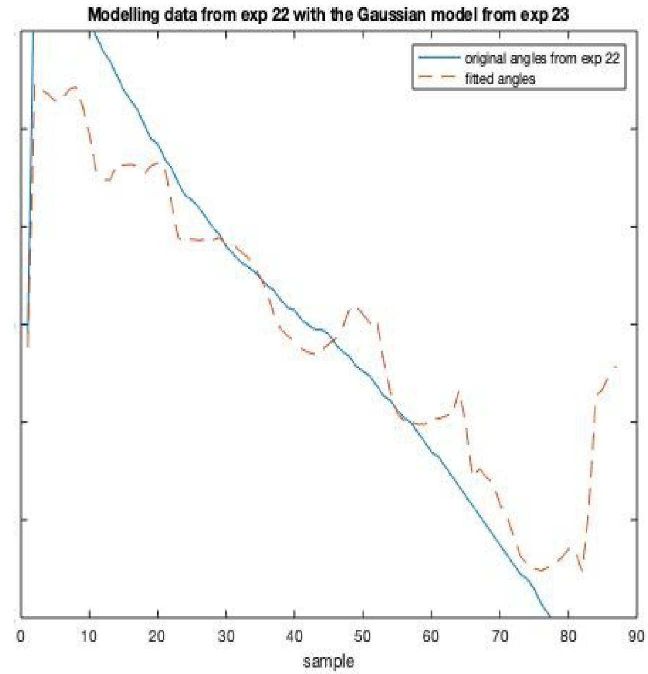


Model from exp. 23

Cross-modelling static experiments Gaussian model modelling exp 22







Model from exp. 21

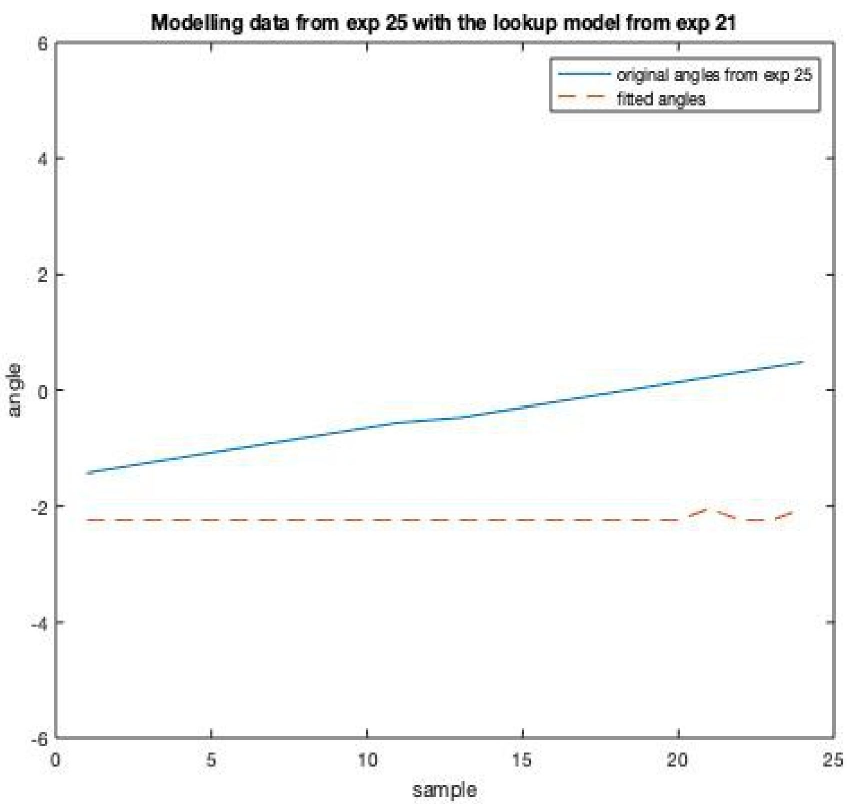
Model from exp. 22
- modelling itself

Model from exp. 23

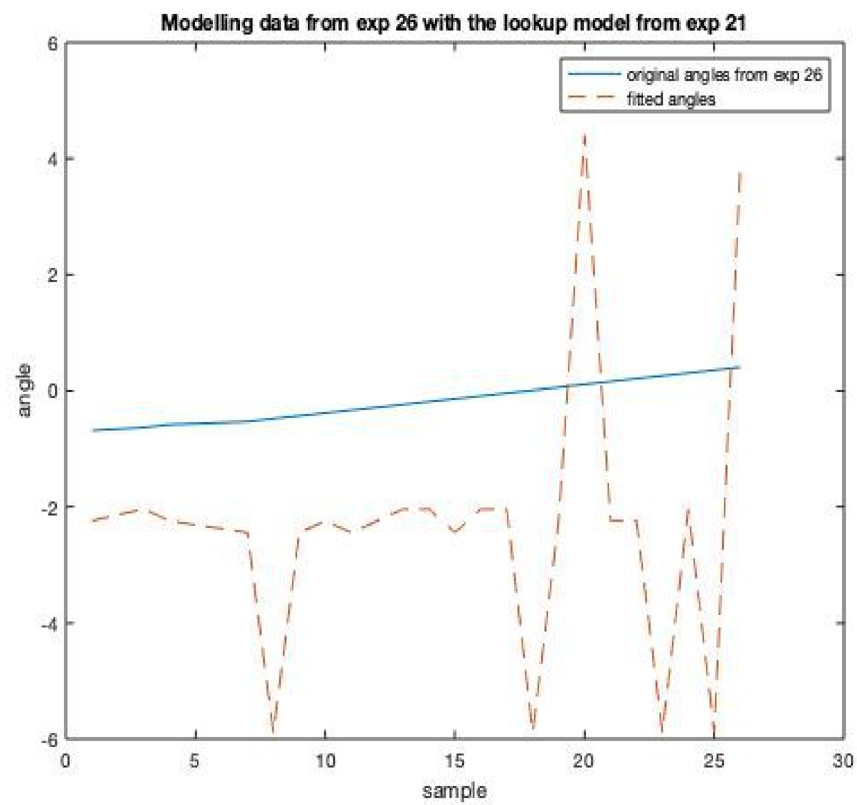
Takeaways from cross-modelling static experiments

- General shape of the modelled data is correct
- It seems the data is more consistent between exp 22 and exp 23, exp 21 differs more
- Lots of errors around the edges of the models
- Where lookup models fail, the errors are quite large, whereas Gaussian models produce smaller errors

Modelling rotating experiments through static data Lookup model from exp 21

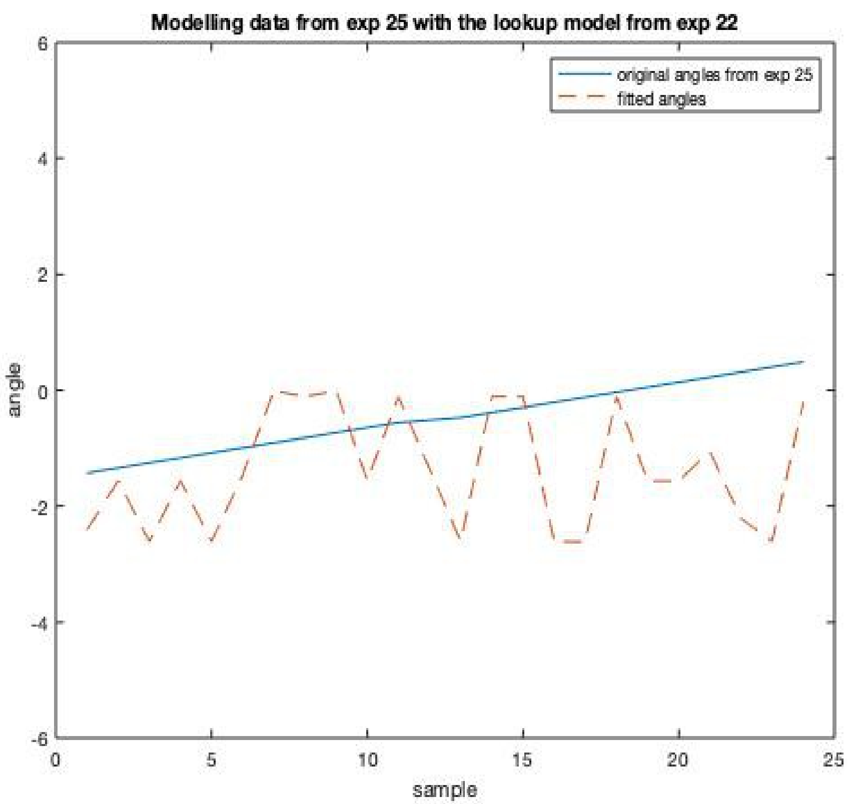


exp. 25

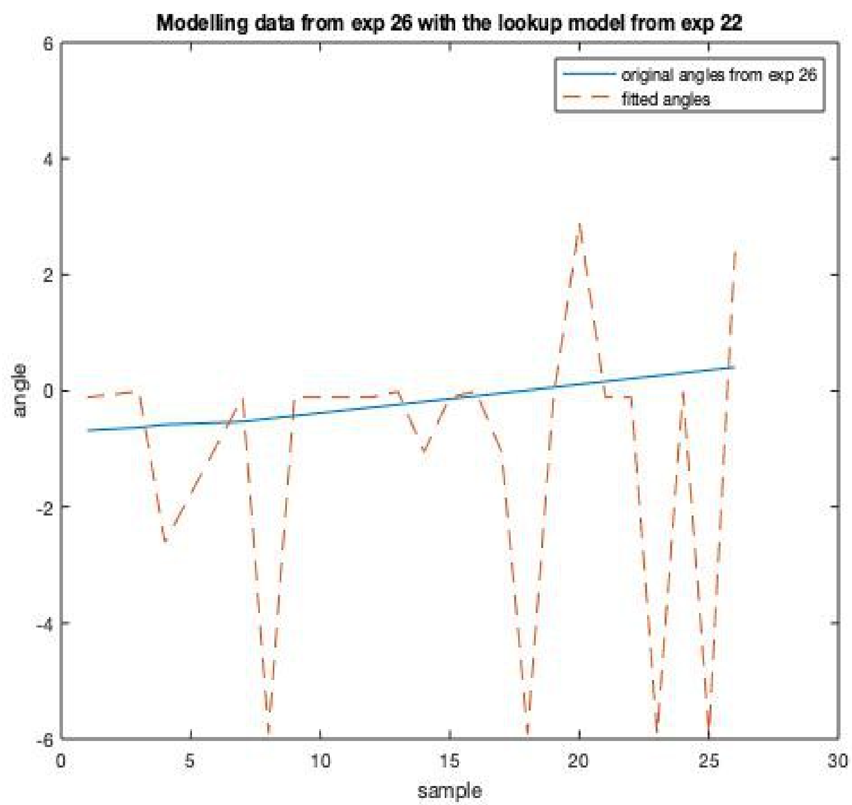


exp. 26

Modelling rotating experiments through static data
Lookup model from exp 22

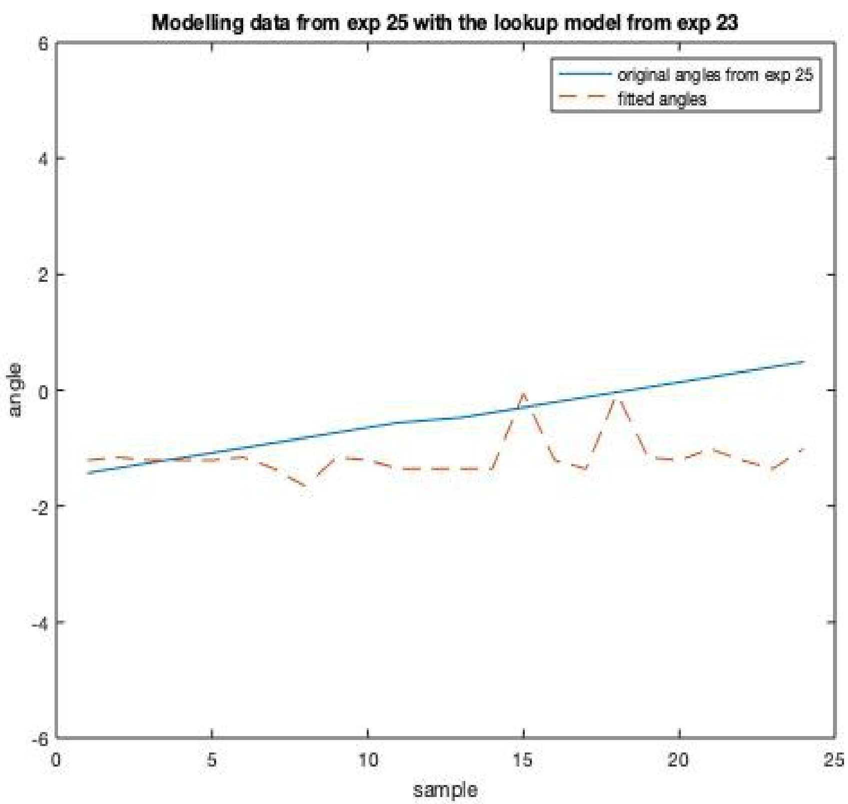


exp. 25

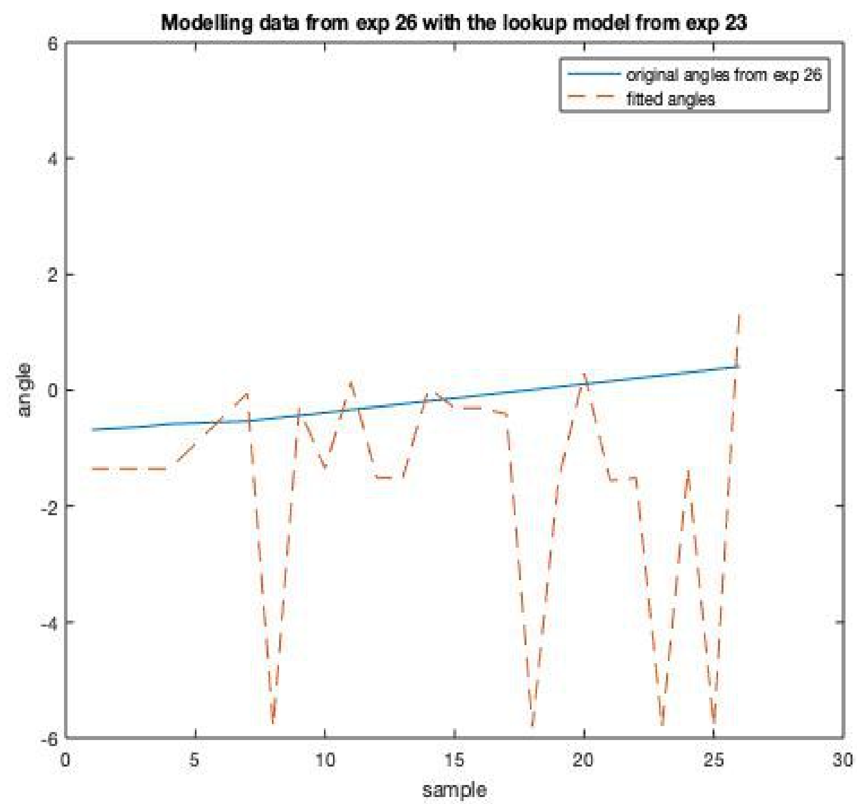


exp. 26

Modelling rotating experiments through static data
Lookup model from exp 23

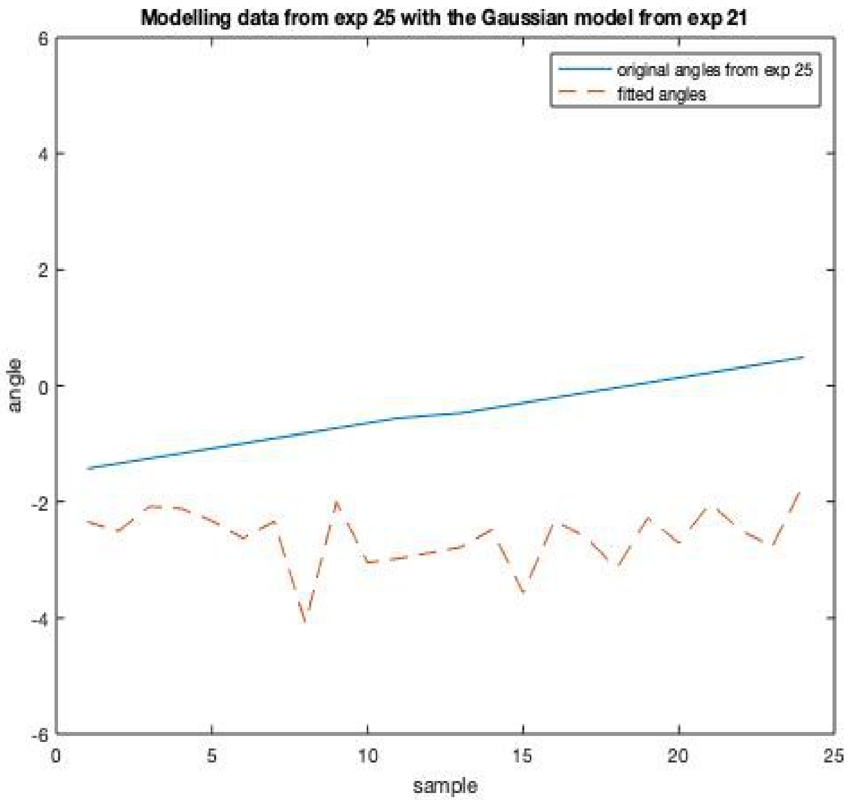


exp. 25

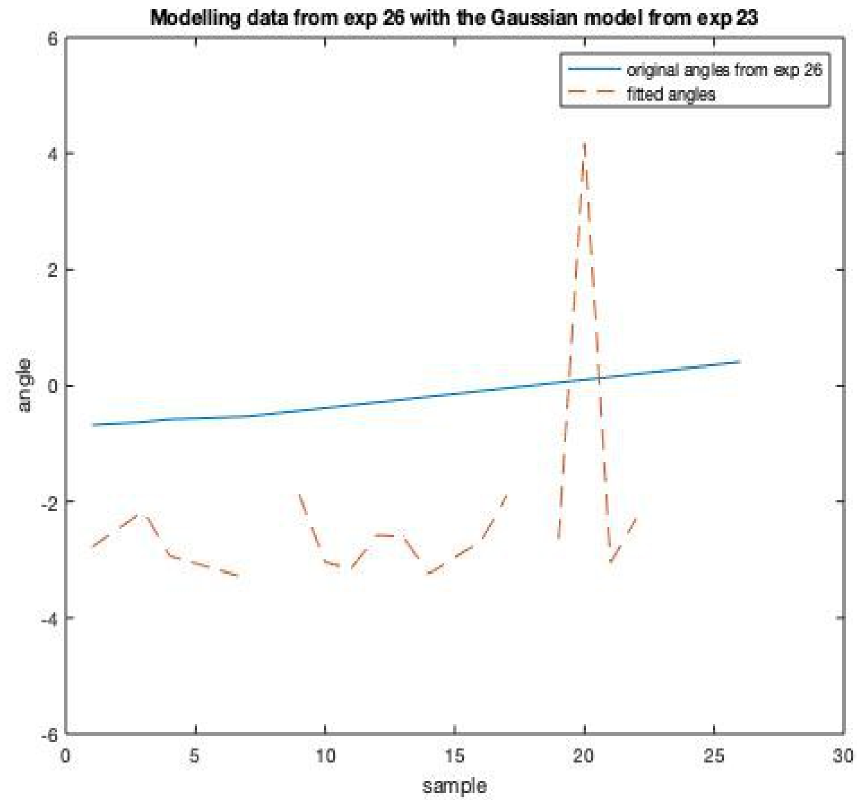


exp. 26

Modelling rotating experiments through static data Gaussian model from exp 21

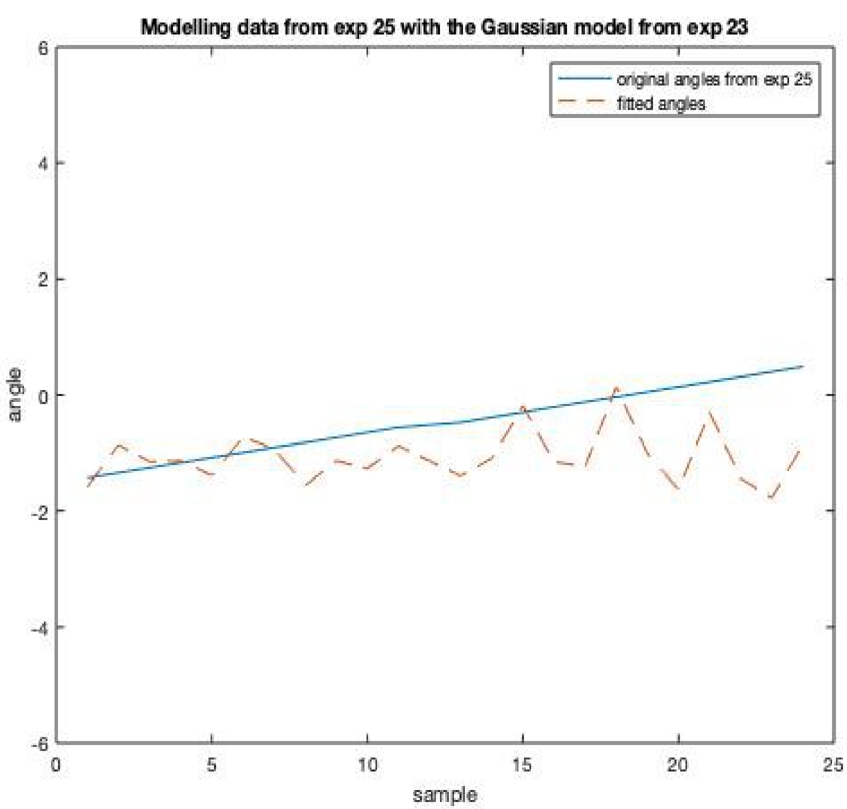


exp. 25

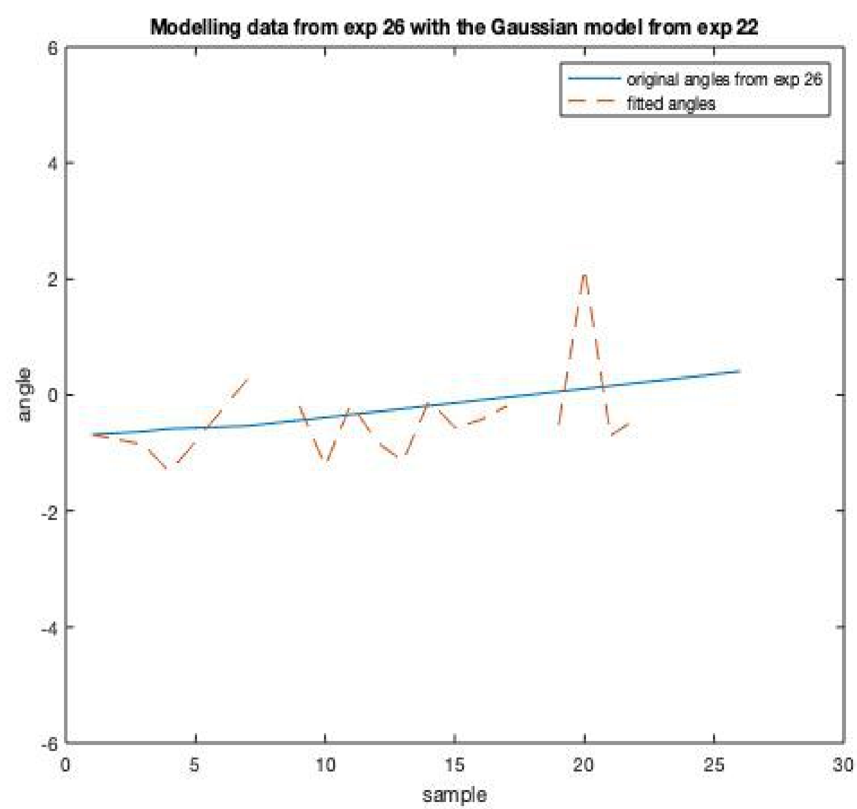


exp. 26

Modelling rotating experiments through static data Gaussian model from exp 22

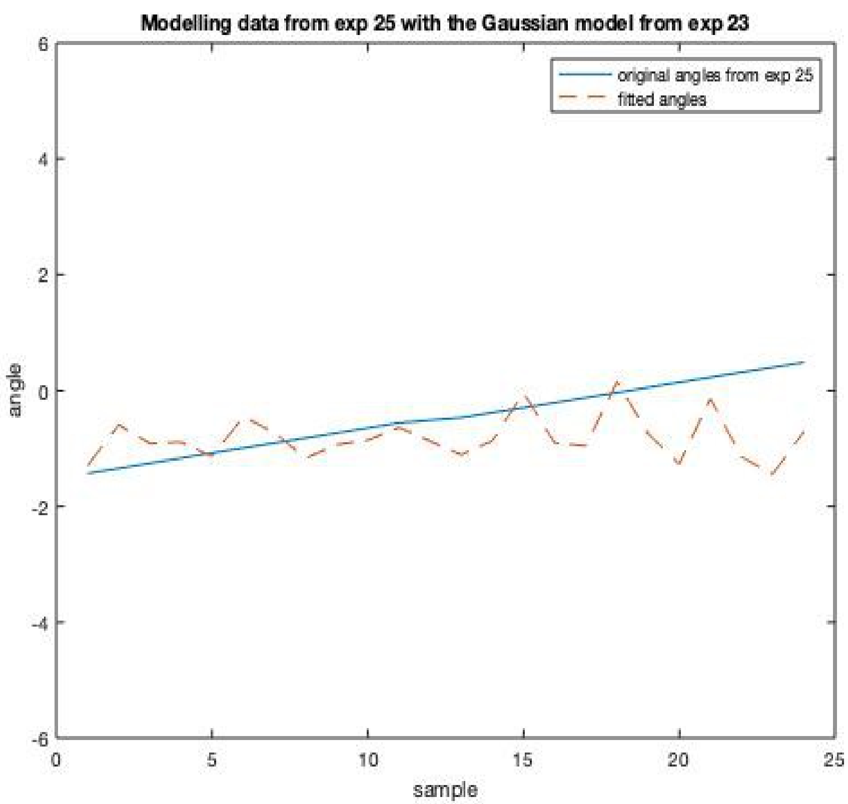


exp. 25

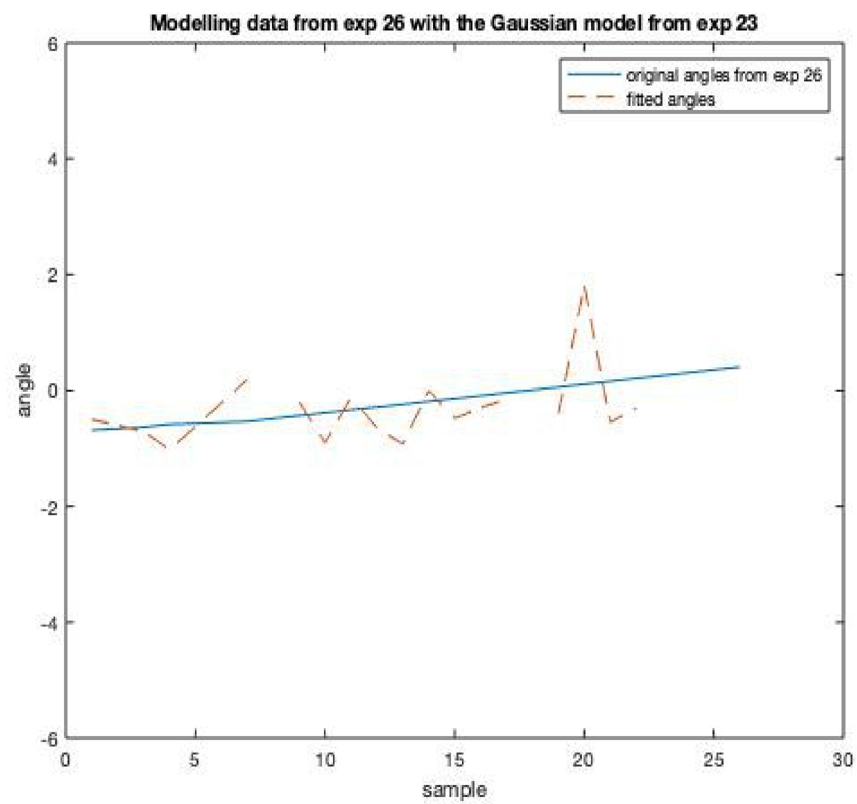


exp. 26

Modelling rotating experiments through static data Gaussian model from exp 23



exp. 25

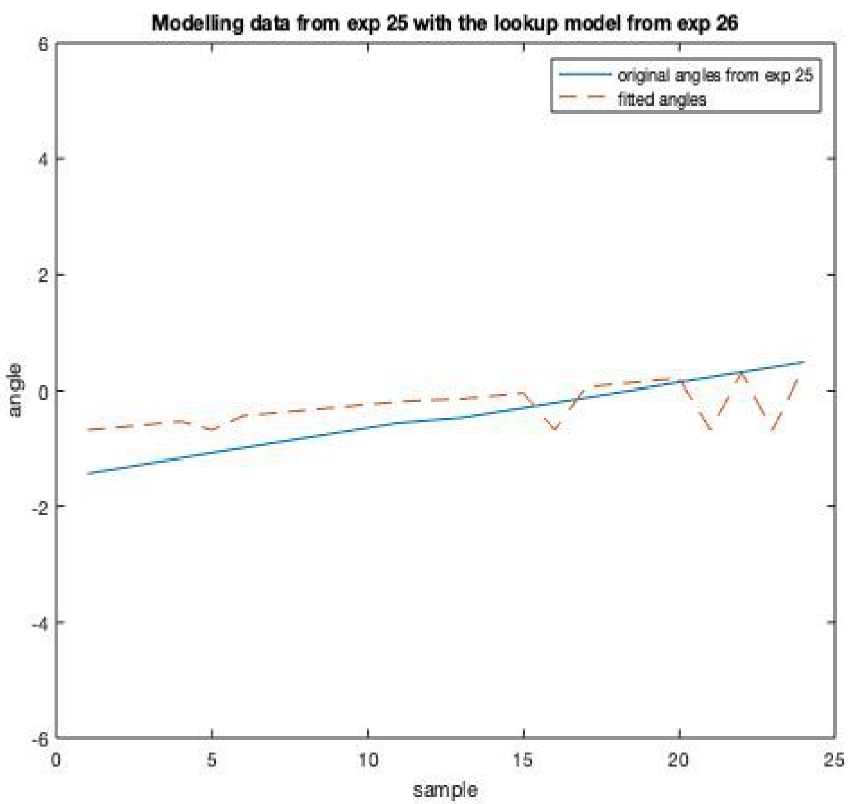


exp. 26

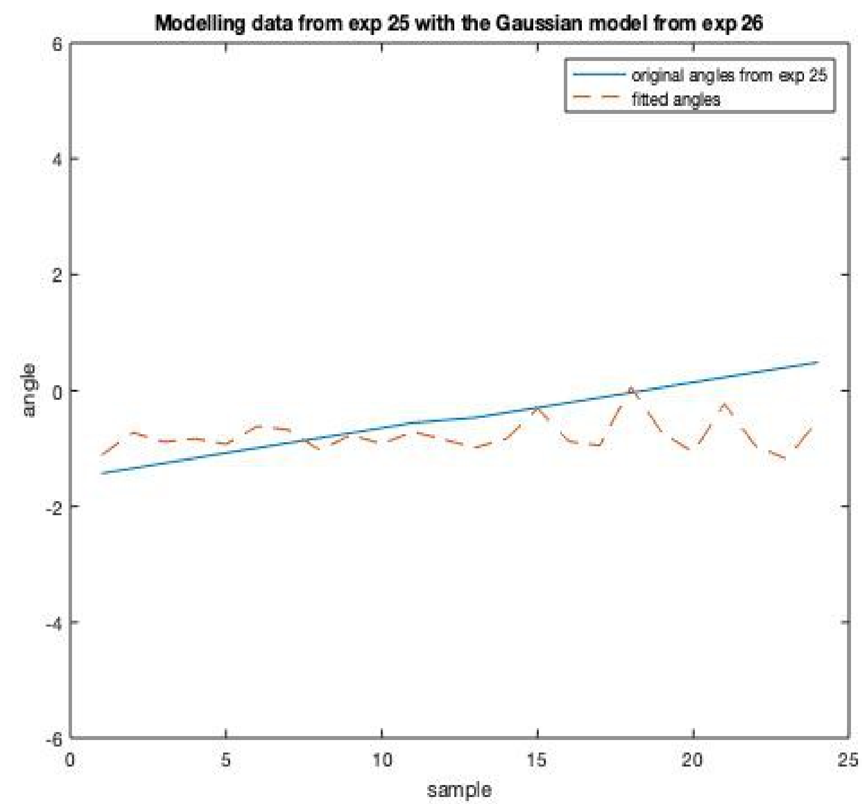
Takeaways from modelling rotating experiments through static data

- Exp. 25 and 26 only cover a narrow range of angles
- While the test data is an ascending line, the overall shape of the predicted data is mostly horizontal: the change in angle is so small that the models do not really notice it
- Exp 25 seems to be easier to model than exp 26, the quality of the data is higher

Cross-modelling rotating experiments modelling exp 25 with the models from exp 26

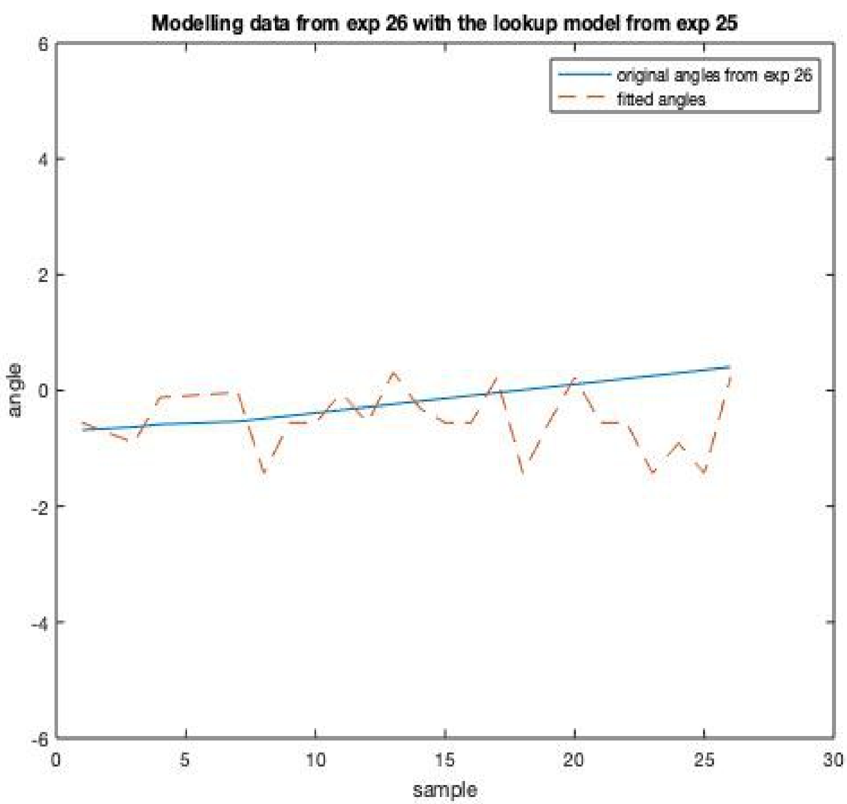


Lookup model

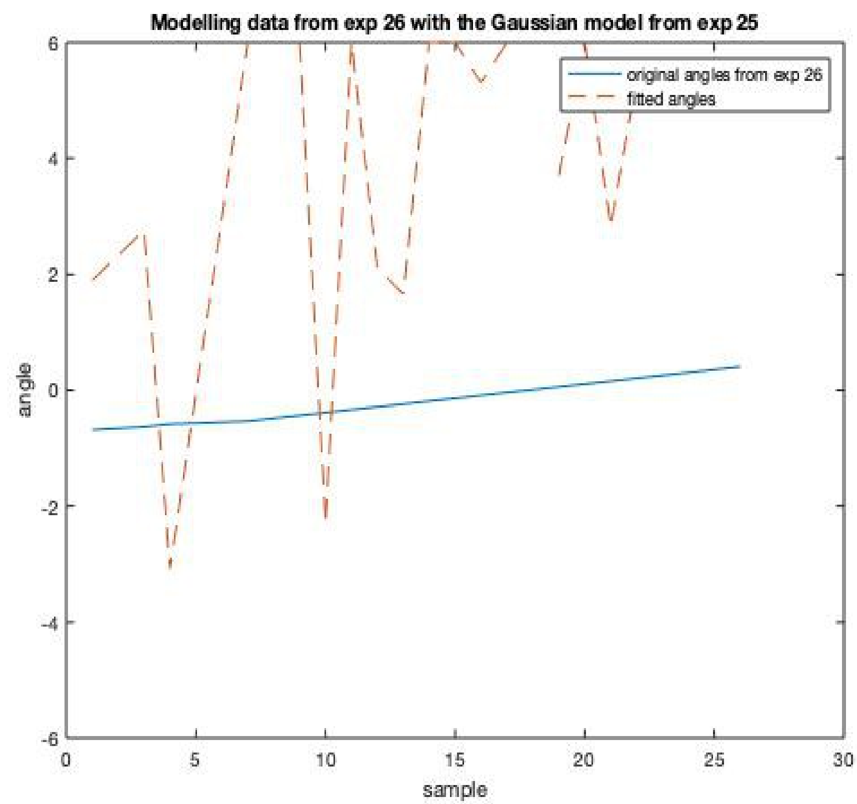


Gaussian model

Cross-modelling rotating experiments modelling exp 26 with the models from exp 25

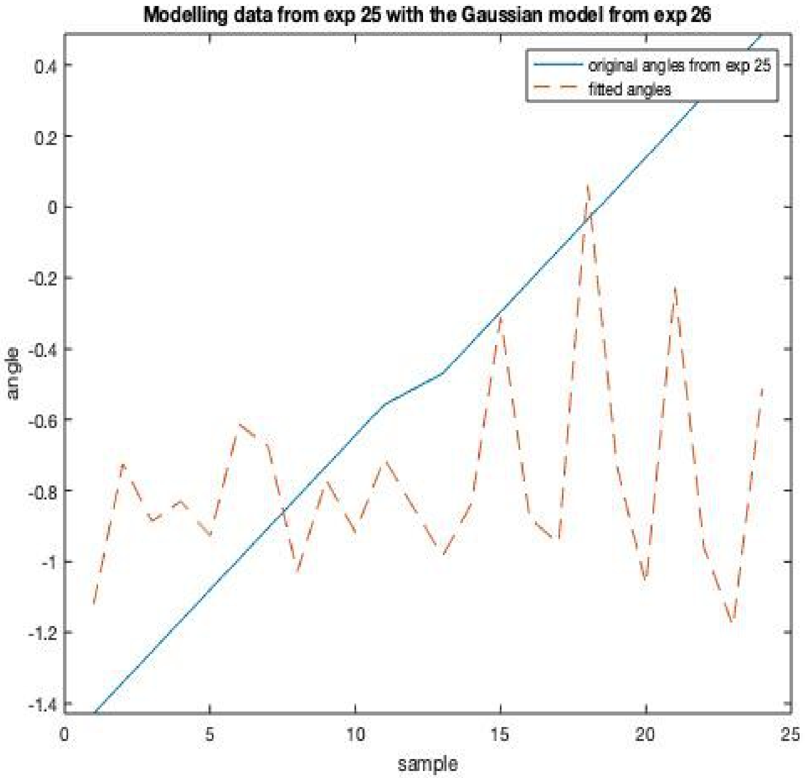


Lookup model



Gaussian model

Prediction error resolution



degree	1º	2º	3º
sin	0.017	0.035	0.052
distance	**elevation error**		
100	1.75	3.49	5.23
200	3.49	6.98	10.47
300	5.24	10.47	15.70
400	6.98	13.96	20.93

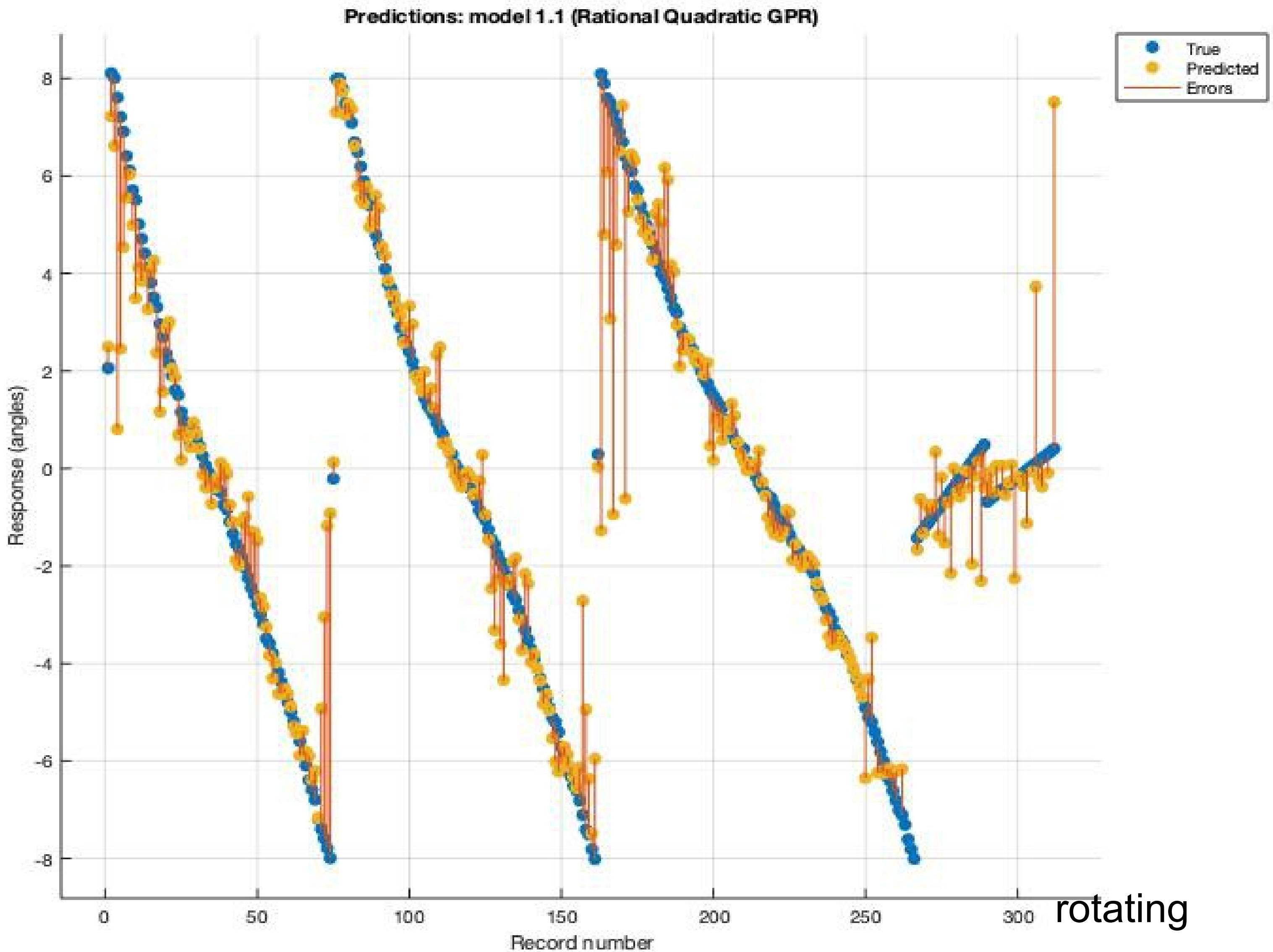
Machine learning

- Generalisation of cross-modelling
- Regression – continuous output
- Tried out various model types – Gaussian processes are best suitable for the task
- Model parameter optimisation vs overfitting

Machine learning

All angles

RMSE = 1.52

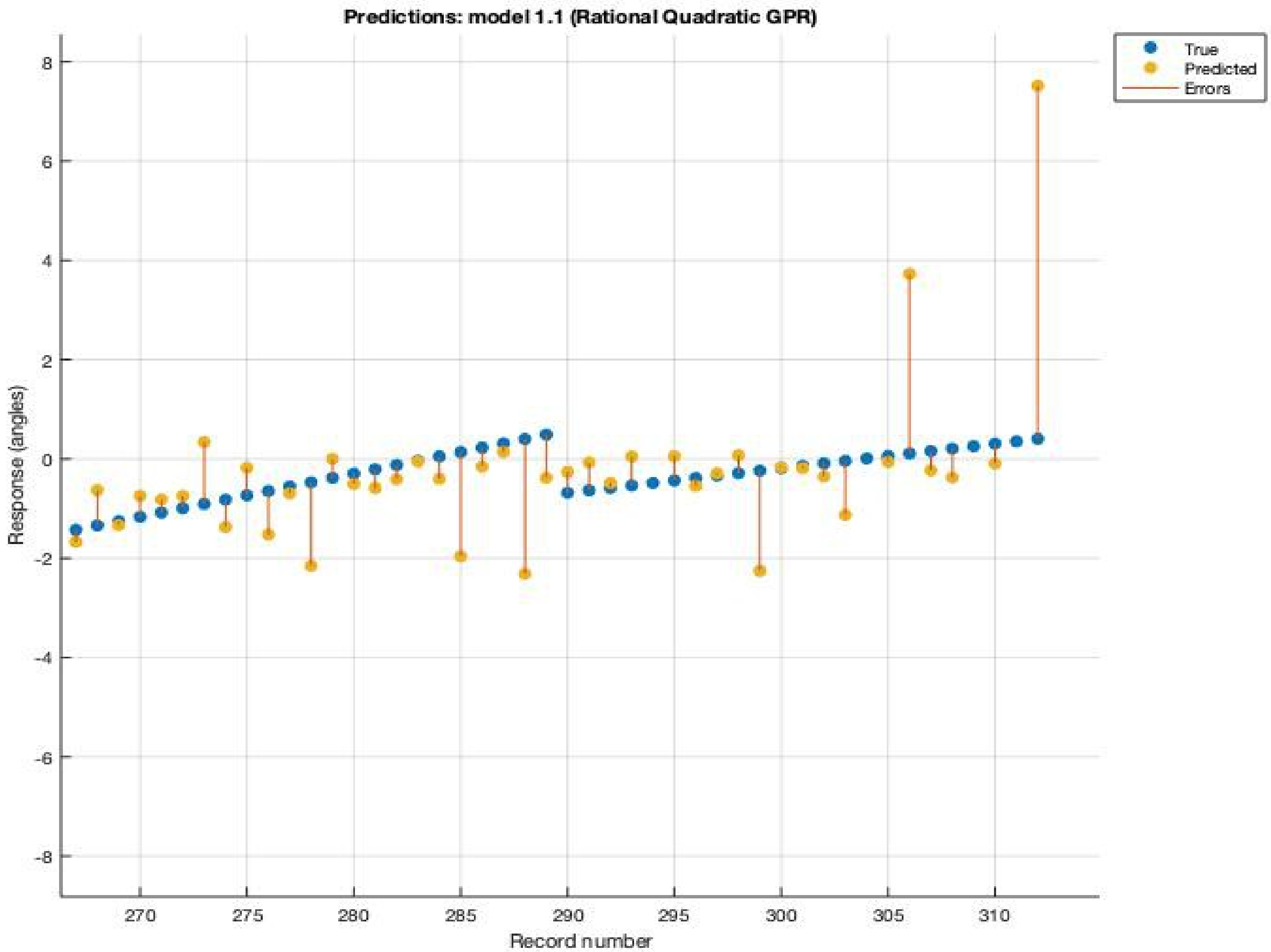


rotating

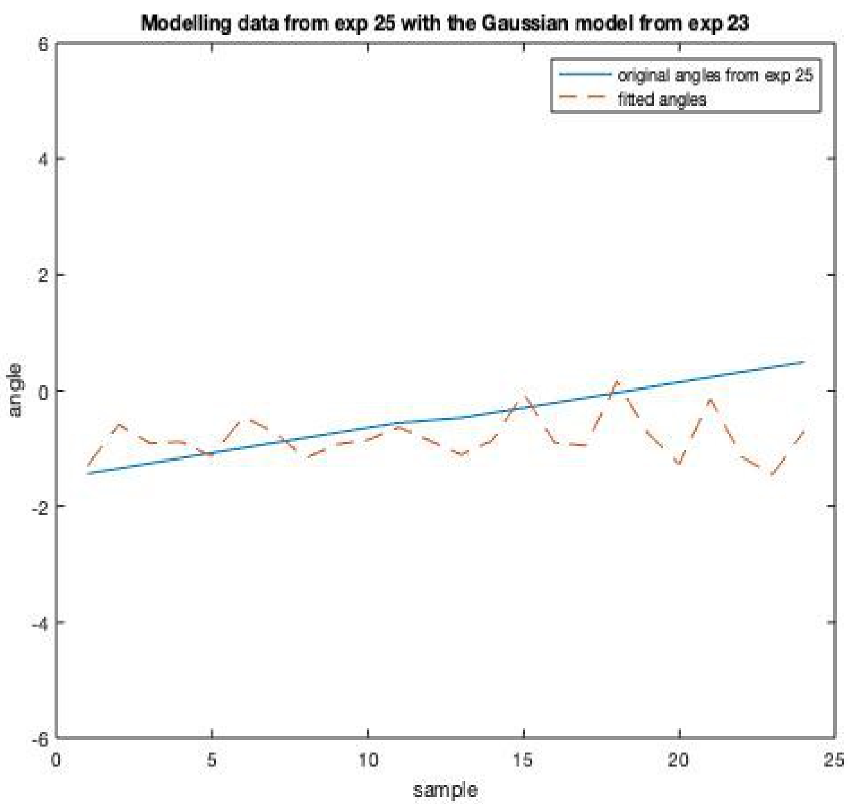
Machine learning

All angles

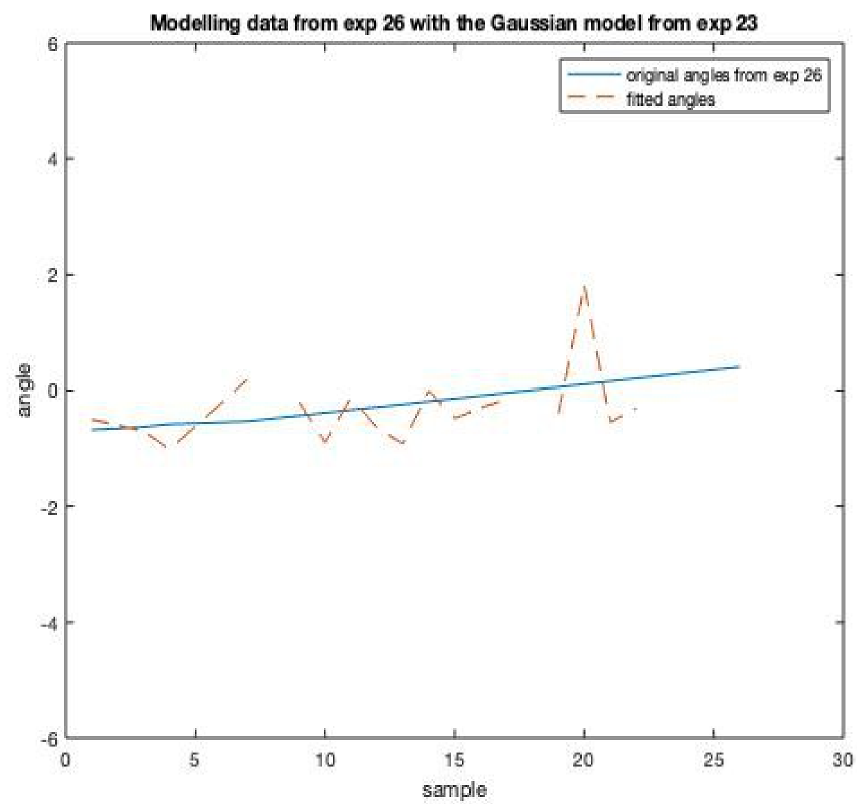
RMSE = 1.52



Modelling rotating experiments through static data Gaussian model from exp 23



exp. 25

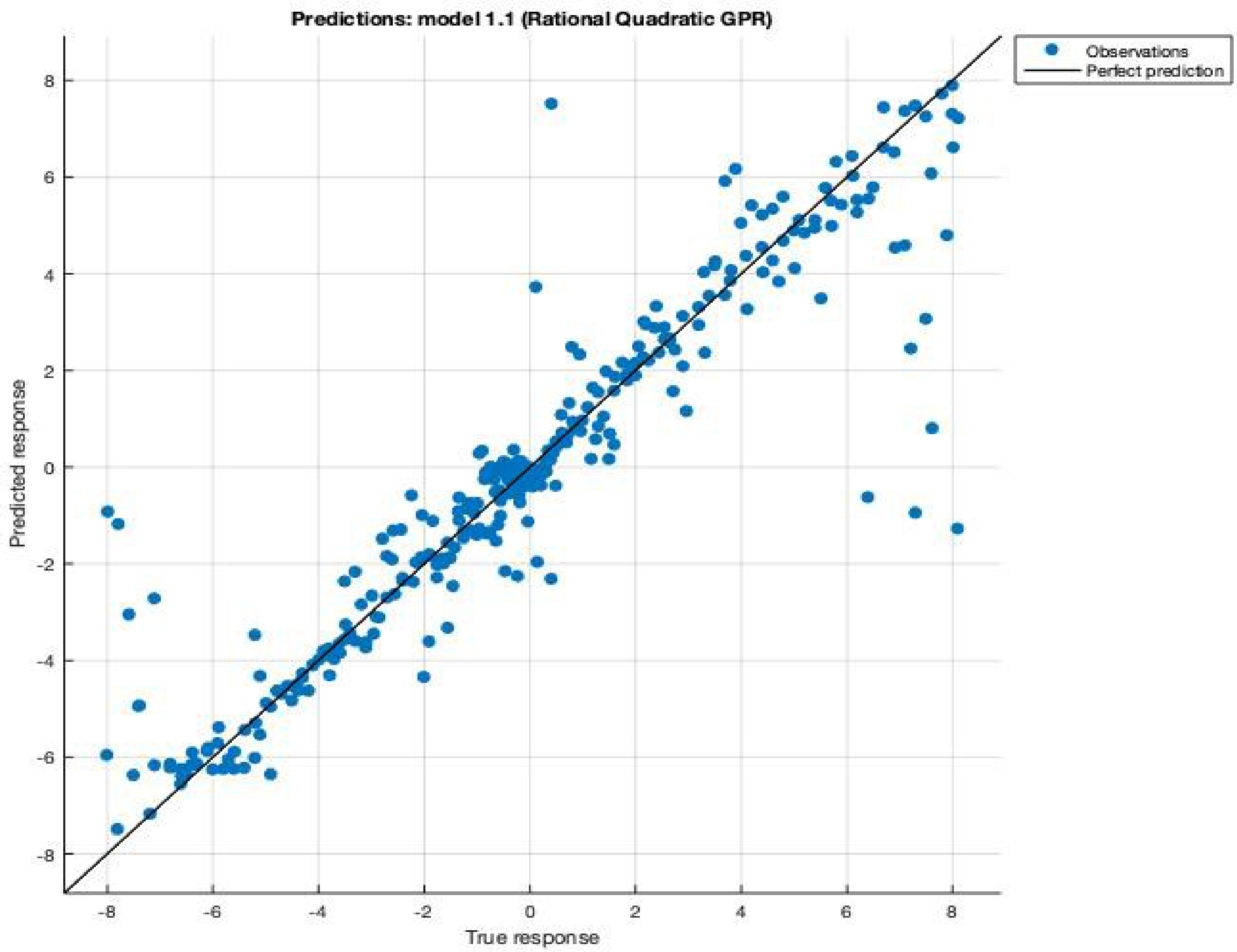


exp. 26

Machine learning

All angles

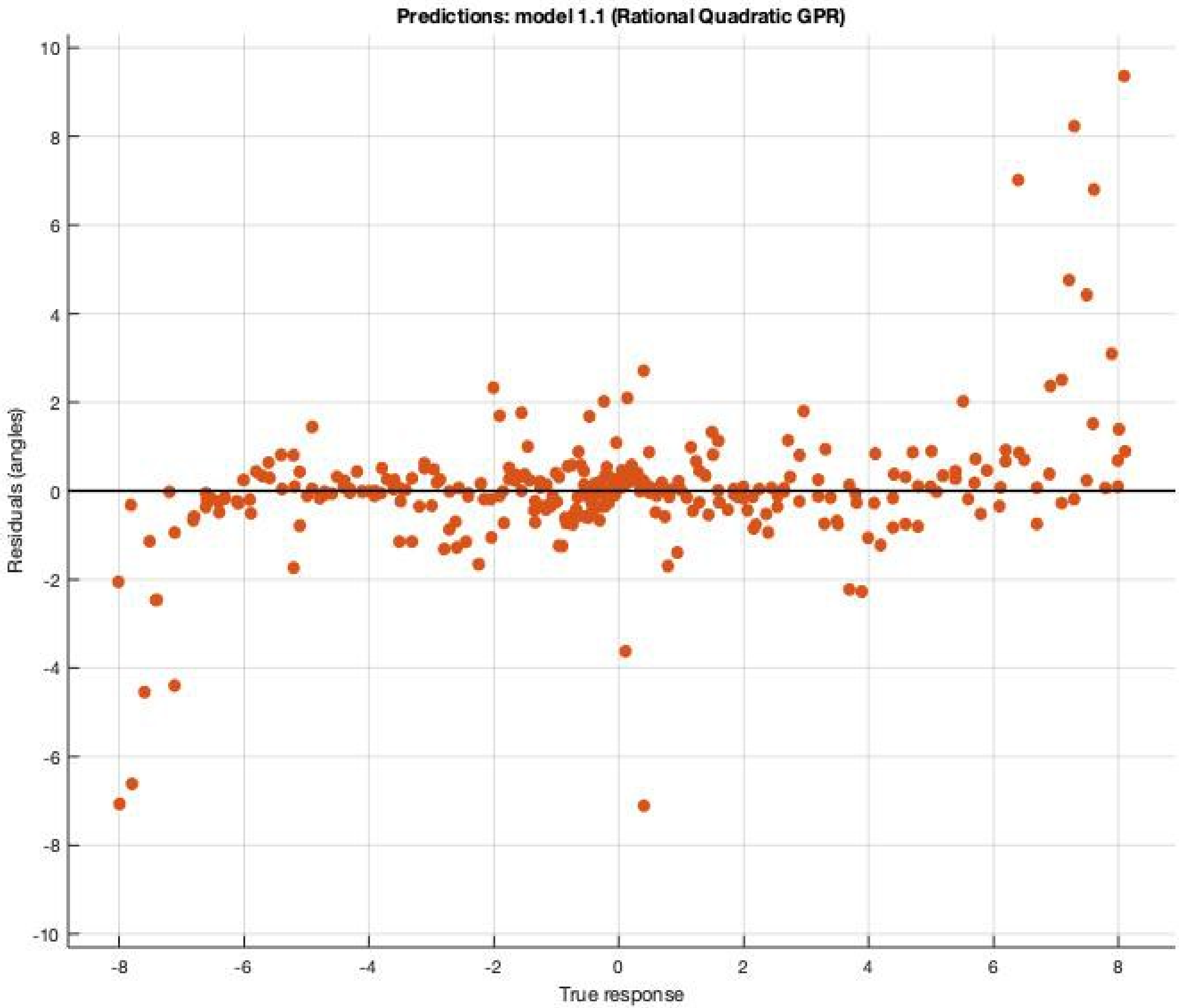
RMSE = 1.52



Machine learning

All angles

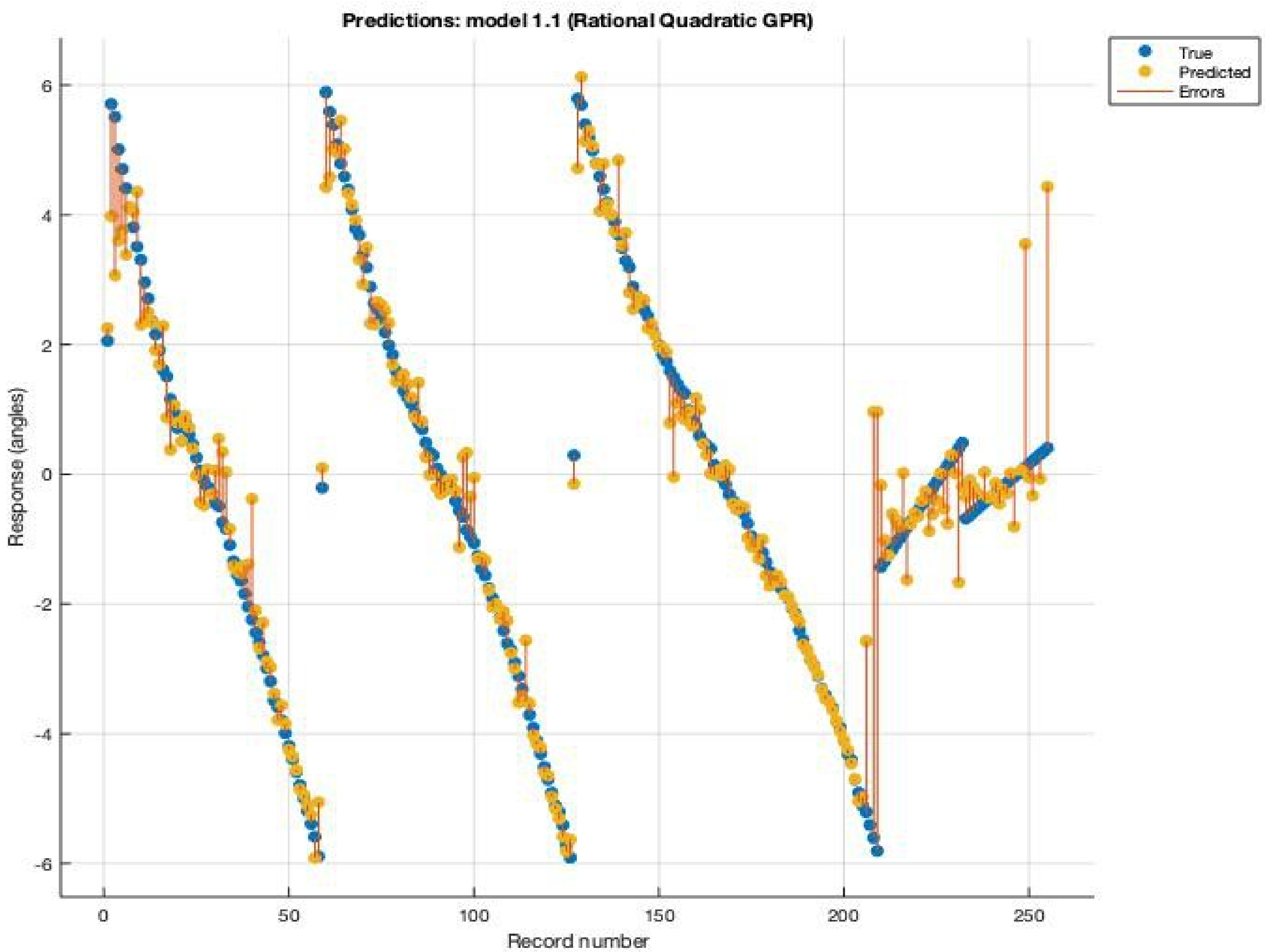
RMSE = 1.52



Machine learning

Angles -6 to 6

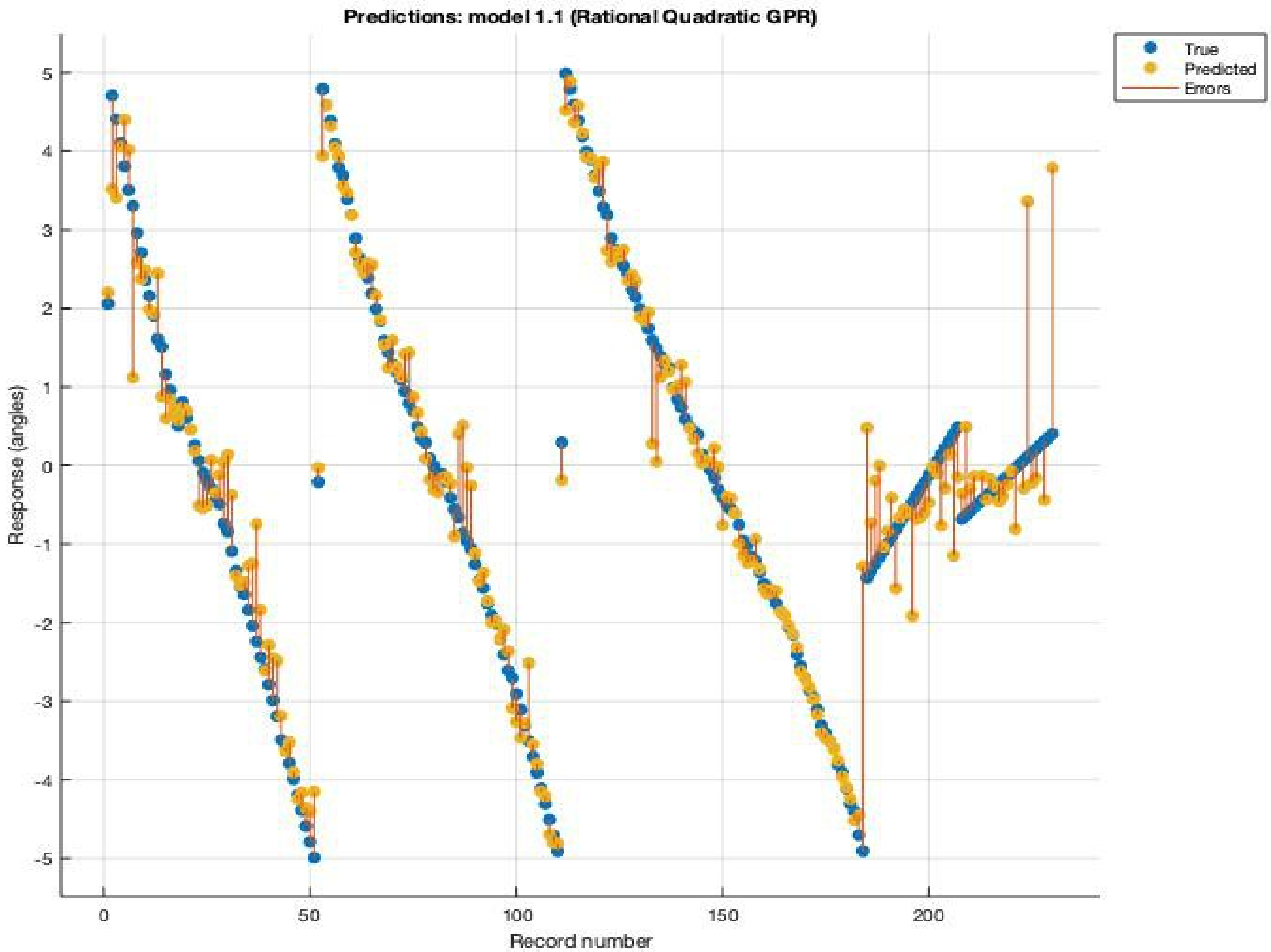
RMSE = 0.66



Machine learning

Angles -5 to 5

RMSE = 0.62

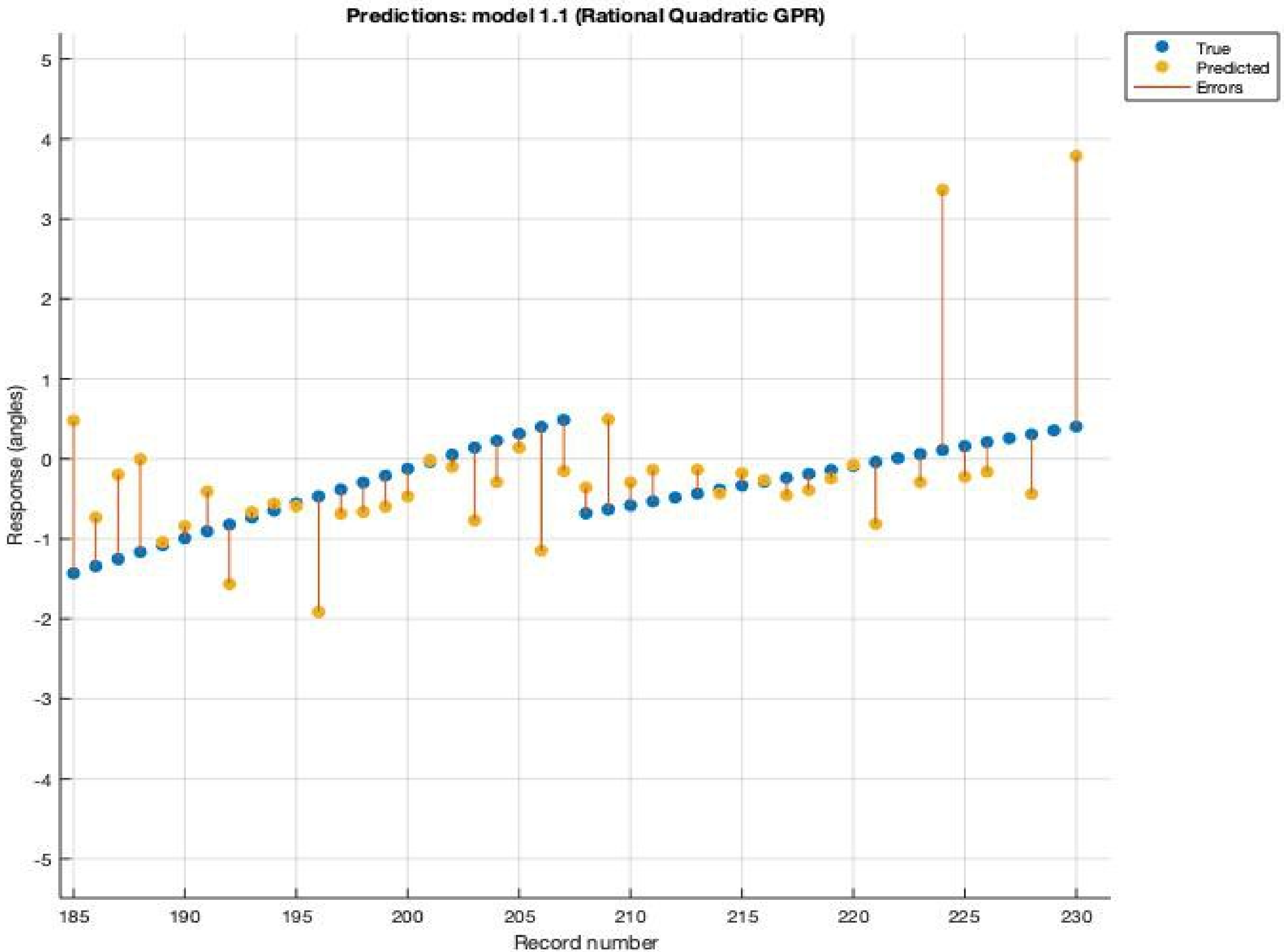


Machine learning

Angles -5 to 5

Validation on rotating experiments

RMSE = 0.26

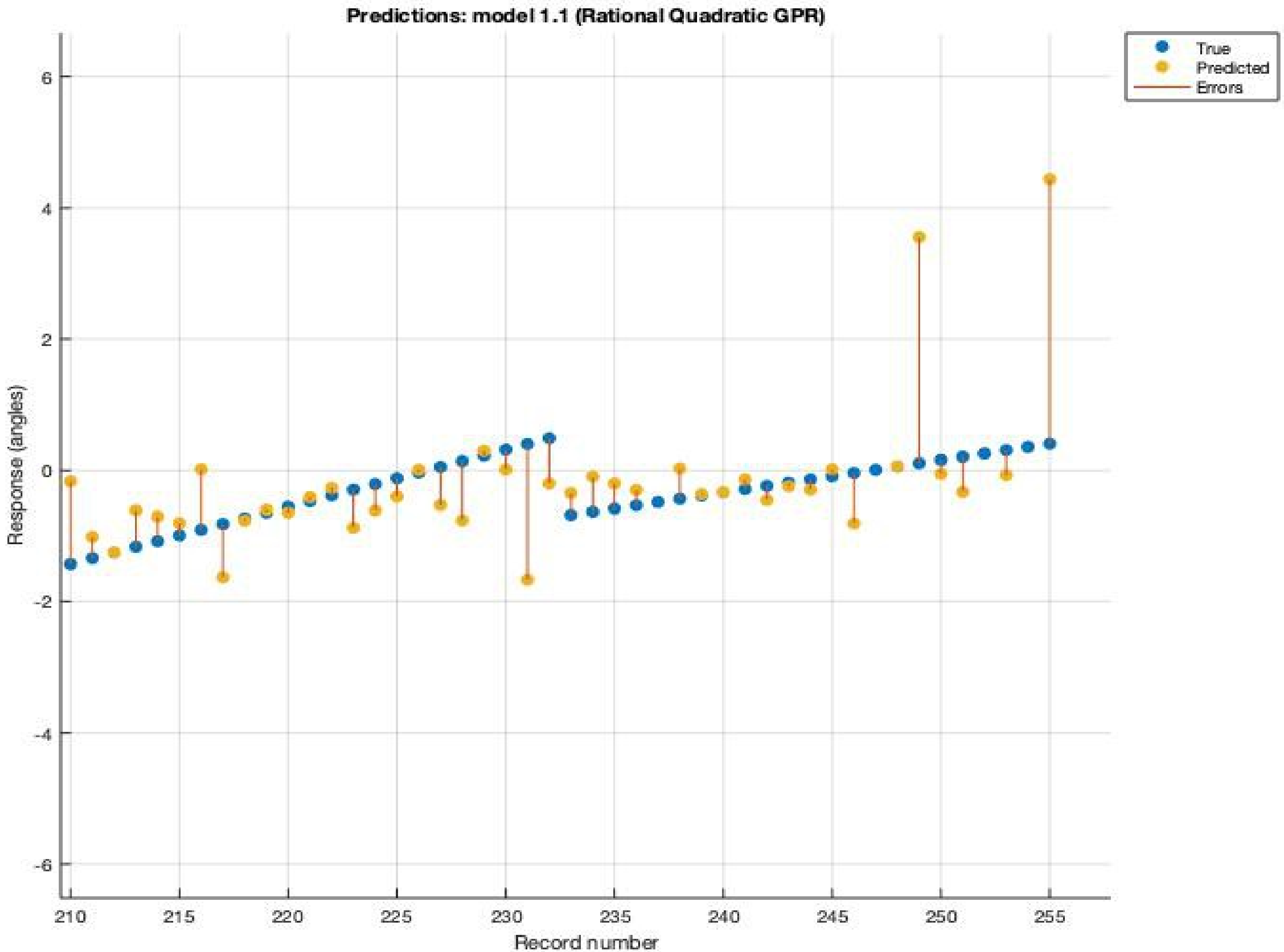


Machine learning

Angles -6 to 6

Validation on rotating experiments

RMSE = 0.19

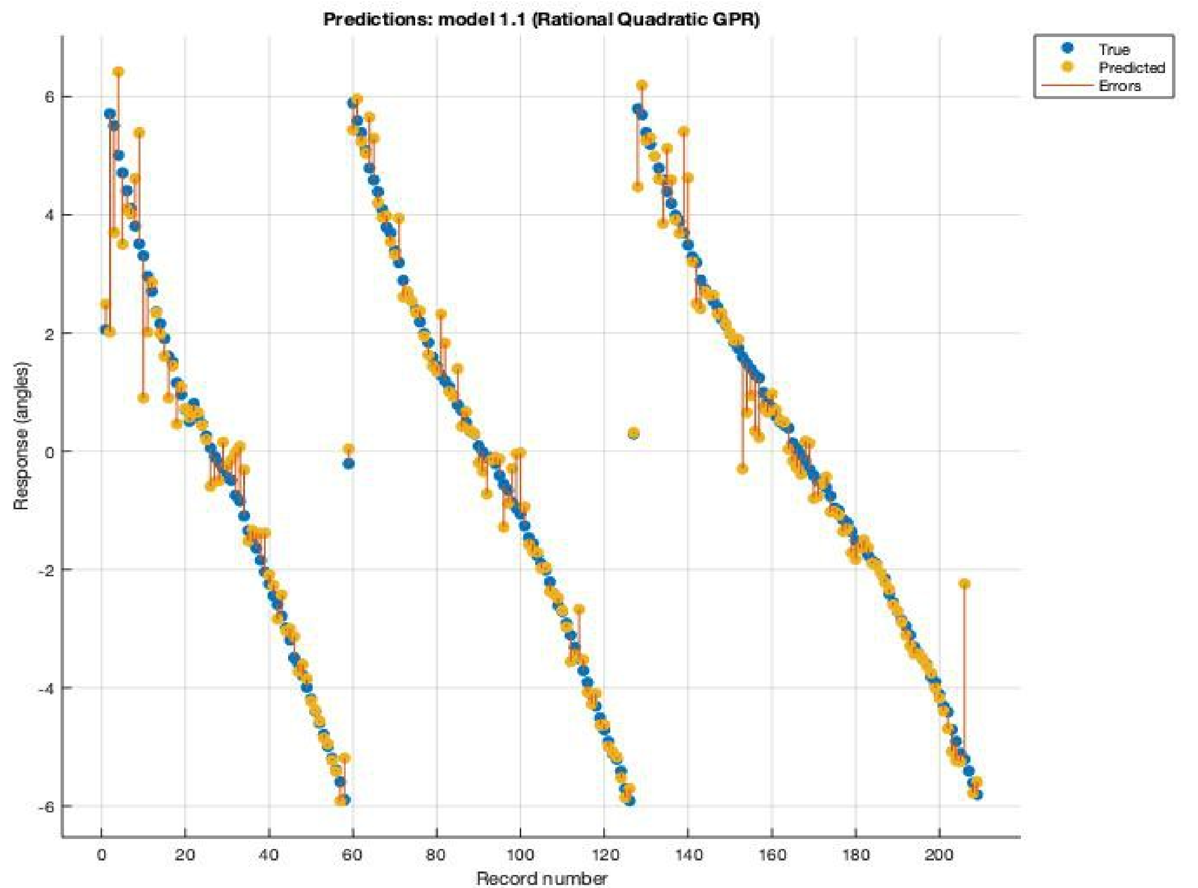


Machine learning

Angles -6 to 6, modelling rotating through static

RMSE = 1.15

degree	1º	2º	3º
sin	0.017	0.035	0.052
distance		elevation error	
100	1.75	3.49	5.23
200	3.49	6.98	10.47
300	5.24	10.47	15.70
400	6.98	13.96	20.93



Takeaways from machine learning

- Generally the approach seems promising
- >1º error is too much, the goal should be to reduce the error to a fraction; at this stage modelling through static data only does not look feasible
- Adding more training data, particularly from rotating experiments with a larger angle range, should improve the modelling (given that the new data does not introduce new errors)
- Gaussian models perform best for this task
- Limiting the angles improved the model
- Limiting the angles too much might reduce the modelling accuracy for Gaussian processes
- Optimising internal model parameters might be explored to improve modelling accuracy, yet that requires a good understanding of Gaussian modelling. Too much optimisation might lead to overfitting.