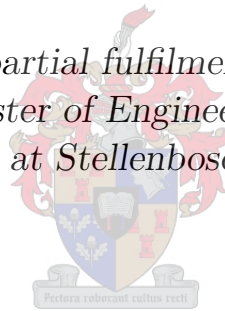


Bayesian Signal Processing of Doppler Radar Data

by

Charl Felix de Villiers

*Thesis presented in partial fulfilment of the requirements
for the degree of Master of Engineering in the Faculty of
Engineering at Stellenbosch University*



Department of Electrical and Electronic Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Supervisor: Prof. J.A. du Preez

December 2016

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date:

Copyright © 2016 Stellenbosch University
All rights reserved.

Abstract

The aim of this thesis is to investigate a Bayesian approach to signal processing of Doppler radar data. The problem of interest involves measured Doppler radar signals measured for golf players' club swings where the frequency shifts are related to the movements of physical objects. Smoothing the frequency shifts of the Doppler signal allows for more accurate estimates of the speeds of the physical objects of interest which is a step towards estimating the velocities of the objects such as the club and ball and can allow one to calculate their trajectories, as their starting points are known. This information would be invaluable to golf players and coaches, who will be able to improve players' skills based on the knowledge of club velocity at impact, the ball spin, and other properties of interest of the golf swing.

We use a Bayesian statistical method called Bayesian spectrum analysis (BSA) to analyse the Doppler signals that were divided into time intervals. BSA allows us to estimate the spectral parameters of the Doppler radar signals in a probabilistic manner, as well as compare competing models in order to select the most probable model from a list of models. We find that the Doppler radar signals contained behaviour that is more complex than our BSA models are able to describe. The BSA results are, however, still useful and can be improved upon by including more prior information.

Our approach is to model the multitarget tracking of the frequency components from BSA in the context of Bayesian probability theory, and to then solve the marginal posterior distributions of the parameters of interest using probabilistic graphical models (PGMs). We compensate for uncertainty in the characteristics of our BSA results by modelling the local signal behaviour, as well as the overall trend of the signal by grouping parts of the signal into segments. These signal segments correspond to different parts of the physical golf swing that contain a different number of objects' Doppler shifts and different signal dynamics. We modelled the segment transition as a left-to-right progression. PGMs are well suited to this modular approach and provide the benefit of deconstructing the problem at hand into a set of local dependencies. We also implemented a "missed-target" model using the PGMs framework. The resulting model resembles a multitarget Kalman filter combined with a

hidden Markov model. We implement the PGMs as both a fully discrete and a hybrid cluster graph and are able to successfully smooth parts of the Doppler radar frequency shifts. We find that the missed-target model and left-to-right segment transition improve upon the conventional multitarget tracking and allow the PGMs to select the correct signal segment and to smooth over regions where a frequency component was missing.

One of the challenges identified in our investigation is estimating both the process noise and measurement noise of the multitarget tracking. Future recommendations include using explicit duration models for the signal segment transitions and using alternative discretisation methods.

Opsomming

Die doelwit van hierdie tesis is die ondersoek na 'n Bayesiese benadering tot seinprosessering van Doppler radar data. Die probleem van belang behels gemete Doppler radarseine wat gerig is op gholfspelers wat gholfstokke swaai. Veranderinge in die weerkaatste seine se frekwensies hou verband met die bewegings van fisiese voorwerpe. Verbeterings op die benaderings van die Doppler-verskuiwings kan lei tot meer akkurate skattings van die spoed van die fisiese voorwerpe. Dit kan lei tot beter beramings van die snelhede van die voorwerpe (soos die gholfstok en -bal) en kan 'n mens toelaat om hul trajekte beter te bereken, aangesien hul beginpunte wel bekend is. Hierdie inligting sal van onskatbare waarde vir gholfspelers en afrigters wees.

Ons maak gebruik van Bayesiese spektrale analise (BSA) om die Doppler-sein, wat in tydstappe opgebreek is, te ontleed. BSA stel ons in staat om die spektrale parameters van die Doppler radarseine met gebruik van waarskynlikheidsleer af te skat, asook om modelle te vergelyk en die mees waarskynlike model te kies. Ons vind dat die Doppler radarseine vervat gedrag wat meer kompleks is as wat ons BSA modelle kan beskryf. Die BSA resultate is egter steeds nuttig en kan verbeter word deur meer inligting in te sluit.

Ons benadering is om die multi-teikenvolging van die frekwensie-komponente van die BSA modelle in die konteks van Bayesiese waarskynlikheidsleer te plaas en om dan die parameters van belang se marginale waarskynlikheidsdigtheidsfunksies te bereken met behulp van waarskynlikheidsgrafiese modelle (PGM'e). Hierdie benadering vergoed vir die inherente statistiese aard van ons BSA resultate deur die modellering van die plaaslike seingedrag, sowel as die algehele tendens van die sein deur die groepering van die seinmonsters in seinsegmente. Hierdie seinsegmente stem ooreen met die verskillende gedeeltes van die fisiese gholfswaai wat verskillende aantal voorwerpe se Doppler-verskuiwings asook verskillende seindinamika bevat. Ons modelleer die segment-oorgange as 'n links-na-regs verloop. PGM'e is goed geskik vir hierdie modulêre benadering en bied voordele aan soos om die probleem te ontbind in plaaslike afhanklikhede. Ons was ook in staat om 'n "gemiste-teiken" model met behulp van die PGM-raamwerk te implementeer. Die model lyk soos 'n multi-teiken Kalman filter gekombineer met 'n verskuilde Markov model. Ons het die PGM'e as beide 'n

diskrete en 'n hibriede bundelgrafiek geïmplementeer en was in staat daartoe om verbeterings te maak op die Doppler-verskuiwings van die radar sein. Ons het gevind dat die gemiste-teiken model en links-na-regs segment-oorgange verbeter op die konvensionele multi-teikenvolging en het toegelaat dat die PGM'e die korrekte seinsegmente kies, asook om frekwensie-komponente op te spoor in gebiede waar 'n frekwensie-komponent vermis was.

Een van die uitdagings wat geïdentifiseer was in ons ondersoek, is die be-raming van beide die proses- en metingsruis van die multi-teikenvolging. Aanbevelings sluit die gebruik van eksplisiete tydsduur modelle vir die sein segment oorgange in, asook die gebruik van alternatiewe diskretiseringsmetodes.

Acknowledgements

I would like to express my sincere gratitude to the following people:

- Prof. Johan du Preez for his excellent advice, boundless knowledge, spirited enthusiasm, and entertaining stories.
- My engineering buddies Colette, Cobus, Sonja, Francois, and Simon for the silly and the serious moments.
- My sister and mother for their love, support and encouragement.
- My father who couldn't see me complete this thesis but taught me the value of learning. "Jy kan nooit verniet 'n boek koop nie".
- My wife Catherine (who has had to share me with this thesis for many years) for her endless patience and love.

Contents

Declaration	i
Abstract	ii
Opsomming	iv
Acknowledgements	vi
Contents	vii
List of Figures	xi
List of Tables	xxi
Nomenclature	xxiv
1 Introduction	1
1.1 Motivation and Research Problem	1
1.2 Background	3
1.3 Objectives	4
1.4 Outcomes and Contributions	5
1.5 Overview of this Work	8
2 Physical Setup	13
2.1 Introduction	13
2.2 Doppler Shift in Radar	13
2.3 Radar Unit	14
2.4 Example Doppler Signal Spectrogram	15
2.5 Conclusion	17
3 Bayesian Spectrum Analysis	18
3.1 Introduction	18
3.2 Bayesian Probability Theory	18
3.3 Introduction to BSA	23
3.4 A BSA Example	24

3.5	BSA Parameter Estimation	24
3.6	BSA Relation to Periodogram	28
3.7	BSA Model Selection	29
3.8	Conclusion	30
4	BSA Experimental Results	31
4.1	Introduction	31
4.2	Signal Models	31
4.3	Finding the Peak	32
4.4	Simulated Examples	32
4.5	Doppler Radar Signal Noise and “Noiseless” Model	36
4.6	Long Segment BSA	37
4.7	Short Time-step BSA	46
4.8	Signal Complexity	48
4.9	Conclusion	50
5	Bayesian State-Space Estimation	51
5.1	Introduction	51
5.2	Single-Target Tracking	52
5.3	Multitarget Tracking	54
5.4	Relation to Kalman Filters	55
5.5	Discrete Markov Chains	55
5.6	Overview of Other Target Tracking Works	57
5.7	Model Developed in Thesis	59
5.8	Conclusion	60
6	Probabilistic Graphical Models	62
6.1	Introduction	62
6.2	Conditional Independence	62
6.3	Clique Trees and Cluster Graphs	63
6.4	Discrete Table Factors	69
6.5	Continuous Gaussian Factors	71
6.6	Hybrid Conditional Linear Gaussian	75
6.7	Comparison of Discretised and Hybrid Computational Tractability	76
6.8	Conclusion	77
7	PGM Model Design	78
7.1	Introduction	78
7.2	General Model Overview	79
7.3	Discrete Model	85
7.4	Hybrid Model	88
7.5	Conclusion	90
8	Implementation Details and Challenges	91

<i>CONTENTS</i>	ix
8.1 Introduction	91
8.2 Discretisation	91
8.3 CLG Factors Implementation Subtleties	92
8.4 Finding the BSA Peak	96
8.5 Non-convergence of PGMs	96
8.6 Conclusion	96
9 PGM Experimental Results	97
9.1 Introduction	97
9.2 Discretise Model Results	100
9.3 Hybrid Model Results	104
9.4 Conclusion	119
10 Conclusion and Recommendations	120
10.1 Conclusion	120
10.2 Recommendations and Future Work	123
Appendices	125
A BSA Detail	126
A.1 BSA Parameter Estimation	126
A.2 BSA Relation to Periodogram	131
A.3 BSA Model Selection	133
A.4 BSA Estimated Noise Variance $\langle \sigma^2 \rangle$	136
B Miscellaneous Integrals and Algorithms	138
B.1 Gaussian Integrals	138
B.2 Laplace's Method	138
B.3 Nelder-Mead Optimisation	139
C Simulated Signal BSA	140
C.1 Introduction	140
C.2 Single Stationary Sinusoid	141
C.3 Two Stationary Sinusoids	143
C.4 Three Stationary Sinusoids	145
C.5 Single Chirped Sinusoid	147
C.6 Two Chirped Sinusoids	150
C.7 Three Chirped Sinusoids	153
C.8 One Quadratic Chirped Sinusoid	156
D Doppler Radar Signal BSA	159
D.1 Introduction	159
D.2 Signal Segments	159
D.3 Segment A: Golf Club Pre-Impact	161
D.4 Segment B: Ball Only	167

*CONTENTS***x**

D.5	Segment C: Club and Ball Post-Impact	170
D.6	Segment D: Club and Stick-Flash Pre-Impact	176
D.7	Segment E: Club and Stick-Flash No Impact	182
D.8	Segment F: Club, Stick-Flash and Ball Post-Impact	188
E	Explicit Duration Model	194
E.1	Explicit Duration	194
E.2	Minimum Duration	195
F	Factor Details	197
F.1	Linear Transitions	197
F.2	Discrete Factor Examples	198
	List of References	200

List of Figures

1.1	An example spectrogram of a golf swing's Doppler radar data, which contains multiple frequency components.	2
1.2	An example of a segmented spectrogram of a golf swing's Doppler radar data. Each segment has either different frequency components or frequency dynamics.	9
1.3	A Bayes network representation of the general model approach.	12
2.1	The setup of the Doppler radar unit in the vertical plane (not to scale). The trajectories over time of the club-head, ball and player are indicated by the red, blue and black dotted directed lines respectively. Example radar signal paths from the objects to the Doppler radar unit are indicated by dashed lines in their corresponding colours. The green dashed path is a possible alternative path that a reflected signal can follow for the club-head.	14
2.2	The spectrogram of a golf swing indicating the absolute frequency shift from the transmitted signal which is proportional to the radial velocity of the object. Imposed on the spectrogram are the frequency paths of the ball, club-head, stick-flash (the radar reflection running up the shaft) and a harmonic frequency caused by the clipping of the signal. There are also low-frequency components created by the player motion or other unwanted effects.	16
2.3	Normalised histogram of a Doppler radar signal part that contains clipping.	17
2.4	Normalised histogram of a Doppler radar signal part that contains no clipping.	17

3.1	This figure has two examples of BSA, one in each column. The left column contains the example of a single sinusoidal signal in noise. From the top the figures are the periodogram, log of the probability density, and probability density functions of the frequency of the signal using a single signal model with unknown noise. The peak of the periodogram coincides with the peak of the BSA analysis. The second column contains an example of the BSA of two signals in noise where the signals are well separated and equally strong and clearly visible. From the top the figures are the periodogram, log of the probability density, and probability density functions of the frequencies of the signal with a two signal model with unknown noise. Note that the two peaks of the probability density are equal and coincide with the top two maxima of the periodogram.	25
4.1	The log of the chirpogram of a simulated signal with a linear chirp: $\omega = 0.3$, $\alpha = -0.001$, $\beta = 0.0$. The density function has a single strong peak by the correct parameter values.	33
4.2	The log of the chirpogram of a simulated signal with a quadratic chirp: $\omega = 0.2$, $\alpha = -0.003$, $\beta = -0.00001$. The density function is more spread out in comparison to Figure 4.1.	33
4.3	The absolute value of a zero-padded DFT of two very close sinusoids ($\omega_0 = 0.3$, $\omega_1 = 0.295$, $\sigma = 0.6$). The DFT cannot distinguish between the two sinusoids.	34
4.4	Log probability density function of two very close sinusoids with peak: $\omega_0 = 0.2998$, $\omega_1 = 0.2951$. BSA separates and distinguishes between the two sinusoids.	35
4.5	Spectrogram of a single sinusoid with quadratic chirp. It is not immediately apparent from the spectrogram that there is only one frequency component present. BSA, however, selects the correct model for this signal with near 100% probability from nine possible models.	36
4.6	Spectrogram of a golf swing hitting a golf ball segmented into different parts: Window Size = 64, Step-size = 16. The frequency components of interest are caused by the club (before and after impact), the stick-flash, and the ball. The low-frequency component near time-step 60 is caused by the motion of the player's body. . .	38
4.7	Spectrogram of a golf swing without hitting a ball segmented into different parts: Window Size = 64, Step-size = 16. The frequency components of interest are caused by the club and the stick-flash. The low-frequency component near time-step 100 is caused by the motion of the player's body.	39

4.8	Segment A (club downward swing) Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram. The two most likely models have near identical instantaneous frequencies which do not completely match the spectrogram.	40
4.9	The chirpogram of Segment A (club downward swing) Example 1. The probability distribution resembles a linear chirped sinusoid despite the quadratic chirped model being the most likely model. . .	41
4.10	Segment B (ball post club swing) Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram. The low-frequency component is likely caused by the low-frequency effects we saw in Figure 2.2. The two very close frequency components could be caused by two very close sinusoids.	42
4.11	Segment D (pre-impact club-head and stick-flash) Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram. Only the stick-flash was picked up by BSA with multiple frequency components being attributed to it.	44
4.12	Segment E (no impact stick-flash) Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram. The club-head, stick-flash and a harmonic of the stick-flash was picked up by BSA.	45
4.13	Scatter-plot of BSA stationary sinusoid frequency values with the unknown noise BSA model. The size of each marker indicates the probability of the specific BSA model. Note that some of the markers are very close. For example, at time-step 40 two of the markers for the three frequency components are close enough to overlap. .	47
4.14	The BSA model probabilities from Figure 4.13	47
4.15	Scatter-plot of BSA stationary sinusoid frequency values with the specified noise BSA model. The size of each marker indicates the probability of the specific BSA model. Note that the less complex models are now selected more often compared to Figure 4.13. . . .	49
4.16	The BSA model probabilities from Figure 4.15	49
5.1	Bayes network representation of state-space estimation.	54
5.2	An uninformative discrete Markov chain. The model is only uninformative if the three edges leaving each node are equally likely. .	56
5.3	A left-to-right discrete Markov chain. States can progress in a strict order.	57
5.4	An example of the left-to-right model probabilities using the transitions in Table 5.1 for 30 time-steps. We plot the probabilities where no other information is known about the states other than the prior transitions.	58
5.5	Switching state-space estimation model	59

5.6	Bayes network as a conceptual representation of the model developed in this thesis. An additional discrete Markov chain (indicated in red) is added to incorporate additional temporal information.	60
6.1	Both PGMs factorise the same joint probability density, $p(A, B, C, D)$, which is the normalised product of the factors: $\Phi_0, \Phi_1, \Phi_2, \Phi_3$	64
6.2	An illustration of the message-passing algorithm used in BP. The cluster \mathcal{C}_3 sends a message to the cluster to its right using the incoming messages from the left.	66
6.3	An illustration of the RIP. If the sepset (indicated in blue) is present, the sepset (indicated in red) will violate RIP if added.	68
7.1	A Bayes network representation of the general model approach. Green: lower-level (likelihoods), Blue: mid-level (multitarget parameter dynamics), Red: high-level (Segment transitions).	80
7.2	Discrete Doppler radar PGM. The red, blue, and green clusters indicate the high-level segment transitions, mid-level parameter dynamics, and lower-level likelihood functions respectively.	88
7.3	Hybrid Doppler radar PGM. The red, blue, and green clusters indicate the high-level segment transitions, mid-level parameter dynamics, and lower-level likelihood functions respectively.	89
8.1	Subset of Figure 7.3, the hybrid Doppler radar PGM. It indicates the messages passed to a mid-level transition cluster containing a CLG factor.	94
9.1	Images can be represented using varying brightness in red green and blue. Figure (a) shows red, green, and blue of varying brightness corresponding to the relative probability density of a frequency component. Figure (b) displays different full intensity colour combinations of red, green, and blue with the intensity indicated in the format: {R;G;B}	99
9.2	In Figure (a) we see simulated examples of summed marginal probability density functions for a given time-step. Each colour represents a BSA model. Using the values from a graph such as Figure (a), we can represent the graph as an RGB image for each time-step in order to visualise the graphs as they change over time. Figure (b) is an example RGB time-step of Figure (a).	99

9.3	The spectrogram of Segment C and B of Doppler radar data. There are two strong sinusoids visible, one stationary (the ball) and one with a negative chirp (the club head). Note the presence of low-frequency components, which are stronger than the club head at time-step 0. There is also a slight “wobble” in the frequency component of the ball caused by a reflective material that was placed in the ball.	101
9.4	PGM analysis of Segment C and B of Doppler radar data. The segment transitions are correctly estimated. However, the lower-frequency noise is incorrectly seen as part of the club head between time-step 0 and 5.	101
9.5	PGM analysis of Segment C and B of Doppler radar data with a restriction of a negative chirp on the lower-frequency component. The segment transitions are correctly estimated. Since the lower-frequency component is restricted to having a negative chirp, the model can now correctly smooth the club head’s frequency backwards in time.	102
9.6	The spectrogram of Segments A, E and G of Doppler radar data. There is one frequency component (club head), then two (club head and stick-flash) and then one again (club head). (We ignore the low-frequency component near time-step 35.) However, the club head is obscured during the stick-flash. There are also harmonics present in the spectrogram as there was clipping present in the signal.	103
9.7	PGM analysis of Segment A, E and G of Doppler radar data without left-to-right segment transitions. The segment transitions are not correctly estimated, with the single-frequency model being chosen as the most likely for the most of the signal. Note the magenta parts of the analysis, which indicate a mixture of single- and two-frequency models near time-step 0 and 40.	103
9.8	PGM analysis of Segment A, E and G of Doppler radar data which includes the left-to-right segment transitions. The segment transitions are correctly estimated with high certainty. The analysis smooths over the part where the club head was not visible in the spectrogram.	104
9.9	The spectrogram of Segment C and B of Doppler radar data. This signal is similar to Figure 9.3. There are two strong sinusoids visible, one stationary (the ball) and one with a negative chirp (the club head). Note the presence of low-frequency components, which are stronger than the club head at one point.	105

9.10	PGM analysis of Segment C and B of Doppler radar data. The segment transitions are correctly estimated. However, the lower-frequency noise is incorrectly seen as part of the club head at time-step 0. This is similar to the discrete model analysis in Figure 9.4. For the hybrid model, the peaks are very sharp and almost not visible as a result. This is caused by the sharp BSA factor peaks.	106
9.11	An enlargement of the top region of Figure 9.10. Note the sharp peaks.	107
9.12	An enlargement of the bottom region of Figure 9.10. Note the sharp peaks and the “jump” of the frequency component between time-step 0 and 1.	107
9.13	PGM analysis of Segment C and B of Doppler radar data using the BSA means but with artificially created covariances. The segment transitions are correctly estimated. The peaks are much less sharp than Figure 9.10 and the signal dynamics much more closely resemble the linear Gaussian dynamics.	108
9.14	The spectrogram of Segment F, C, and B of Doppler radar data. There are three strong sinusoids visible, one stationary (the ball) and two with a negative chirp (the club head in the middle and stick-flash at the bottom). Note the middle frequency component is relatively weak from time-step 0 to 30.	108
9.15	This is an enlargement of the top region of Figure 9.17.	109
9.16	This is an enlargement of the top region of Figure 9.18. Note the sharp peaks	109
9.17	PGM analysis of Segment F, C and B of Doppler radar data without left-to-right segment transitions. For the hybrid model, the peaks are very sharp and almost not visible as a result. This is caused by the sharp BSA factor peaks. Note the uncertainty between the active BSA model indicated by the cyan and magenta colours.	110
9.18	PGM analysis of Segment F, C and B of Doppler radar data. The segment transitions are correctly estimated. For the hybrid model, the peaks are very sharp and almost not visible as a result. This is caused by the sharp BSA factor peaks.	111
9.19	The spectrogram of Segments A, E and G of Doppler radar data. This spectrogram is the same signal as Figure 9.6. There is one frequency component (club head), then two (club head and stick-flash) and then one again (club head). (This is ignoring the low-frequency components between time-step 23 and 33.) During the stick-flash, the club head is obscured.	112
9.20	PGM analysis of Segment C and B of Doppler radar data. The segment transitions were not estimated and were instead specified in order to demonstrate the difference the missed target model makes in Figure 9.22.	113

9.21	PGM analysis of Segment C and B of Doppler radar data using the BSA means but with artificially created covariances. The segment transitions were not estimated and were instead specified in order to demonstrate the difference the missed target model makes in Figure 9.22.	113
9.22	PGM analysis of Segment C and B of Doppler radar data. The missed target model is able to smooth the position of the club head where it was not visible to BSA.	114
9.23	The spectrogram of Segments A and D of Doppler radar data. There is one frequency component (club head), then two (club head and player motion), and then a different two (club head and stick-flash). In this example we distinguish between Segment A without player motion as A, and with player motion as A'. However, during the stick-flash, the club head is obscured.	115
9.24	PGM analysis of Segment A and D of Doppler radar data. The segment transitions were not estimated and specified in order to demonstrate the difference the missed-target model makes in Figure 9.25.	115
9.25	PGM analysis of Segment C and D of Doppler radar data. The missed-target model is able to smooth the position of the club head where it was not visible to BSA. Note the faint colour of the club head. It continues as a linear extension from the previous segment without receiving any information from the BSA factors.	116
9.26	This is an enlargement of the top region of Figure 9.25. We include the maximum values of the club-head position. Note that the PGM becomes more uncertain of the club head position for each time-step.	117
9.27	The spectrogram of Segments A, D, F, C, and B of Doppler radar data. This spectrogram contains the frequency components from Figure 9.23 and Figure 9.14.	117
9.28	PGM analysis of Segments A, D, F, C, and B of Doppler radar data. The missed-target model is able to smooth the position of the club head where it was not visible to BSA from time-steps 34 to 51. The smoothed position is, however, so faint that one cannot see it in this figure. The segment transitions were not estimated and needed to be specified. The large variance near time-step 30 (the last step for $M = 1$) is likely due to a combination of the faint frequency component which can be seen in the spectrogram and the jump in frequency to the new BSA model.	118
9.29	This is an enlargement of the first region of Figure 9.28. The smoothed club head is very faint, and we therefore included the error bars and the mean. Note that the PGM is more certain of the club-head position at the edges of the missing-target model's segment.	119

C.1	DFT of a simulated single stationary sinusoid. $\omega_0 = 0.3, \sigma = 0.6$.	141
C.2	Spectrogram of a simulated single stationary sinusoid. $\omega_0 = 0.3, \sigma = 0.6$	141
C.3	DFT of two simulated stationary sinusoids. $\omega_0 = 0.3, \omega_1 = 0.2, \sigma = 0.6$	143
C.4	Spectrogram of two simulated stationary sinusoids. $\omega_0 = 0.3, \omega_1 = 0.2, \sigma = 0.6$	143
C.5	DFT of three simulated stationary sinusoids. $\omega_0 = 0.3, \omega_1 = 0.2, \omega_2, \sigma = 0.6$	145
C.6	Spectrogram of three simulated stationary sinusoids. $\omega_0 = 0.3, \omega_1 = 0.2, \omega_2, \sigma = 0.6$	145
C.7	DFT of a single simulated chirped sinusoid. $\omega_0 = 0.3, \alpha = -0.001$.	147
C.8	Spectrogram of a single simulated chirped sinusoid. $\omega_0 = 0.3, \alpha = -0.001$	149
C.9	Theoretical instantaneous frequency of the top two hypotheses. .	149
C.10	DFT of two simulated chirped sinusoids.	150
C.11	Spectrogram of two simulated chirped sinusoids.	152
C.12	Theoretical instantaneous frequency of the top two hypotheses. .	152
C.13	DFT of three simulated chirped sinusoids.	153
C.14	Spectrogram of three simulated chirped sinusoids.	155
C.15	Theoretical instantaneous frequency of the top two hypotheses. .	155
C.16	DFT of a single simulated quadratic chirped sinusoid.	156
C.17	Spectrogram of a single simulated quadratic chirped sinusoid . . .	158
C.18	Theoretical instantaneous frequency of the top two hypotheses. .	158
D.1	Spectrogram of a golf swing hitting a golf ball segmented into different parts: Window Size = 64, Stepsize = 16.	160
D.2	Spectrogram of a golf swing without hitting a ball segmented into different parts: Window Size = 64, Stepsize = 16.	160
D.3	Segment A Example 1: DFT of Segment A containing a non-linear chirped signal around $\omega = 0.1$ and low frequency noise from player movement around $\omega = 0.01$	161
D.4	Segment A Example 1: Spectrogram of Segment A containing a non-linear chirped signal around $\omega = 0.1$ and low frequency noise from player movement around $\omega = 0.01$ time-step	162
D.5	Segment A Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram	162
D.6	Segment A Example 2: DFT of Segment A containing a non-linear chirped signal around $\omega = 0.1$ and low frequency noise from player movement around $\omega = 0.03$	164
D.7	Segment A Example 2: Spectrogram of Segment A containing a non-linear chirped signal around $\omega = 0.1$ and low frequency noise from player movement around $\omega = 0.01$ time-step	165

D.8 Segment A Example 2: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram	165
D.9 Segment B Example 1: DFT of Segment B a chirped signal around $\omega = 0.24$ and a low frequency component around $\omega = 0.02$	167
D.10 Segment B Example 1: Spectrogram of Segment B containing a chirped signal around $\omega = 0.24$	168
D.11 Segment B Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram	168
D.12 Segment C Example 1: DFT of Segment C a signal around $\omega = 0.24$, a chirped single around $\omega = 0.12$ and low frequency noise from player movement around $\omega = 0.04$	170
D.13 Segment C Example 1: Spectrogram of Segment C containing a signal around $\omega = 0.24$, a chirped single around $\omega = 0.12$ and low frequency noise from player movement around $\omega = 0.04$ time-steps 1 to 6	171
D.14 Segment C Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram	171
D.15 Segment C Example 2: DFT of Segment C a signal around $\omega = 0.24$, a chirped single around $\omega = 0.07$ and low frequency noise around $\omega = 0.02$	173
D.16 Segment C Example 2: Spectrogram of Segment C containing a signal around $\omega = 0.24$, a chirped single around $\omega = 0.07$ and low frequency noise around $\omega = 0.02$	174
D.17 Segment C Example 2: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram	174
D.18 Segment D Example 1: DFT of Segment D a chirped signal around $\omega = 0.1$ and $\omega = 0.17$ and possibly a harmonic of that signal around $\omega = 0.3$	176
D.19 Segment D Example 1: Spectrogram of Segment D containing a chirped signal around $\omega = 0.2$ and $\omega = 0.1$ and a harmonic of that signal around $\omega = 0.4$ and $\omega = 0.3$	177
D.20 Segment D Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram	177
D.21 Segment D Example 2: DFT of Segment D a chirped signal around $\omega = 0.13$ and $\omega = 0.1$ and possibly a harmonic of that signal around $\omega = 0.27$	179
D.22 Segment D Example 2: Spectrogram of Segment D containing a chirped signal around $\omega = 0.13$ and $\omega = 0.1$ and a harmonic of that signal around $\omega = 0.3$ and $\omega = 0.27$	180
D.23 Segment D Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram	180
D.24 Segment E Example 1: DFT of Segment E containing multiple chirped signals	182

D.25 Segment E Example 1: Spectrogram of Segment E containing a chirped signal from around $\omega = 0.12$ to $\omega = 0.01$ and its harmonic from around $\omega = 0.22$, time-step 2, to $\omega = 0.12$, time-step 12 and a signal $\omega = 0.12$ from time-step 12 to 17.	183
D.26 Segment E Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram	183
D.27 Segment E Example 2: DFT of Segment E containing multiple chirped signals	185
D.28 Segment E Example 2: Spectrogram of Segment E containing a chirped signal from around $\omega = 0.11$ to $\omega = 0.04$ and its harmonic from around $\omega = 0.2$, time-step 2, to $\omega = 0.11$, time-step 12 and a signal $\omega = 0.12$ from time-step 12 to 14.	186
D.29 Segment E Example 2: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram	186
D.30 Segment F Example 1: DFT of Segment F containing three signals.	188
D.31 Segment F Example 1: Spectrogram of Segment F containing a signal around $\omega = 0.21$, a chirped signal around $\omega = 0.1$ and a chirped signal around $\omega = 0.04$	189
D.32 Segment F Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram.	189
D.33 Segment F Example 2: DFT of Segment F containing three signals	191
D.34 Segment F Example 2: Spectrogram of Segment F containing a signal around $\omega = 0.29$, a chirped signal around $\omega = 0.12$ and a chirped signal around $\omega = 0.03$	192
D.35 Segment F Example 2: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram.	192
E.1 Left-to-right model prior probabilities.	194
E.2 Minimum Duration model prior probabilities.	194
E.3 Ferguson Explicit Duration Markov chain	195
E.4 Ferguson Minimum Duration Markov chain	196

List of Tables

4.1	The BSA model probabilities for two very close sinusoids $\omega_0 = 0.3$, $\omega_1 = 0.295$ in Figure 4.4. The most likely model is correctly selected as the two stationary-frequency model.	35
4.2	The most likely signal parameters of Segment A Example 1. The signal model is a single quadratic chirped sinusoid selected with near 100% probability.	40
4.3	The most likely signal parameters of Segment B Example 1. The signal model is three stationary sinusoids selected with near 100% probability.	42
4.4	The most likely signal parameters of Segment D Example 1. The signal model is three chirped sinusoids selected with near 100% probability.	43
4.5	The most likely signal parameters of Segment E Example 1. The signal model is three chirped sinusoids selected with near 100% probability.	45
5.1	Transition table describing the probabilities of the left-to-right discrete Markov chain from Figure 5.3.	57
6.1	An example discrete factor representing $p(A, B C)$	70
7.1	The state-space estimation parameters of interest describing the signal's frequency dynamics	80
7.2	A discrete factor describing an example left-to-right segment transition of three segments.	82
7.3	A discrete factor describing an example relationship between signal segments and the active BSA model	82

7.4	This table describes how we calculate the possible values for the discrete factors of the form $p(\boldsymbol{\omega}_{t+1} \boldsymbol{\omega}_t, \boldsymbol{\Delta}\boldsymbol{\omega}_{t+1})$. The random variables on the right of the conditioning bar (denoted with the double line), determine the possible values of the random variables on the left of the conditioning bar for the given model transition. The place-holder values in the factor are ϕ , \mathfrak{A} , \mathfrak{B} , \mathfrak{C} , and \mathfrak{D} . The ϕ value denotes that there is no target, \mathfrak{A} and \mathfrak{C} indicate a transition of the top target, and \mathfrak{B} and \mathfrak{D} of the bottom target.	87
C.1	Estimated parameters and model selection of a simulated single stationary sinusoid. The most likely model is indicated in blue. . .	142
C.2	Estimated parameters and model selection of two simulated stationary sinusoids. The most likely model is indicated in blue. . . .	144
C.3	Estimated parameters and model selection of three simulated stationary sinusoids. The most likely model is indicated in blue. . . .	146
C.4	Estimated parameters and model selection of a single simulated chirped sinusoid. The most likely model is indicated in blue. . . .	148
C.5	Estimated parameters and model selection of two simulated chirped sinusoids. The most likely model is indicated in blue.	151
C.6	Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	154
C.7	Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	157
D.1	Segment A Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	163
D.2	Segment A Example 2: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	166
D.3	Segment B Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	169
D.4	Segment C Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	172
D.5	Segment B Example 2: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	175
D.6	Segment D Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	178

D.7	Segment D Example 2: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	181
D.8	Segment E Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	184
D.9	Segment E Example 2: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	187
D.10	Segment F Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	190
D.11	Segment F Example 2: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.	193
F.1	This table describes how we calculate the possible values for the discrete factors of the form $p(\boldsymbol{\omega}_{t+1} \boldsymbol{\omega}_t, \boldsymbol{\Delta}\boldsymbol{\omega}_{t+1})$. Note that there are two scenarios each for switching between the signal models; one involving the top frequency component and one the bottom. . . .	199
F.2	This table describes how we calculate the possible values for the discrete factors of the form $p(\boldsymbol{\Delta}\boldsymbol{\omega}_{t+1} \boldsymbol{\Delta}\boldsymbol{\omega}_t, \mathbf{a}_{t+1})$. Note that it takes two time-steps for the dynamics to switch between different signal models.	199

Nomenclature

List of Abbreviations

BSA	Bayesian Spectrum Analysis.
BP	Belief Propagation.
BU	Belief Update.
CLG	Conditional Linear Gaussian.
CPD	Conditional Probability Density.
DFT	Discrete Fourier Transform.
FFT	Fast Fourier Transform.
HMM	Hidden Markov Model.
JMPDT	Joint Multitarget Probability Density tracking.
LBP	Loopy Belief Propagation.
LBU	Loopy Belief Update.
MHT	Multiple Hypothesis Tracking.
PGM	Probabilistic Graphical Model.
RBP	Residual Belief Propagation.
RGB	Red Green Blue.
RIP	Running Intersection Property.
SSSE	Switching State-Space Estimation
TOMHT	Track-Oriented Multiple Hypothesis Tracking.

List of Symbols

A	A random variable.
\mathbf{A}	Orthonormalised signal model component amplitudes $\{A_0, \dots, A_m\}$.
\mathbf{a}	Difference between two sets of $\Delta\boldsymbol{\omega}$.
B	A random variable.
\mathbf{B}	Signal model component amplitudes $\{B_0, \dots, B_m\}$.
\mathcal{B}	The belief of a cluster, the product of the factors in the cluster and all incoming converged messages.
\mathbf{b}	Vector used to denote linear relationship between Gaussian densities.

C	A random variable.
$C(\omega)$	The Schuster periodogram, defined as $C(\omega) = \frac{R(\omega)^2 + I(\omega)^2}{N}$, with $R(\omega)$ and $I(\omega)$ respectively the real and imaginary parts of the Fourier transform.
\mathcal{C}	A Gaussian probability density represented in canonical form with parameters K, \mathbf{h}, g . $\mathbf{x} \sim \mathcal{C}(K, \mathbf{h}, g)$.
\mathcal{C}	A cluster from a cluster graph.
c_o	The speed of light in air.
D	The set of data measured at discrete points in time, $\{d(t_1), \dots, d(t_N)\}$
$\overline{d^2}$	The mean of the squared data, $\frac{1}{N} \sum_{i=1}^N d(t_i)^2$
d_{KL}	Kullback-Leibler divergence.
$e(t)$	Function of the noise values present in the measured signal.
$E[]$	The expected value function.
\mathcal{F}	A sufficient statistic function.
f_t	Transmitted frequency.
f_r	Received frequency.
f_D	Doppler shift in frequency.
\mathcal{G}	A PGM graph.
$G(\cdot)$	Signal model component, takes the time-step and model parameters as input arguments.
g	Scalar used to describe Gaussian density in Canonical form.
\mathcal{H}	Measurement linear relation matrix.
H	A discrete hypothesis space.
$H(\cdot)$	An orthonormalized model component. Orthonormalized from the $G()$ components.
h	A continuous hypothesis space.
$h(\cdot)$	Projection of the data onto the orthonormalised model function $H()$. Defined in Equation 3.5.4.
\mathbf{h}	Vector used to describe Gaussian density in Canonical form.
$\overline{h^2}(\{\omega\})$	The mean-square of the m projections of the data onto the orthonormalized model components.
I	Explicit and implicit prior knowledge in a Bayesian analysis.
$I(\omega)$	The imaginary part of the Fourier Transform of a signal.
i, j, k, l	Indexing values.
\mathcal{J}	The total number of targets for multitarget tracking.
K	Matrix used to describe Gaussian density in Canonical form.
\mathfrak{K}	The number of values that a discrete random variable can assume.

$L()$	A likelihood function.
M_j	The j th signal model function.
m	Total number of signal model component.
\mathbf{m}	The number of discrete random variables in a hybrid factor.
m_{mass}	Probability mass of a probability density.
N	The number of samples in a set of data.
\mathcal{N}	A Gaussian probability density represented in covariance form with mean and covariance matrix, $\boldsymbol{\mu}$ and Σ respectively. $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$.
Nb_i	The neighbouring clusters of cluster i .
\mathbf{n}	The number of random variables in a factor.
\mathbf{P}	The total state-space.
P	A state-space.
$p(\cdot)$	Probability density function.
\mathcal{Q}	Covariance matrix of process noise.
$R(\omega)$	The real part of the Fourier Transform of a signal.
r	Total number of Θ parameters.
S	A random variable of the active signal segment.
\mathbf{S}	Set of segment discrete random variables $\mathbf{S} = \{S_0, S_1, \dots\}$.
$\mathcal{S}_{i,j}$	A sepsset, consisting of the subset of random variables shared between two clusters $\mathcal{S}_{i,j} \subseteq \mathcal{C}_i \cap \mathcal{C}_j$.
T	The last time-step.
t	Time.
t_i	The i th time-step.
\mathbf{u}	Measurement noise.
\mathbf{v}_r	Radial speed.
w_i	The weight of the i th Gaussian probability density function.
\mathbf{w}	Process noise.
\mathbf{X}	A set of random variable of the state-space $\{X_0, X_1, \dots\}$.
\mathbf{Y}	A set of random variables $\mathbf{Y} = \{Y_0, Y_1, \dots\}$
$y(t)$	Function of the true signal.
α	The linear chirp parameter of a sinusoid.
β	The quadratic chirp parameter of a sinusoid.
Γ	The Gamma function.
Θ	A subset of parameters for parameter estimation.
θ	Angle.
$\Delta \mathbf{a}$	Difference between two sets of \mathbf{a} .

$\Delta\omega$	Difference between two sets of frequency parameters ω .
$\delta_{i \rightarrow j}$	A message sent from cluster \mathcal{C}_i to cluster \mathcal{C}_j .
Λ	Matrix used to denote linear relationship between Gaussian densities.
λ	The damping factor used in message passing.
μ	The mean vector of a Gaussian density.
Σ	The covariance matrix of a Gaussian density.
σ	The standard deviation of the noise present in the signal.
Φ	A factor in a cluster graph.
ϕ	A place-holder value that is used when a target is not present.
Ψ	The set of all factors in a graph $\Psi = \{\Phi_0, \Phi_1, \dots\}$ containing a conditional probability density each.
Ω	A set of parameters for parameter estimation.
ω	A frequency parameter, measured in cycles/sample.
$\boldsymbol{\omega}$	Set of frequency parameters $\boldsymbol{\omega} = \{\omega_0, \omega_1, \dots\}$.

Chapter 1

Introduction

Signal processing deals with the analysis of information-bearing signals that often have complex characteristics. The time-varying behaviour in these signals can develop in ways that are not well understood or can be difficult to model. A typical signal-processing problem is that one has a set of noisy measurements and wishes to refine them, or estimate the hidden parameters of an underlying model. Being able to incorporate different types of knowledge into one's signal model can allow one to extract more useful information and can benefit from a probabilistic and systematic approach.

1.1 Motivation and Research Problem

The problem of interest involves measured Doppler radar signals measured for golf players' club swings. The Doppler radar frequency shifts relate to the movements of physical objects which we wish to estimate using a probabilistic approach to signal processing. The sampled signal is noisy with multiple objects present in the same measurement, the objects and received signals interact in complex ways, there are multi-path signals present, and the signal is sometimes clipped causing harmonics to appear. Smoothing the frequency shifts of the Doppler signal will allow more accurate estimates of the speeds of the physical objects of interest. Accurately estimating the speeds is a step towards estimating the velocities of the objects such as the club and ball and can allow one to calculate their trajectories, as their starting points are known. This information is invaluable to golf players and coaches, who will be able to improve players' skills based on knowledge about club velocity at impact, the ball spin, and other properties of interest of the golf swing. In order to smooth the signal, multiple objects present in the signal such as the club-head, the ball or the divot need to be tracked. Figure 1.1 contains a spectrogram of an example golf swing's Doppler radar data. The spectrogram illustrates the multiple frequency components present in the Doppler radar signal with some frequency components that are weaker than others and barely visible.

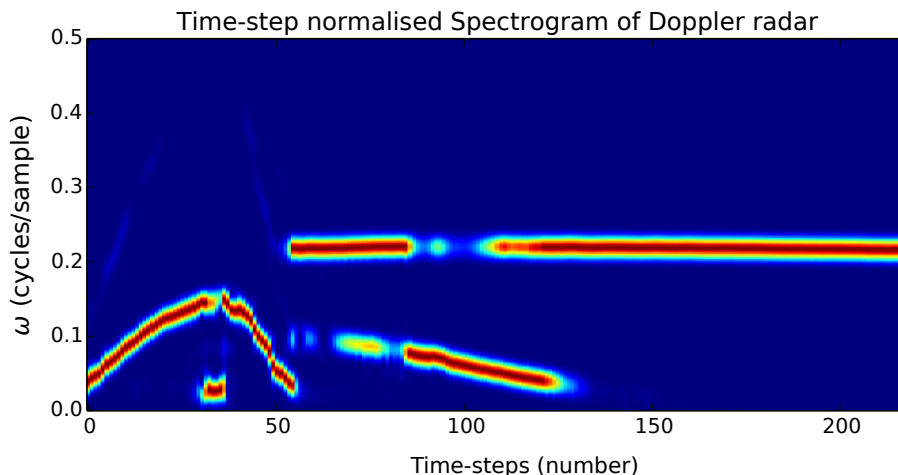


Figure 1.1: *An example spectrogram of a golf swing’s Doppler radar data, which contains multiple frequency components.*

Being able to estimate our objects of interest’s frequency parameters can be formulated as *multitarget tracking*, where the number of targets and their states are jointly or separately estimated from sensor data using *state-space estimation* [1]. In our application the Doppler frequency shifts are the targets being tracked. Many of the difficulties of multitarget tracking are related to handling the uncertainty around the measurement noise, process noise, data association between targets, missed detections, false alarms, and unknown birth and death rates of targets.

When tracking single or multiple targets using noisy measurements it is difficult to model the entire signal, particularly if the target behaviour is complex. It can often be the case that the local interactions and behaviour over short time-periods of a signal are better understood, which makes piecewise modelling an appealing alternative. Quantifiable domain knowledge about the targets’ behaviour is also sometimes available, but it is not always certain how to meaningfully incorporate it.

Our approach will be to model the multitarget tracking in the context of Bayesian probability theory, and then to solve the marginal posterior distributions of the parameters of interest using probabilistic graphical models (PGMs). We compensate for uncertainty in the characteristics of our measured Doppler radar signal by breaking it up into time-steps and modelling the short-time signal, local signal behaviour, as well as the overall trend of the signal by grouping time-steps into signal segments. These signal segments correspond to different parts of the physical golf swing that contain a different number of objects’ Doppler shifts and different signal dynamics. PGMs are well suited to this modular approach and provide the benefit of decon-

structuring the problem at hand into a set of local dependencies. PGMs are often used in machine learning problems where they show promise as seen by the improvements made - for example, in speech-recognition where deep belief networks are being used [2]. In conjunction, we will use a Bayesian statistical method called Bayesian spectrum analysis (BSA) to analyse the Doppler signal in each time-step. Bayesian statistics provide a powerful way of including prior information explicitly and have been successfully applied by using BSA in astrophysics [3, 4], target identification using radar [5], and frequency detection in biological time series [6].

1.2 Background

1.2.1 Bayesian Spectrum Analysis

The spectrogram is a useful signal-processing tool that provides frequency information by analysing short time-windows of a typically time-varying signal. For each time-window, the periodogram is calculated over short, consecutive and often partially overlapping signal windows to produce the spectrogram. The discrete Fourier transform (DFT) is used to calculate the periodogram. If the signal is semi-stationary for short time periods, the implicit assumptions made by using the periodogram to find frequencies hold and provide meaningful and accurate information for the spectrogram. Using Bayesian probability theory, Jaynes [7] has shown that a scaled version of the periodogram gives the ideal frequency estimate of discrete data containing a single sinusoid in white noise. This gives a probabilistic justification for using the periodogram. Jaynes also developed the *chirpogram* which estimates both the frequency and chirp-rate of a chirped sinusoid signal [7]. Working on the foundation of Jaynes' work, Bretthorst [8] developed a method, BSA, that generalises estimating the parameters of a signal given a signal model. BSA specifies a customisable signal model that can contain multiple parameters, such as the change in frequency or signal amplitude decay, and calculates their joint probability density given the data. Bretthorst shows that the periodogram's implicit assumption of a single sinusoid in white noise does not hold for more complex, time-varying signals. The periodogram and DFT, however, still prove useful in many cases, as is evident from its continued wide-scale use in engineering and other applications.

As an example, the probability of the frequency of a single sinusoid with white noise, as shown in [8], is

$$p(\omega|\sigma, D) \propto \exp\left\{\frac{C(\omega)}{\sigma^2}\right\} = \exp\left\{\frac{|DFT|^2}{N\sigma^2}\right\},$$

with ω the frequency in cycles per sample, σ the standard deviation of the noise, D the data that we have measured with N samples, and $C(\omega)$ the

periodogram. BSA gives one the tools to create more complex signal models and to include prior information about the parameters contained in these signal models. Concepts such as noise, which is not present in the DFT, can now be incorporated probabilistically.

1.2.2 State-Space Estimation

State-space estimation models are well-known techniques that include temporal domain knowledge by building it into the models. Among these are the hidden Markov model (HMM) [9], where a set of discrete state-spaces are estimated at discrete time-steps, and Kalman filters [10], where the continuous state-space at each time-step is modelled using Gaussian densities with linear relationships between parameters in each time-step. In [11], Stone formulates the problem of multitarget tracking in terms of Bayesian probability theory. State-space estimation has also been formulated in terms of PGMs such as in [12, 13, 14].

1.2.3 Probabilistic Graphical Models

PGMs include well-known examples such as HMMs [9], Bayes networks [15], Markov random fields [16], and factor graphs [17] and have found use in fields as diverse as speech recognition, image classification, and simultaneous localisation and mapping [14]. An advantage of the PGM framework is the ability to break a problem into smaller, more manageable parts. Instead of solving the joint probability density or distribution of a model, the PGM solves the marginal distributions of the random variables allowing otherwise intractable problems to be solved. Some PGM algorithms only solve the most likely values of the model parameters. The piecewise modular approach absolves one of modelling the entire system at once. A PGM inference algorithm, *belief propagation* (BP), gives an exact solution to the marginal distribution of *clique trees*, an undirected PGM with no loops [14]. *Cluster graphs*, clique trees containing loops, can be solved using a variation on BP called *loopy belief propagation* (LBP), which gives an approximate solution [14].

1.3 Objectives

With regard to the measured Doppler radar signal of golf swings, the objectives of this study are to:

- Model the characteristics of short time-steps of the Doppler signal and estimate the spectral parameters of the model in a probabilistic manner.
- Compare competing models of the short time-steps of the Doppler signal in order to select the most probable model from a list of models.

- Use a PGM to incorporate knowledge of the physical behaviour of the signal to better estimate the signal behaviour within a given segment of the golf swing.
- Model the changing dynamics and number of frequency components present in the signal.
- Do the above with the aim of smoothing the estimates of Doppler frequency shifts.
- Compare different implementations, namely between a discrete PGM and a hybrid discrete-Gaussian PGM.

1.4 Outcomes and Contributions

The objectives of the research were met by employing the methods and approach set out below. These methods were used on simulated data and on parts of measured Doppler radar data of golf swings where multiple objects' Doppler frequency components were smoothed:

- We analyse the Doppler radar data that has been segmented into parts that typically appear in the signal (such as the downward swing of the club or the flight of the ball) using BSA to gain a better understanding of the behaviour of the frequency components.
 - (a) The BSA models used had up to three frequency components and up to three frequency parameters (frequency, linear chirp, and quadratic chirp) making a total of nine different models. We explicitly model compound signals containing multiple frequencies with behaviour more complex than only stationary frequencies. The models optionally contain the noise present in the signal as a random variable.
 - (b) The models are made comparable using Bayesian probability theory.
 - (c) The more complex BSA models are selected more often than an inspection of the spectrogram would suggest. We hypothesise that there is more complexity present in the Doppler radar signals than our BSA models can explain.
 - (d) The above results are congruent with simulated results where multiple, close frequency components can have a single peak in the DFT but BSA can detect the correct number of frequency components.
- BSA is used to model the Doppler radar signals for short time-periods of the signal, similar to a spectrogram, in order to estimate the frequency parameters with no information being shared between time-steps. These analyses will form the basis of our PGMs and will be analogous to state-space estimation observation likelihoods in Kalman filters.

- (a) Analysing shorter parts of the signal increases the likelihood that the less complex BSA models are selected and match what we would expect from the spectrogram. The improvement is to be expected, as fewer non-linearities are likely to be present in shorter parts of the signal.
 - (b) We include the energy of the signal noise as a parameter in the BSA models to compensate for the difference between the underlying Doppler shift behaviour and our BSA models. This noise parameter is adjusted by inspection to increase the accuracy of our BSA model selection. This is done in order to increase the likelihood of selecting the correct number of frequency components.
 - (c) We hypothesise that our understanding of the underlying Doppler radar signal is incomplete. Despite this, BSA still gives meaningful results and it still proves useful in making comparisons between different parts of the signal, as can be expected thanks to the relationship between BSA and the DFT.
 - (d) We use BSA not only to give us an estimate of the parameters that we are modelling but also to determine the full probability density function. This is crucial, as it enables us to use BSA in a probabilistic framework such as PGMs.
- PGMs are designed that model the time-varying properties of the Doppler radar signal within a segment of a golf swing.
 - (a) We smooth the frequency probability density functions calculated with BSA for each time-step using a discrete PGM and a continuous PGM. This can be seen as a state-space estimation smoothing of the signal.
 - (b) This approach resembles other multitarget tracking methods.
 - We include a discrete Markov chain which runs parallel with the state-space estimation in our PGMs, thus giving a weighted quality to the underlying analyses. Each state of the Markov chain describes a different segment of the golf swing (for example, the downward swing of the club) and therefore allows the system to decide in which segment of the swing it is for each underlying BSA analysis. Each segment of the signal has a relevant BSA hypothesis and signal dynamics.
 - (a) We show that a simple multitarget state estimation, without information about segment progression, is not enough for the PGM to select the correct BSA models and smooth the Doppler radar data. Without specifying a strong prior probability distribution on the number of targets or their dynamics, the underlying BSA model selection is not sufficiently accurate for the PGMs to select the correct BSA model.

- (b) Enforcing a left-to-right progression of the signal segments in the discrete Markov chain, similar to a HMM, improves the PGM and allows it to select the correct BSA model more often. This differs from the typical multitarget tracking methods, which encode target births and deaths instead of this segment state progression. The PGM now only has a limited number of segment changes available and is better able to select the correct signal dynamics. Where there is insufficient or incorrect information about a target's Doppler frequency from the BSA analysis, the PGM is now better able to calculate the position based on the probability density functions of the surrounding time-steps. For even more complex parts of the Doppler radar signal, such as during the club impact, the left-to-right progression is not sufficient for correct signal segment selection without more temporal information or the correct model selection of BSA.
 - (c) We include a "missed target" model, which allows our PGMs to smooth over parts of the signal where the BSA models incorrectly fail to detect a frequency component in the signal.
 - (d) We hypothesise that the lack of explicit duration modelling of the signal segments is detrimental to estimating the frequency parameters.
- The above PGMs were implemented using both a fully discrete model, where the continuous BSA parameters were discretised, and a hybrid network where the parameters were approximated as multivariate Gaussian density functions.
 - (a) We find that the discrete model allows more flexibility than the Gaussian hybrid network in the design of the conditional probability density functions. One is not restricted to one type of distribution, but only the discretisation of the distribution as one can approximate arbitrary probability density functions.
 - (b) Evenly discretising the BSA results can cause us to miss the sharp peaks, which leads to problems in estimating the relative probabilities between the different BSA models. We use Laplace's method to approximate integrating the model parameters out as a Gaussian probability density function integral. This allows us to avoid numerical integration and is a better approximation of the relative model probabilities.
 - (c) We find that, as expected, the discrete PGM is significantly slower computationally than the hybrid network. The number of parameters increase exponentially with the number of random variables, and this leads to restrictions on the resolution of the discretisation.

- (d) The hybrid model is unforgiving when the process noise is not correctly estimated. In the discrete model one can model zero process noise, whereas this leads to singularities in the case of the hybrid model.
- (e) The Gaussian is uni-modal and one needs a multi-modal probability density function in order to model the joint probability of multiple targets' parameters without associating the parameters with a target identity. If there is no target identity, the joint probability densities are symmetrical with multiple peaks. We therefore used an implicit target identity where frequency components were arbitrarily assigned to the random variables in the Gaussian from the highest to the lowest frequency.
- (f) There are different limitations of the discrete and discrete-Gaussian hybrid PGMs, and there is a trade-off between accuracy and computational tractability. Both were sufficient to explore the concepts presented in this work, with the hybrid network more likely to be practically applicable in a portable Doppler radar unit.

1.5 Overview of this Work

The approach in this work consists of three levels, namely the bottom-level BSA, the mid-level temporal state-space estimation that refines the signal parameters and can reason about competing hypotheses, and the top-level signal segment estimation describing which mid-level behaviour is active for a given time-step. With BSA we estimate the frequency parameters of the Doppler radar signal for short time-steps, similar to a spectrogram. However, we use different competing hypotheses that have relative probabilities for the signal. The other two levels are set up using prior knowledge about the local and general signal behaviour.

1.5.1 Physical Signal and Setup

In Chapter 2 we discuss the physical setup and signal. The underlying signal is measured by a Doppler radar unit placed behind a golf player as the player swings a golf club and hits a golf ball. The Doppler radar unit emits a sinusoidal radar signal of an effectively constant frequency and measures the reflected signal. The shift in frequency of the received signal relative to the transmitted signal, the Doppler shift, is proportional to the speed of the object from which it is reflected. Measuring the Doppler shift can thus give us the radial velocity of objects relative to the Doppler radar unit. There are multiple reflected sinusoids present in the signal that are reflected from different objects or different parts of the same objects. The reflected sinusoids are distorted by the shape of the object, change frequency over time, have non-linear changes

in behaviour, and can be the result of unwanted targets such as the player's legs moving, different parts of the golf club, or harmonics caused by signal clipping. The challenge is therefore to correctly estimate the Doppler shift despite these unwanted effects.

1.5.2 Bayesian Modelling of the Underlying Signal

1.5.2.1 BSA

In Chapter 3 we discuss Bayesian parameter estimation and model selection and present Bretthorst's results for BSA in the general case. One of the interesting results of BSA, as found in [7] by Jaynes, is that the periodogram is a sufficient statistic for finding the frequency of a single sinusoid with white noise. We discuss how BSA allows for other signal models such as multiple frequency components being present or sinusoids with non-stationary frequencies.

1.5.2.2 Experimental Results

Our experimental results using simulated data and the measured Doppler radar signals are presented in Chapter 4. We find that BSA performs well on the simulated signals with added white noise for both parameter estimation and model selection. In order to analyse the Doppler radar signals, we segment them into parts based on the number of frequency components present and their dynamics. Figure 1.2 contains an example of a segmented Doppler radar signal.

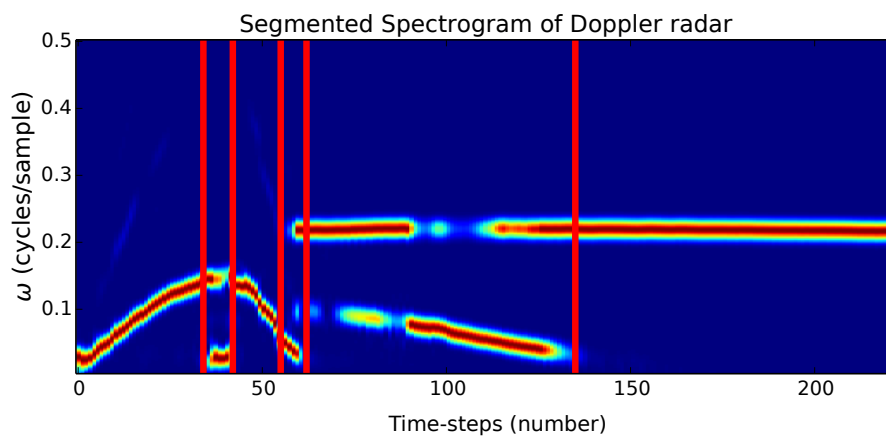


Figure 1.2: An example of a segmented spectrogram of a golf swing's Doppler radar data. Each segment has either different frequency components or frequency dynamics.

Using stationary frequency models, the parameter estimation of simpler parts of a Doppler radar signal gives results consistent with an inspection of the spectrogram of the signal. The more complex signal segments do not fit the models as expected and there seems to be more complexity than our current BSA models can explain. This is likely due to the non-linear signal frequency behaviour, the indirect measurement of the target, or other effects in the signal we do not yet model. To help reduce the effect of the non-linearities and to better capture the transient nature of the signal, we split the segments into smaller windows for analysis such as is done in a spectrogram. For this analysis we do not have much prior knowledge about the error in our signal model and thus we cannot rely heavily on the sharpness of the peaks of the parameter estimation results. BSA does not have the benefit of seeing the data that is contained in the other time-steps using this strategy as-is. We suggest that modelling the signal as multiple sinusoids with linear or quadratic chirp rates is not enough to explain the complexity of the signal. This is likely due to a single target reflecting a range of different frequencies that are close to one another that relate to the different velocities of its different parts. Our assumption of a maximum of three well-separated signals is an approximation of the more complex underlying physical signal.

The model selection favours the more complex models, as they tend to explain the data better despite the penalties Bayesian probability theory imposes for having more parameters. This is not a failure of BSA but a case of our assumptions about the signal being too strict. One can attempt to model these multiple, complex frequencies or somehow measure or calculate the error one makes in using the simpler models. We manually increase the error factor in our models for better model selection results. This could be automated in future work. There tends to be a single strong frequency or group of very close frequencies created by the object that we are interested in. If this were not so then the spectrogram would not be as useful as it is in identifying the objects present.

1.5.3 Refinement of the BSA Results

1.5.3.1 Bayesian State-Space Estimation

In Chapter 5 we frame the problem of state-space estimation as a problem of Bayesian inference. Stating the problem within the Bayesian framework allows us to use PGMs to solve the marginal posterior probabilities of the states we wish to estimate. We briefly discuss three methods used in literature, namely switching state-space estimation [18], track-oriented multiple hypothesis tracking [12], and joint multitarget probability density [19].

1.5.3.2 PGM Theory

In Chapter 6 we give background information regarding the PGM approach and the type of models we use. We use both a fully discrete and a hybrid PGM that are made up of different types of factors. We implement the factors in cluster graphs [14], which are undirected PGMs. The discrete factors in our graphs take the form of a key-value paired table with the values that random variables can assume as the key, and their probabilities the values. The continuous factors are Gaussian probability density functions that are manipulated in canonical form. The hybrid factors are a mixture of the discrete and continuous factors; this allows for a mixture of discrete and continuous random variables used within the same factor. They are described by a table similar to the discrete factors where, instead of a probability for each key-value pair, there is a Gaussian probability density function. When one uses linear Gaussian probability density functions, these hybrid factors are called conditional linear Gaussians (CLGs) [14]. We describe LBP and loopy belief update (LBU), two similar algorithms that both allow the factors to share knowledge about the random variables that they have in common. Using these algorithms, one can calculate the marginal probabilities of the joint probability density function that a graph factorises.

1.5.3.3 PGM Model Design

Chapter 7 contains our modelling of the Doppler radar signal behaviour. Figure 1.3 is a Bayes network representation of the lower- (green), mid- (blue), and higher-level (red) transitions we used to describe the different sources of domain knowledge we incorporated into our PGMs. The model is only used conceptually in order to create the factors needed to implement the PGMs using cluster graphs. In our PGMs the three levels of knowledge we used were:

- The different BSA estimations of the parameters at each time-step (lower-level, indicated in green).
- The behaviour of the physical targets' motion, which is described by the parameter dynamics (mid-level, indicated in blue).
- The fact that each discrete signal segment occurs in a certain order and other temporal information about signal segments (high-level, indicated in red).

1.5.3.4 PGM Experimental Results

In Chapter 8 we discuss the challenges faced and lessons learned in this thesis. The challenges include selecting process noise for the hybrid model, the effects of discretisation for our discrete model, and non-convergence of some

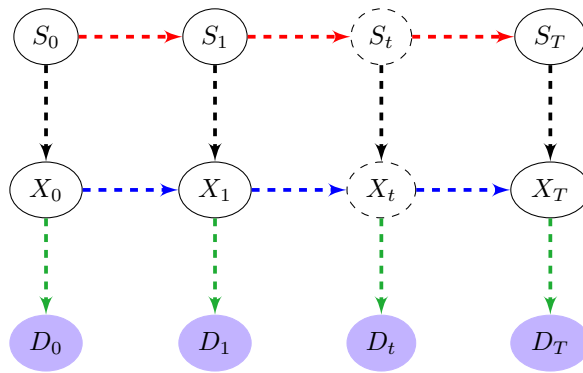


Figure 1.3: *A Bayes network representation of the general model approach.*

of the PGMs. We also compare the hybrid and discrete models where we find that the discrete model offers more flexibility in creating factors, but that the hybrid model is much less computationally expensive.

Chapter 9 contains examples of the results of our PGM models on Doppler radar data. The experiments include intra-segment parameter smoothing for different signal models, inter-segment transitions, and a missed target BSA model. Restricting parts of the signals' frequency component dynamics can improve the results. For certain parts of the signal we needed to specify the signal segment transitions where for other parts the models were able to automatically select the segments. We find that the missed-target model and left-to-right segment transition improved upon the conventional multitarget tracking and allowed the PGMs to select the correct signal segment and to smooth over regions where a frequency component was missing.

1.5.4 Conclusions and Recommendations

In Chapter 10 we discuss the conclusions reached from our research, summarise the results of this thesis, and make recommendations for future work. We find that both the discrete and hybrid models were sufficient for exploring the concepts presented in this thesis. The models allowed us to incorporate our domain knowledge and improve our noisy measurements of the frequency components present in the signal. Our recommendations include using different BSA models, exploring the signal noise further in the BSA models, finding better methods for selecting the PGM process noise, exploring other methods for factor representations, and using explicit duration models.

Chapter 2

Physical Setup

2.1 Introduction

The Doppler effect has found many practical uses in engineering and the sciences whereby one can calculate the relative speed of objects from the frequencies of reflected or emitted electromagnetic waves. In this chapter we will discuss the physical setup of the Doppler radar unit that produces the radar signal and measures the reflected Doppler shifted signals. The Doppler radar unit is used to measure the velocity of a golf player's swing and of the golf ball.

2.2 Doppler Shift in Radar

The Doppler effect causes a difference in frequency from the transmitted signal f_t and the signal reflected from a moving object with a resulting frequency of f_r . We can calculate the the reflected signal's frequency as

$$f_r = \left(\frac{1 + \mathbf{v}_r/c_0}{1 - \mathbf{v}_r/c_0} f_t \right)$$

where the radial velocity of the target object, \mathbf{v}_r , is relative to the source of the transmitted signal, and c_0 is the speed of light in air [20]. For an approaching target, \mathbf{v}_r is positive, and it is negative for a receding target.

The difference between the transmitted frequency f_t and the received frequency f_r is the Doppler shift, f_D :

$$f_D = f_r - f_t = \frac{2\mathbf{v}_r}{c_0 - \mathbf{v}_r} f_t.$$

Since the speed of light in air is much greater than the radial speed of the object, $\mathbf{v}_r \ll c_0$, we can simplify the above equation to obtain

$$f_D \approx \frac{2\mathbf{v}_r}{c_0} f_t.$$

If the target is moving with speed \mathbf{v} and an angle θ with respect to the radar line-of-sight, where $\theta = 0^\circ$ for a directly approaching target and $\theta = 180^\circ$ for a receding target, the radial velocity, \mathbf{v}_r , will be projected as $\mathbf{v} \cos \theta$. The Doppler shift is thus

$$f_D \approx \frac{2\mathbf{v}}{c_0} f_t \cos \theta,$$

which is proportional to the radial velocity of the target object [20].

2.3 Radar Unit

We have data from a Doppler radar unit with the unit setup as seen from the side as shown in Figure 2.1. The ball is placed 3m in front of the unit, with the ball trajectory away from the unit. The Doppler radar unit emits a sinusoidal signal with a near constant frequency in the range 10.5 - 10.6 GHz. The unit receives the transmitted signal's reflections from objects. This received signal is mixed down with the transmitted signal and passed through a low-pass filter. The resulting compound signal contains the absolute frequency shifts of the received signal relative to the transmitted signal's frequency.

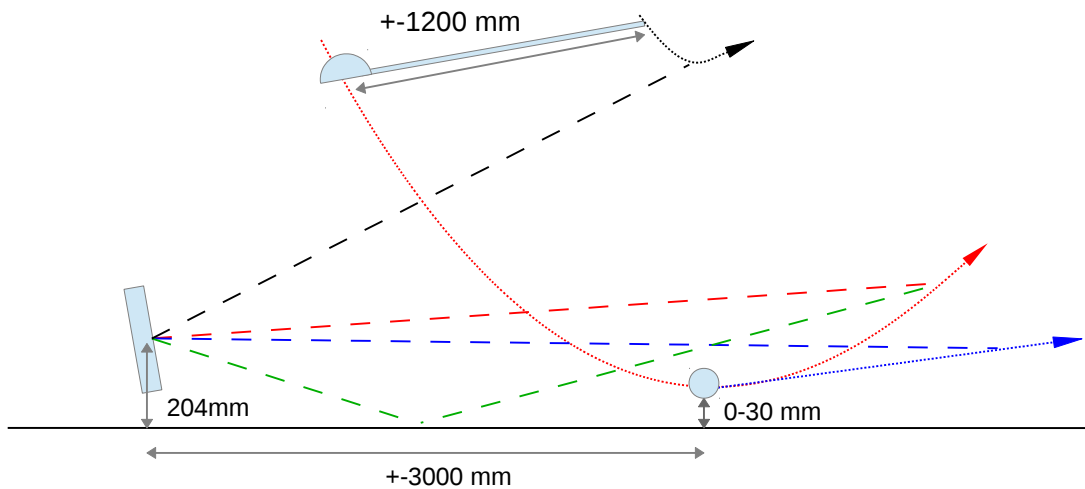


Figure 2.1: *The setup of the Doppler radar unit in the vertical plane (not to scale). The trajectories over time of the club-head, ball and player are indicated by the red, blue and black dotted directed lines respectively. Example radar signal paths from the objects to the Doppler radar unit are indicated by dashed lines in their corresponding colours. The green dashed path is a possible alternative path that a reflected signal can follow for the club-head.*

In Figure 2.1 the object trajectories that create Doppler shifts are indicated with dotted directed lines with the club in red, the ball in blue and the player body in black. Example paths of the reflected signals caused by the objects are shown by the dashed lines. The dashed green line indicates a possible *multi-path* signal where a signal can be reflected against the ground before being measured.

2.4 Example Doppler Signal Spectrogram

Figure 2.2 is a spectrogram of a Doppler radar signal of a golf swing such as in Figure 2.1. It shows the resulting absolute Doppler frequency shifts caused by the movement of the target objects. For each time-step, a window of 64 samples, multiplied with a Hamming window, is applied with a step-size of 16 samples between windows. We used the Hamming window to suppress the side lobes of the periodogram which would otherwise be clearly visible in the log scale. We do not use any window functions other than a flat window in the rest of this thesis. Superimposed over the spectrogram are dashed lines indicating which objects caused the frequency shifts, namely the ball, club-head, the stick-flash phenomenon, and an example of a harmonic caused by clipping. Other objects visible but not indicated are low-speed objects such as the body of the player or other low-frequency noise components. These targets of interest are explained in more detail in the following sections.

2.4.1 Low-frequency Components/Noise

There are low-frequency components present during the typical golf swing which are clearly visible in the log spectrogram in Figure 2.2. The effect is most pronounced during the club's downward swing between time-steps 60 and 80. They are not as visible to this degree when the log of the spectrogram is not taken.

2.4.2 Ball

The ball component has a relatively constant frequency compared to the other components. Outside of the scope of the spectrogram, the frequency component of the ball slowly drops to zero.

2.4.3 Club-head

The club is the object with the most complex behaviour where multiple parts of the club cause reflected signals, most notably the club-head and the shaft. The discontinuity around time-step 80 is caused by the momentum transfer from the club-head to the ball during the golf-swing impact. The club-head

reaching zero Hz/sample near time-step 40 is when its velocity is perpendicular to the radar reflection path at that time-step. Before time-step 40, the Doppler shift of the club-head is negative as it is moving towards the Doppler radar unit – however, it appears positive on the spectrogram as it displays the absolute frequency shift. The club-head is one of the strongest components, with the stick-flash phenomenon sometimes being stronger. The high amplitude of the club-head frequency component causes clipping of the measured signal, which can cause harmonic frequencies to appear such as during time-steps 40 to 60.

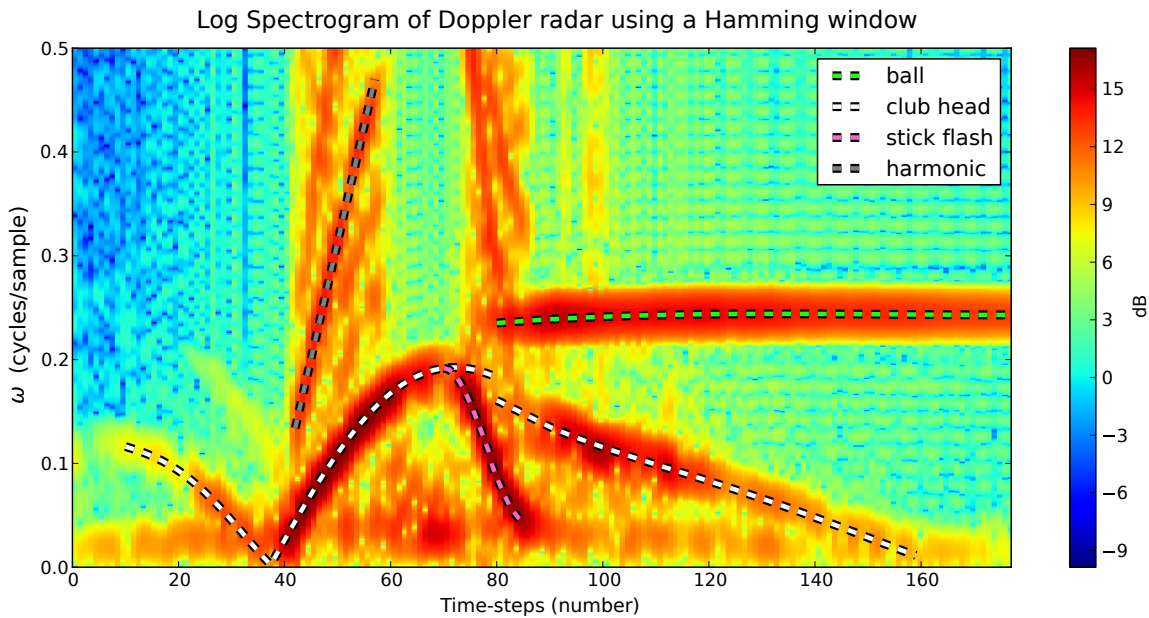


Figure 2.2: *The spectrogram of a golf swing indicating the absolute frequency shift from the transmitted signal which is proportional to the radial velocity of the object. Imposed on the spectrogram are the frequency paths of the ball, club-head, stick-flash (the radar reflection running up the shaft) and a harmonic frequency caused by the clipping of the signal. There are also low-frequency components created by the player motion or other unwanted effects.*

2.4.4 Stick-Flash Phenomenon

The stick-flash phenomenon occurs when the club shaft moves into a position perpendicular to the face of the radar unit, creating a larger target profile and consequently a higher amplitude reflected signal. This part on the club shaft, which has the largest target profile relative to the radar unit, moves along different parts of the shaft for a short time before, during, and after the moment of impact. There is a discontinuity in the stick-flash at the moment of impact at time-step 80, which is similar to the club-head discontinuity. This target also produces a strong frequency component that obscures the club-

head at the moment of impact and clips the signal, thus causing harmonics to appear.

2.4.5 Signal Clipping

Signal clipping occurs when the measured signal amplitude is larger than the measuring equipment or the digital representation's maximum amplitude. Figure 2.3 and 2.4 both contain the normalised histograms of the measured signal amplitudes of different parts of the same signal from Figure 2.2. Figure 2.3 contains a part where the signal clipped and Figure 2.4 contains a part where there was no clipping. We normalised the amplitudes to the maximum measured amplitude for both the clipped and non-clipped parts. One can see the effects of clipping in Figure 2.3 where the (now normalised) maximum amplitude is significantly more prevalent than in Figure 2.4. The signal clipping causes multiple higher harmonic frequency components to appear in the spectrogram such as shown in the spectrogram in Figure 2.2.

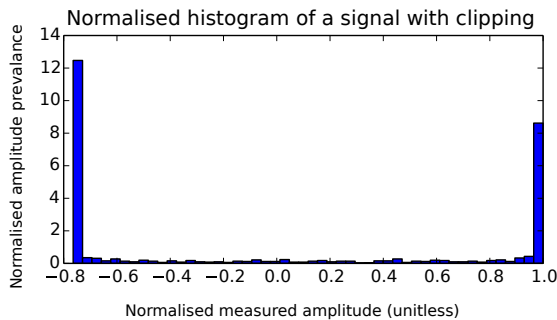


Figure 2.3: *Normalised histogram of a Doppler radar signal part that contains clipping.*

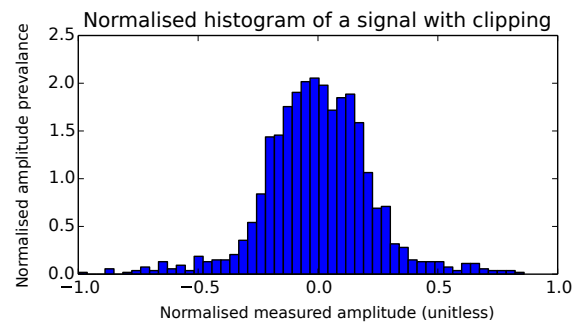


Figure 2.4: *Normalised histogram of a Doppler radar signal part that contains no clipping.*

2.5 Conclusion

In this chapter we discussed the physical setup that measures the Doppler radar signals. The spectrogram is very useful for inspecting the reflected, measured signal as the relative speed of the targets of interest is proportional to the Doppler shift of the reflected sinusoids relative to the transmitted sinusoid. It does not, however, account for signal noise or an explicit signal model. We also discussed effects in the signal that obscure our targets of interest, namely the stick-flash phenomenon, multi-path sinusoids, and clipping. In this thesis we will focus only on estimating the radial velocity of the objects which we compare to the spectrogram.

Chapter 3

Bayesian Spectrum Analysis

3.1 Introduction

Bayes' theorem can be used as a method of inference. It allows one to update one's prior knowledge or "belief" about a parameter of interest with a *likelihood* function which incorporates observed data. The parameters used in the probability calculations are called *random variables*. Bayesian probability theory is made possible by the product and sum rules and can be applied to concepts such as nuisance parameters, parameter estimation, model selection and selecting prior probabilities [8]. In this chapter we will discuss Bayes' theorem and introduce the theory behind Bayesian spectrum analysis (BSA). We also demonstrate the relation between the periodogram and BSA for a single sinusoid in white noise. The general theory discussed here will be applicable to the rest of this thesis, and specifically the BSA will form the lower level of our PGMs.

3.2 Bayesian Probability Theory

In Bayesian probability theory one makes use of Bayes' theorem to update one's prior state of knowledge by incorporating data using a statistical model. This is done by using the product,

$$p(A, B|C) = p(B|A, C)p(A|C), \quad (3.2.1)$$

and sum rules,

$$p(A) = \sum_B p(A, B), \quad (3.2.2)$$

of probability theory. The product rule allows us to relate random variables to each other and the sum rule allows us to systematically remove them from consideration by *marginalising* them. Marginalising random variables is done by summing the probabilities of the unwanted discrete random variables and

integrating out unwanted continuous random variables.

We have a hypothesis or hypothesis space, H , under consideration and we wish to update our belief about this hypothesis with the data we have collected, D . One can perform the necessary probability calculations on the random variables using the above product and sum rules. To calculate the probability density of the hypothesis space using Bayes' theorem, we have:

$$p(H|D, I) = \frac{p(D|H, I)p(H|I)}{p(D|I)}. \quad (3.2.3)$$

In this chapter we emulate Jaynes [7] and Bretthorst [8] by adding I to the right of the conditioning bar “|”. Jaynes used this parameter to explicitly denote that there is certain information that we take as given when setting up an experiment. Examples of this, sometimes implicit, information include the class of alternative hypotheses being considered, the statistical model which denotes the relationship between hypotheses, and the data and other information we might have about the data or problem domain [7]. Our prior probability for the hypothesis is $p(H|I)$. This is our knowledge about the hypothesis space H given the knowledge of the problem setup or other domain knowledge contained in I before seeing the data. For a fixed value of our data, D , the function that we use to update our state of knowledge about the hypothesis is the likelihood function or statistical model $p(D|H, I)$, the probability of the data given the hypothesis. This function is what allows us to update our prior knowledge with the data that we have observed. When H is fixed it is known as the sampling distribution. $p(D|I)$ serves as the *normalising* factor and is also known as the *global likelihood*. It can also be written as

$$p(D|I) = \sum_H p(D|H, I)p(H|I).$$

Bayes' theorem gives us the *posterior* probability of our hypothesis given the data $p(H|D, I)$ which is our knowledge about H modified by obtaining new information in the form of the data D .

The Bayesian approach allows us to pose hypotheses such as by Bretthorst in BSA [8]: “What is the probability density of a frequency ω of a sinusoid given the data, independent of the amplitude and phase?” and “Given several possible models of the data, which model is more probable?”. The first hypothesis assumes the model to be correct and is parameter estimation [21], and the latter is model selection [22] where the correct model to be selected is another parameter and is also handled as a random variable by Bayesian probability theory.

3.2.1 Nuisance Parameters

We want to focus our attention on a subset of parameters. Historically, parameters that one did not wish to estimate but were part of the model were referred to as “nuisance” parameters. Using Bayesian probability theory, one can allocate a prior probability density function to them and marginalise them by either integrating or summing them out. The global likelihood is a special case where all parameters are marginalised.

3.2.2 Discrete and Continuous Hypotheses

For a discrete hypothesis space H one has a probability distribution, $p(H_j|D, I)$. Marginalising all the parameters as nuisance parameters, the hypothesis space is constrained to the following, with k discrete hypotheses under consideration:

$$\sum_{i=0}^{k-1} p(H_i|D, I) = 1.$$

For a continuous hypothesis space h one can consider the problem as the limiting case for an arbitrarily large number of discrete hypotheses [23]. The rule for marginalising the hypothesis space as applied above then becomes:

$$\int dh p(h|D, I) = 1$$

for a continuous hypothesis space.

3.2.3 Sufficient Statistic

A function is a *sufficient statistic* of a random variable when, given the data, it summarises all the relevant information for computing the random variable’s likelihood function [14]. From [24] the sufficient statistic is a function where no other function calculated from the same sample would provide any additional information as to what the value of a parameter is. The following has to hold for the function \mathcal{F} to be a sufficient statistic of the parameter ω for the data observed, D :

$$p(D|\mathcal{F}, \omega) = p(D|\mathcal{F}).$$

This indicates that the likelihood function is conditionally independent of the probability of ω , as all the relevant information is already contained in the sufficient statistic \mathcal{F} [24]. An example is a function which calculates the mean of a set of data. It would be the sufficient statistic for the mean μ of a Gaussian probability density function with known variance.

3.2.4 Parameter Estimation

In parameter estimation our hypotheses are the possible values that a given model's parameters can assume. The model under consideration can be implicitly included in I or explicitly stated on the right of the conditioning bar. One aims not only to obtain the most likely values for a set of parameters, but the probability density function of a range of possible values [21].

Rewriting Bayes' theorem where our model parameters are Ω and M_j is the j th model under consideration, we have:

$$p(\Omega|D, M_j, I) = \frac{p(D|\Omega, M_j, I)p(\Omega|M_j, I)}{p(D|M_j, I)}. \quad (3.2.4)$$

We might only be interested in a subset of the parameters $\Theta \subset \Omega$ – then we have:

$$p(\Theta|D, M_j, I) = \int_{d(\subseteq \Omega)} \frac{p(D|\Omega, M_j, I)p(\Omega|M_j, I)}{p(D|M_j, I)},$$

where $d(\subseteq \Omega)$ indicates that a subset of Ω is integrated out.

3.2.5 Model Selection

We can have more than one parameterised model under consideration and wish to compare them. The probability of a model being the true model can be treated as another random variable and we can calculate the probability distribution of the models using Bayesian probability theory. Considering the probability of the j th model M_j from k different models, we can rewrite Bayes' theorem as:

$$p(M_j|D, I) = \frac{p(D|M_j, I)p(M_j|I)}{p(D|I)} = \frac{p(D|M_j, I)p(M_j|I)}{\sum_{i=0}^{k-1} p(D|M_i, I)p(M_i|I)}.$$

The model likelihoods can be calculated by marginalising the parameters of the models:

$$p(D|M_j, I) = \int d\Omega p(D|\Omega, M_j, I)p(\Omega|I)p(M_j|I).$$

According to Occam's Razor [23] as applied to the sciences, if there is uncertainty between hypotheses, the hypothesis with fewer assumptions should be preferred. Bayesian probability has an analogous principle, the Occam factor, in which more complex models are implicitly penalised proportional to their complexity by their prior probabilities [23]. The more complex hypotheses with more parameters or spanning over larger parameter ranges require higher likelihood functions in order to overcome the Occam factor penalty.

It is worth considering that when only comparing the likelihood functions of different models, if one is contained as a special case of the other, then the simpler model will never be favoured. The Occam factor is thus crucial in ensuring that the more complex model is favoured only when its likelihood ratio is large enough to justify the added complexity. This is especially true when we integrate out the parameters of our models, Ω , in order to calculate the model probabilities.

In this work, we will typically assume that we have no information leading to a prior preference for one model over another, so the ratio between each of the prior probabilities will be unity. The determining factor between which model is more likely will be determined by the models' global likelihoods $p(D|M_j)$.

3.2.6 Prior Probability Selection

According to MacKay, one cannot do inference without assumptions [25]. Using Bayesian Probability theory, we can quantify our assumptions using prior probability density functions. If we do not have strong prior information about a random variable, we would preferably keep the prior probability density function as *uninformative* as possible given the knowledge about the random variable.

Uniform Prior: If we know the range of values that a parameter can assume, but not the likelihood of the values within that range, a uniform probability density is appropriate. It denotes ignorance of the value of the random variables bounded within the allowed range of values. The probability density is a constant within the range of interest and zero otherwise. We can have an unbounded uniform prior that forms an *improper prior*, which cannot be normalised [7].

Gaussian Prior: Gaussian priors are typically used for location parameters. A Gaussian prior is appropriate when the variance and mean of a random variable are known. It may also be the case that the range of possible values the random variable can assume are unknown or unbounded. In [7], Bretthorst justifies the use of Gaussian prior probability density functions by using the principle of maximum entropy. The other well-known justification for Gaussian prior probabilities is the central limit theorem [25]. The central limit theorem states that the sum of a number of independent random variables with finite variances will begin to approximate a Gaussian probability density function as their number approaches infinity. If there are many small, random or pseudo-random effects on a random variable, then its probability distribution may approach a Gaussian density function.

Jeffrey's Prior: Jeffrey's prior is a function which is uniform in the log scale and is typically used for scale parameters [7]. It denotes ignorance of the scale or magnitude of the values that the random variables can assume. The improper and unbounded Jeffrey's prior is formulated as:

$$p(\sigma|I) = 1/\sigma.$$

One needs to bound the prior with a range to be able to normalise it.

Hyperparameters: Sometimes our prior probability density functions contain parameters themselves called *hyperparameters*, such as the standard deviation σ for Gaussian priors. The term is used to distinguish them from the random variables of the model under consideration and they may differ depending on the state of one's prior knowledge. If we know their values we can use them as they are; otherwise we can treat them as random variables that are handled by probability theory and allocate prior probability density functions to them also. Hyperparameters are often treated as nuisance parameters and marginalised out [7].

3.3 Introduction to BSA

Depending on the prior information, the periodogram is a sufficient statistic for the frequency of a single sinusoid present in white noise. The periodogram can be used to calculate the probability density function of the frequency to an arbitrary resolution, depending on the amount of zero-padding. This matches one's intuition that the DFT can be used to find the frequency components in a signal, as the DFT is often used in frequency analyses in engineering and the sciences. BSA makes apparent the limitations of the DFT when dealing with more complex signals that contain multiple sinusoids or complex frequency behaviour for which the periodogram is not a sufficient statistic. Often when a signal is excessively noisy or there are changes in complex behaviour of the frequency components, the DFT may not be useful for analysing the entire signal, and methods such as the spectrogram are employed.

In the following sections we give a brief overview of Bretthorst's generalised approach to BSA in the case of a general signal model. A more detailed look at Bretthorst's derivation of BSA can be found in Appendix A. We will show how this method allows one to model an arbitrary signal given a signal model and amplitude, estimate its parameters, and find the most likely model given the data.

3.4 A BSA Example

In this section, we discuss an example analysis in order to clarify the end goal. We have two simple simulated signals which we will analyse using BSA. In Figure 3.1 we have two examples of BSA, one in each column. The left column shows the periodogram of a single sinusoid with an amplitude of 1 and 64 evenly spaced samples with white noise added ($\sigma = 1$) with the true frequency of 0.2 Hz/sample. The periodograms have been zero-padded to 4096 samples. The second and third graphs are the log probability density and probability density functions of the model of a single frequency with unknown noise. Note that the BSA log probability density is proportional to the periodogram, the sufficient statistic.

In the second column we have the sum of two sinusoids with frequencies of 0.2 and 0.25 Hz/sample respectively with the same energy noise added. The two other graphs in the right column are the log probability density and probability density of two frequencies with unknown noise. Since we have two parameters for the model, the probability density is also two-dimensional. The probability density is symmetric along the $\omega_0 = \omega_1$ axis due to the permutations of the frequencies where ω_0 and ω_1 could be either frequency in the signal.

Note how sharply peaked both probability densities are in the third row, indicating the confidence level in the parameter estimates. This does not, however, indicate the confidence in the models themselves. One would need to carry out model selection in order to calculate the model probabilities. Intuitively, one would expect the peaks in the periodogram to be the most likely frequencies as the periodogram is often used to find the frequencies in a measured signal. It is only when the noise becomes significant, the underlying signals are complex, or frequency components are very close to one another, that the periodogram starts to yield unreliable results.

3.5 BSA Parameter Estimation

This section gives a brief overview of the of BSA parameter estimation. It is based on [8], unless otherwise specified. A more detailed overview of BSA parameter estimation can be found in Section A.1.

With N data points $D = \{d_1, \dots, d_N\}$ sampled from $y(t)$, the continuous signal, at discrete times $\{t_1, \dots, t_N\}$, the signal model $M(t)$ and signal noise $e(t)$, we have the following equation to describe our data:

$$d_i = M(t_i) + e(t_i), \quad (1 \leq i \leq N). \quad (3.5.1)$$

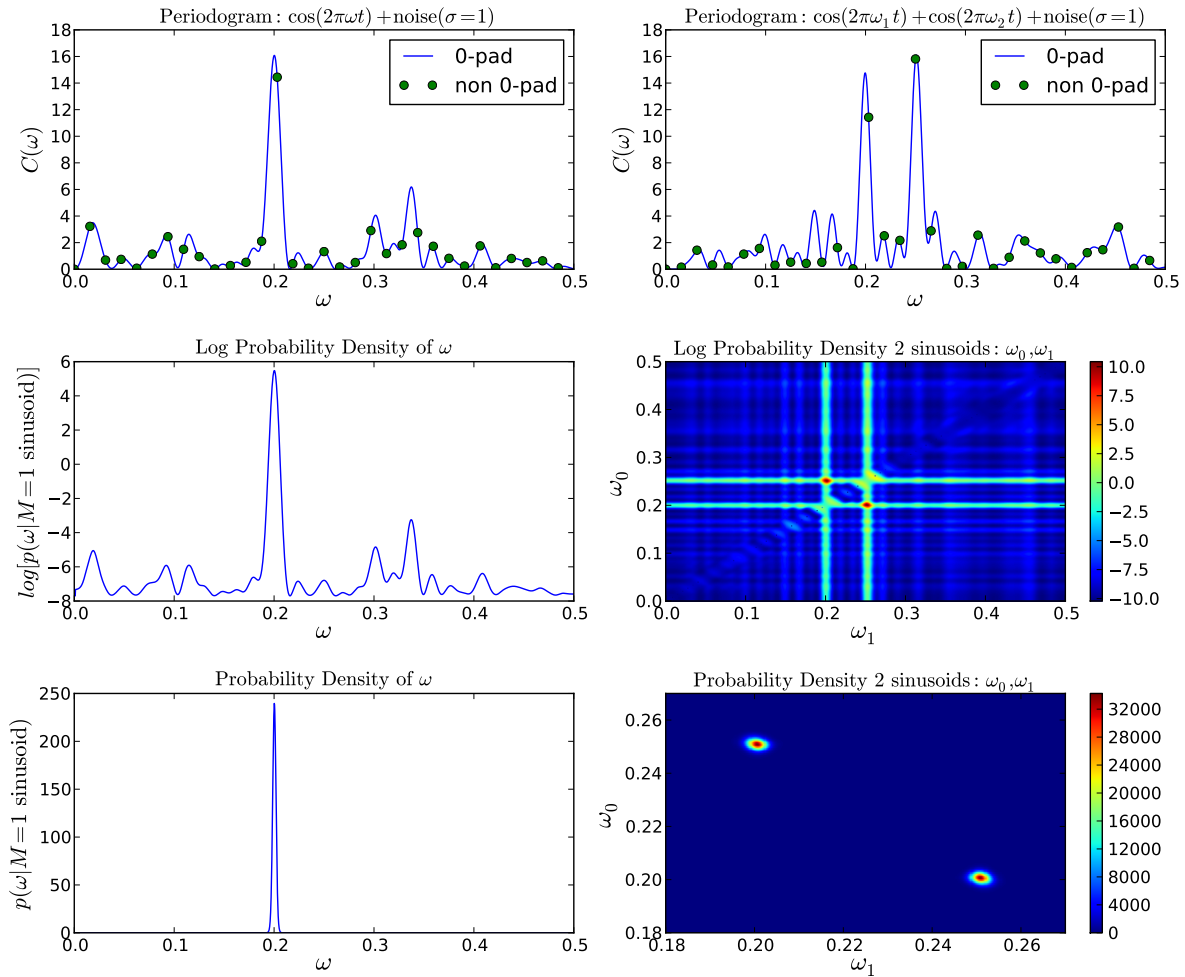


Figure 3.1: This figure has two examples of BSA, one in each column. The left column contains the example of a single sinusoidal signal in noise. From the top the figures are the periodogram, log of the probability density, and probability density functions of the frequency of the signal using a single signal model with unknown noise. The peak of the periodogram coincides with the peak of the BSA analysis. The second column contains an example of the BSA of two signals in noise where the signals are well separated and equally strong and clearly visible. From the top the figures are the periodogram, log of the probability density, and probability density functions of the frequencies of the signal with a two signal model with unknown noise. Note that the two peaks of the probability density are equal and coincide with the top two maxima of the periodogram.

All of the functions in the above equation are functions of time that have been sampled at discrete times. The noise is assumed to be additive and non-correlated.

By following Bretthorst's approach, we would like to calculate the parameters we are interested in, independent of those we do not wish to estimate. We will generally assume that little prior information is available and that the most informative prior information is in fact the functional form of the model – for example, the number and types of parameters as well as the model functions. This will also be true for our specific Doppler radar problem, as we know little more about the underlying signal than that it is the sum of reflected sinusoidal signals with Doppler-shifted frequencies. We assume that the measured data can be separated into a signal and noise. The noise models effects such as thermal or measurement noise. The parts of the model that we cannot accurately specify or that are incorrectly specified can also be seen as part of the noise [21].

The general form of the model from [8] is as follows:

$$M(t_i) = \sum_{j=1}^m B_j G_j(t_i, \Theta) \quad (3.5.2)$$

with B_j the amplitude multiplied with the j th model function $G_j(t_i, \Theta)$, out of m possible model functions. The set of r parameters that are going to be estimated are

$$\Theta = \{\Theta_0, \dots, \Theta_{r-1}\}.$$

The model parameters can be frequencies, chirp rates, signal decay or any arbitrary set of parameters used to describe the signal.

Referring to the parameter estimation from Equation 3.2.4, the probability of the model parameters, Θ , and nuisance amplitude parameters, \mathbf{B} , given the data and other prior information is

$$p(\Theta, \mathbf{B} | D, M_j, I) = \frac{p(D | \Theta, \mathbf{B}, M_j, I) p(\Theta, \mathbf{B} | M_j, I)}{p(D | M_j, I)}.$$

We wish to calculate the probability density of the random variables of interest Θ independent of the amplitudes \mathbf{B} to get

$$p(\Theta | D, M_j, I) = \int d\mathbf{B} p(\Theta, \mathbf{B} | D, M_j, I),$$

which is the probability of the parameters given the data. Because any information we have about our model parameters Θ will not influence our knowledge about the amplitudes \mathbf{B} we can write our prior knowledge about these independent parameters as:

$$p(\mathbf{B}, \Theta | M_j, I) = p(\mathbf{B}, | M_j, I) p(\Theta | M_j, I).$$

We use a uniform prior for $p(\Theta|M_j, I)$ for both parameter estimation and model selection. The prior for the amplitudes is approximated as an improper uniform prior for simplicity. We discuss the details of Bretthorst's approximation for parameter estimation in Appendix A.1.3.

The starting point to BSA parameter estimation is using a Gaussian prior for the noise, which is the difference between the model function $M(t_i)$ and the data d_i from Equation 3.5.1. Jaynes [7] and Bretthorst [8] justify using a Gaussian prior for the noise by using the principle of maximum entropy and the central limit theorem, as discussed in Section 3.2.6.

Bretthorst simplifies the calculations by calculating *orthonormal* (orthogonal and normalised) model functions from the original model functions:

$$M_{orth}(t_i) = \sum_{j=1}^m A_j H_j(t_i, \Theta), \quad (3.5.3)$$

where the new model $M_{orth}(t_i)$ now consists of orthonormalised amplitudes A_j and model functions $H_j(t_i, \Theta)$. From [8], the projection of the data onto the orthonormal model function is calculated as

$$h_j(\Theta) \equiv \sum_{i=1}^N d_i H_j(t_i, \Theta), \quad (1 \leq j \leq m), \quad (3.5.4)$$

and the sufficient statistic is the mean-square of the observed projections:

$$\bar{h}^2(\Theta) \equiv \frac{1}{m} \sum_{j=1}^m [h_j(\Theta)]^2. \quad (3.5.5)$$

BSA contains the concept of noise as a hyperparameter in the form of the noise variance σ^2 . If σ is known, then the joint probability of the Θ parameters conditional on the data and the noise is:

$$p(\Theta|D, I) \propto \exp \left\{ \frac{m\bar{h}^2(\Theta)}{2\sigma^2} \right\}. \quad (3.5.6)$$

If σ is not known, it is removed as a nuisance parameter. It is a scale parameter and restricted to positive values. The uninformative prior probability for scale parameters is the Jeffrey's prior. Marginalising σ given the Jeffrey's prior ($1/\sigma$) gives us

$$p(\Theta|D, I) \propto \left[1 - \frac{m\bar{h}^2(\Theta)}{Nd^2} \right]^{\frac{m-N}{2}}, \quad (3.5.7)$$

where the mean-square of the data $\overline{d^2}$ is calculated as follows:

$$\overline{d^2} = \frac{1}{N} \sum_{i=1}^N d_i^2.$$

Equation 3.5.7 only becomes singular if one has perfect knowledge of the Θ parameters. Equations 3.5.6 and 3.5.7 will allow us to calculate the frequency and chirp parameters of an arbitrary number of sinusoids.

3.6 BSA Relation to Periodogram

In this section we will briefly relate the periodogram with BSA. The more detailed derivation can be found in Appendix A.2. We would construct the model as follows for a single sinusoid:

$$M_{1 \sin}(t_i) = B_1 \sin(2\pi\omega) + B_2 \cos(2\pi\omega),$$

where $m = 2$ is the number of model functions. The projection of the data on the orthonormal model functions from [8] are

$$h_1(\omega) \approx \sum_{i=1}^N d_i \frac{\sqrt{2} \cos(2\pi\omega t_i)}{\sqrt{N}},$$

and

$$h_2(\omega) \approx \sum_{i=1}^N d_i \frac{\sqrt{2} \sin(2\pi\omega t_i)}{\sqrt{N}}.$$

Note the similarities to the projection of the data for the periodogram that can be calculated as:

$$C(\omega) = \frac{1}{N} [R(\omega)^2 + I(\omega)^2] = \frac{1}{N} \left| \sum_{j=1}^N d_j e^{2\pi i \omega t_j} \right|^2 = \frac{1}{N} |\text{DFT}|^2$$

and where the real and imaginary projections are:

$$R(\omega) = \sum_{j=1}^N d_j \cos(2\pi\omega t_j),$$

and

$$I(\omega) = \sum_{j=1}^N d_j \sin(2\pi\omega t_j).$$

From Equation 3.5.5 we then calculate

$$\overline{h^2}(\omega) \approx \frac{1}{2} \left[\sum_{i=1}^N d_i \frac{\sqrt{2} \cos(2\pi\omega t_i)}{\sqrt{N}} \right]^2 + \frac{1}{2} \left[\sum_{i=1}^N d_i \frac{\sqrt{2} \sin(2\pi\omega t_i)}{\sqrt{N}} \right]^2 = C(\omega)$$

and arrive at the periodogram $C(\omega)$ as the sufficient statistic for a single sinusoid in white noise.

3.7 BSA Model Selection

If one has a set of possible BSA models, how does one determine which one is most likely to be the correct model? Given the data and the available prior information, how strongly is that model supported relative to the alternatives? The questions asked for model selection are different to those asked for parameter estimation but are solved using the same methods. This section is based on [22], except where otherwise indicated. The details on BSA model selection can be found in Appendix A.3.

The questions can be solved by calculating the joint probability density function of the data and the parameters, and then integrating out the parameters. For BSA this cannot be done analytically. Using Bayes' theorem we have the likelihood function for our model under consideration:

$$p(D|M_j, I) = \int d\Theta p(\Theta|M_j, I) \times \int d\mathbf{A} p(\mathbf{A}|\Theta, M_j, I)p(D|\mathbf{A}, \Theta, M_j, I), \quad (3.7.1)$$

where M_j is the j th model function under consideration with the parameters of interest in the previous section Θ and the model function amplitudes \mathbf{A} being marginalised out.

Unlike our method of parameter estimation from the previous section, all the numerical factors must be kept as they do not necessarily cancel out during model selection. One can consider all the parameters as nuisance parameters which are to be marginalised. Since we are ignorant as to which model is the correct one, we use a uniform prior for the models. The following equation for the model probability has the constants cancelling out:

$$p(M_j|\sigma, D, I) = \frac{p(D|\sigma, M_j, I)p(M_j|I)}{\sum_{k=1}^s p(D|\sigma, M_k, I)p(M_k|I)} = \frac{p(D|\sigma, M_j, I)}{\sum_{k=1}^s p(D|\sigma, M_k, I)},$$

where s is the number of hypotheses under review and the prior for our models is $p(M_j|I) = 1/s$.

From [22] we then have a model likelihood that is comparable between models:

$$p(D|M_j, I) \approx \Gamma\left(\frac{m}{2}\right) \left(\frac{N-m}{2}\right) v_1^{-0.5} \dots v_r^{-0.5} \times \left[\frac{m\bar{h}^2}{2}\right] \left[\frac{N\bar{d}^2 - m\bar{h}^2}{2}\right]^{\frac{m-N}{2}} \Bigg|_{\hat{\Theta}}, \quad (3.7.2)$$

where defining

$$b_{kl} \equiv - \left. \frac{\partial^2 m \bar{h}^2}{2 \partial \Theta_k \partial \Theta_l} \right|_{\hat{\Theta}}, \quad (3.7.3)$$

and v_j is the j th eigenvalue of the matrix b_{jk} . For models with multiple maxima due to symmetry, we can multiply the model likelihood by the number of peaks. For example, the two-peaked example in Figure 3.1 would be multiplied by 2, otherwise we would only approximate the integral of one of the two identical peaks.

3.8 Conclusion

BSA is a generalised method used to calculate the joint probability density function of the parameters of a signal model given the signal data. It can be used to calculate the parameters of arbitrary model functions if one can carry out the necessary operations of Bayesian probability theory. We showed the relation between BSA and the periodogram. The concepts of Bayesian probability theory that were introduced before the derivation of BSA will also be used in later chapters of this thesis.

Chapter 4

BSA Experimental Results

4.1 Introduction

In this chapter we use BSA to do parameter estimation and model selection on the Doppler radar data. The Doppler radar signal is broken up into long segments that relate to the number of frequency components and the different dynamics of each part of the signal. We analyse these longer segments to gain a better understanding of the application of BSA to our data. We also break up the signal into even smaller parts similar to a spectrogram. Both the “noiseless” and specified noise power models are used in our analyses.

4.2 Signal Models

From the spectrogram there would appear to be a maximum of three sinusoids present at a time in the Doppler radar signal. We model single-, two- and three-frequency models with three different types of behaviour for the signals – namely stationary, linear chirp, and quadratic chirp frequencies – which brings it to nine models in total. The stationary frequency model assumes the sinusoids of interest have a constant frequency. Signals with a linear chirp in frequency are also known as sweep signals. The quadratic chirp model is an extension of the linear chirp model and models the instantaneous frequency forming a quadratic polynomial function.

For r stationary sinusoids the model function is:

$$M(t_i) = \sum_{j=0}^{r-1} B_{j,1} \cos(2\pi\omega_j t_i) + B_{j,2} \sin(2\pi\omega_j t_i),$$

where then the two stationary sinusoid model function is

$$M(t_i) = B_{0,1} \cos(2\pi\omega_0 t_i) + B_{0,2} \sin(2\pi\omega_0 t_i) + B_{1,1} \cos(2\pi\omega_1 t_i) + B_{1,2} \sin(2\pi\omega_1 t_i).$$

For a sinusoid with linear chirp the model is

$$M(t_i) = \sum_{j=0}^{r-1} B_{j,1} \cos(2\pi\omega_j t_i + \alpha_j t_i^2) + B_{j,2} \sin(2\pi\omega_j t_i + \alpha_j t_i^2),$$

which for a single linear chirp is the chirpogram model from [7]:

$$M(t_i) = B_{0,1} \cos(2\pi\omega_0 t_i + \alpha_0 t_i^2) + B_{0,2} \sin(2\pi\omega_0 t_i + \alpha_0 t_i^2). \quad (4.2.1)$$

One can extend the models to an arbitrary amount of frequency parameters, but we only do so up until the quadratic polynomial chirp model – specifically

$$M(t_i) = \sum_{j=0}^{r-1} B_{j,1} \cos(2\pi\omega_j t_i + \alpha_j t_i^2 + \beta_j t_i^3) + B_{j,2} \sin(2\pi\omega_j t_i + \alpha_j t_i^2 + \beta_j t_i^3). \quad (4.2.2)$$

4.3 Finding the Peak

In order to approximate the probability density functions of BSA for our model selection from Section 3.7, we need to be able to find the peak. As used in [6], where BSA is implemented for biological time series, we used the Nelder-Mead optimisation technique [26]. The Nelder-Mead technique is a heuristic search method to find the peak of a function that does not make use of derivatives (see Appendix B.3).

For the model functions in the previous section we could choose that for our models,

$$0 \leq t_i < N,$$

where N is the number of samples in the signal under analysis. Instead we choose,

$$-\frac{N}{2} \leq t_i < \frac{N}{2},$$

which simply means that the $\omega = \{\omega_0, \dots, \omega_r\}$ parameters are the instantaneous frequencies in the middle of the measured signal as opposed to the starting frequency. It allows us to use the peaks of the absolute value of the DFT as a starting point for the Nelder-Mead technique, as they tend to be close to the instantaneous frequencies in the middle of a measured signal.

4.4 Simulated Examples

Figure 4.1 is an example of a chirpogram (with a signal model such as in Equation 4.2.1) for a simulated signal with linear chirp, and Figure 4.2 is a chirpogram of a different simulated signal with quadratic chirp. In Figure 4.2

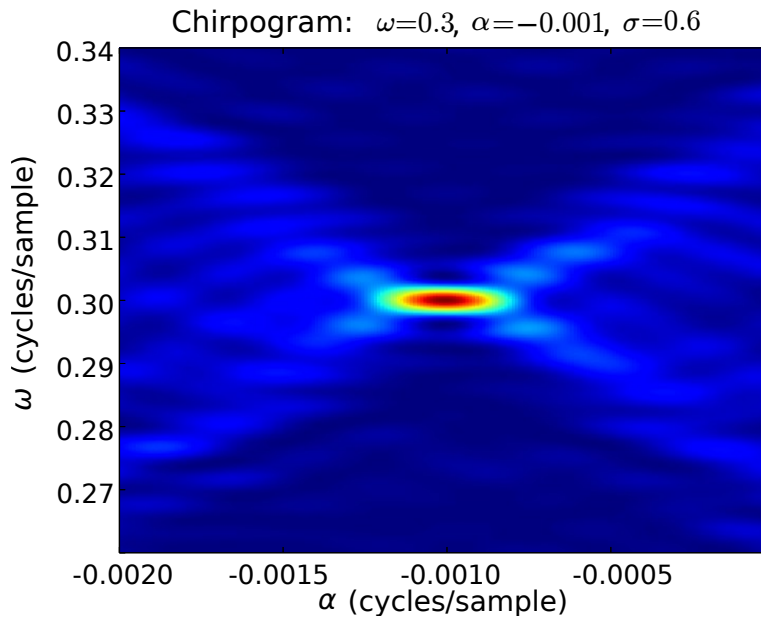


Figure 4.1: *The log of the chirpogram of a simulated signal with a linear chirp: $\omega = 0.3, \alpha = -0.001, \beta = 0.0$. The density function has a single strong peak by the correct parameter values.*

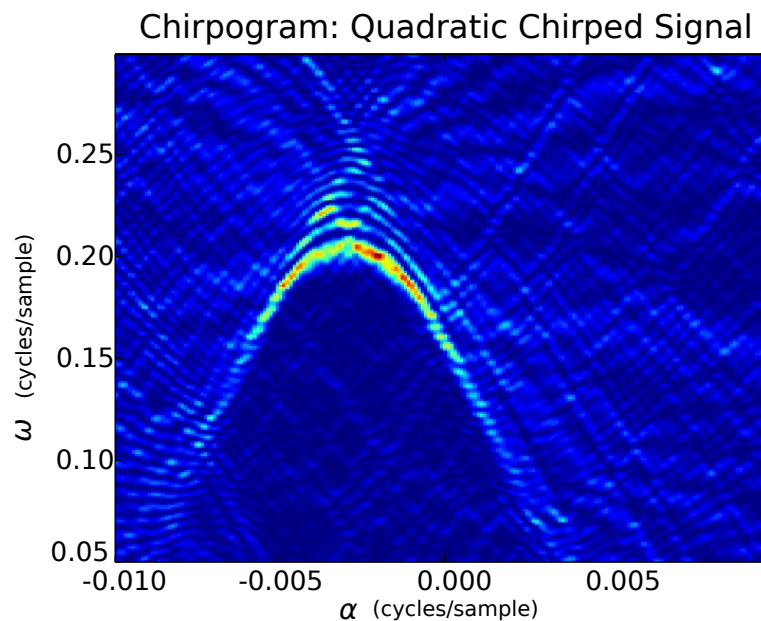


Figure 4.2: *The log of the chirpogram of a simulated signal with a quadratic chirp: $\omega = 0.2, \alpha = -0.003, \beta = -0.00001$. The density function is more spread out in comparison to Figure 4.1.*

the instantaneous frequency of the signal forms a quadratic polynomial func-

tion but we only model it as a linear chirp. The signal with linear chirp has a single strong peak where the quadratic chirp simulated signal's peak shows that there is more complexity present in the signal. We will compare these chirpogram with Doppler radar data in Section 4.6.1.

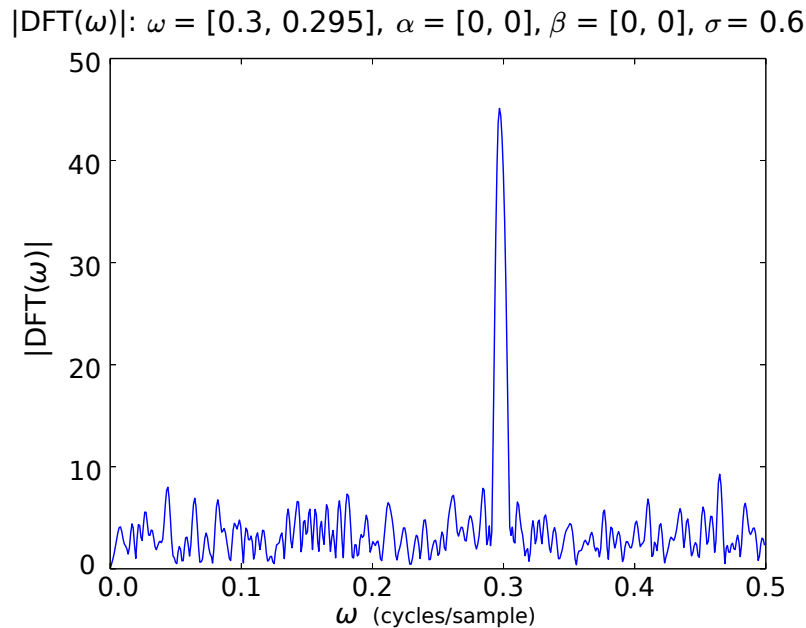


Figure 4.3: *The absolute value of a zero-padded DFT of two very close sinusoids ($\omega_0 = 0.3$, $\omega_1 = 0.295$, $\sigma = 0.6$). The DFT cannot distinguish between the two sinusoids.*

Figure 4.3 and 4.4 are both from a third simulated signal. They indicate the absolute value of the DFT and the log probability density of the two-stationary-sinusoid model respectively. This simulated signal contains two sinusoids that have two very close frequencies with white noise added ($\sigma = 0.6$). In the zero-padded DFT there is a single peak, but with the two-sinusoid model it clearly indicates that two frequency components are present. Doing model selection, the model with two stationary sinusoids is chosen above the other stationary sinusoid models with near certainty, as can be seen in Table 4.1. We will see examples of possible multiple close sinusoids in our Doppler radar data in Section 4.6.

These examples show that the complexity of the underlying signal can either be apparent from an inspection of the probability density function such as can be seen in the difference between the signals from Figure 4.1 and 4.2, or more subtle such as in Figure 4.3 and 4.4, where the same signal with different models can give very different results. One needs to do model selection to find

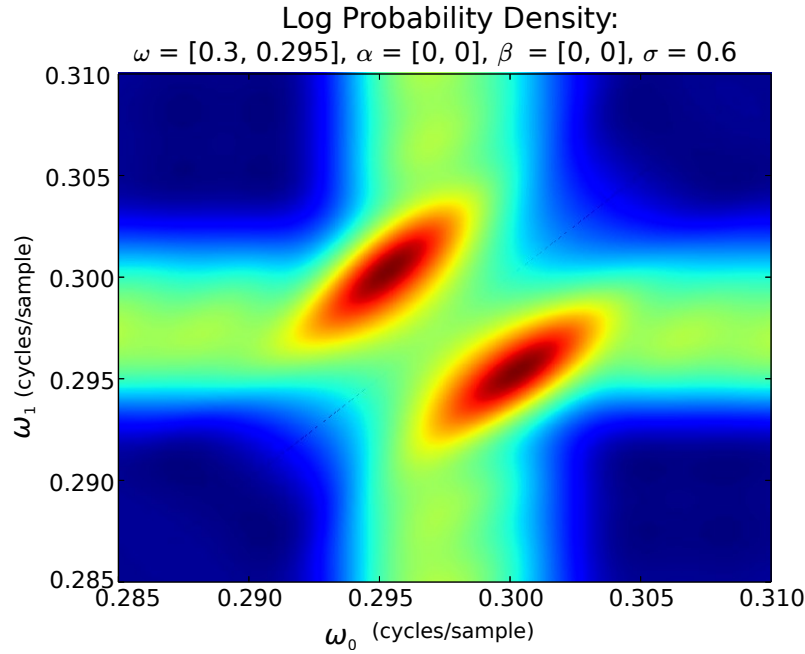


Figure 4.4: Log probability density function of two very close sinusoids with peak: $\omega_0 = 0.2998$, $\omega_1 = 0.2951$. BSA separates and distinguishes between the two sinusoids.

Table 4.1: The BSA model probabilities for two very close sinusoids $\omega_0 = 0.3$, $\omega_1 = 0.295$ in Figure 4.4. The most likely model is correctly selected as the two stationary-frequency model.

	Single Stationary	Two Stationary	Three Stationary
Model Probability	9.1169e-35	9.9537e-01	4.6290e-03

the most likely model, as it is not always apparent from the probability density function.

Figure 4.5 contains the spectrogram of a simulated signal where it is not immediately apparent how many frequency components are present. There appear to be many frequency components before time-step eight when, in fact, there is only one quadratic chirped sinusoid. BSA selects the quadratic chirp model with near 100% probability from the list possible models from Section 4.2. More details on this specific example can be found in Appendix C.8 and other simulated examples can be found in Appendix C. Not only can BSA select the correct model where the signal behaviour is unclear in the periodogram or DFT, it can also do so when the behaviour is unclear in the spectrogram.

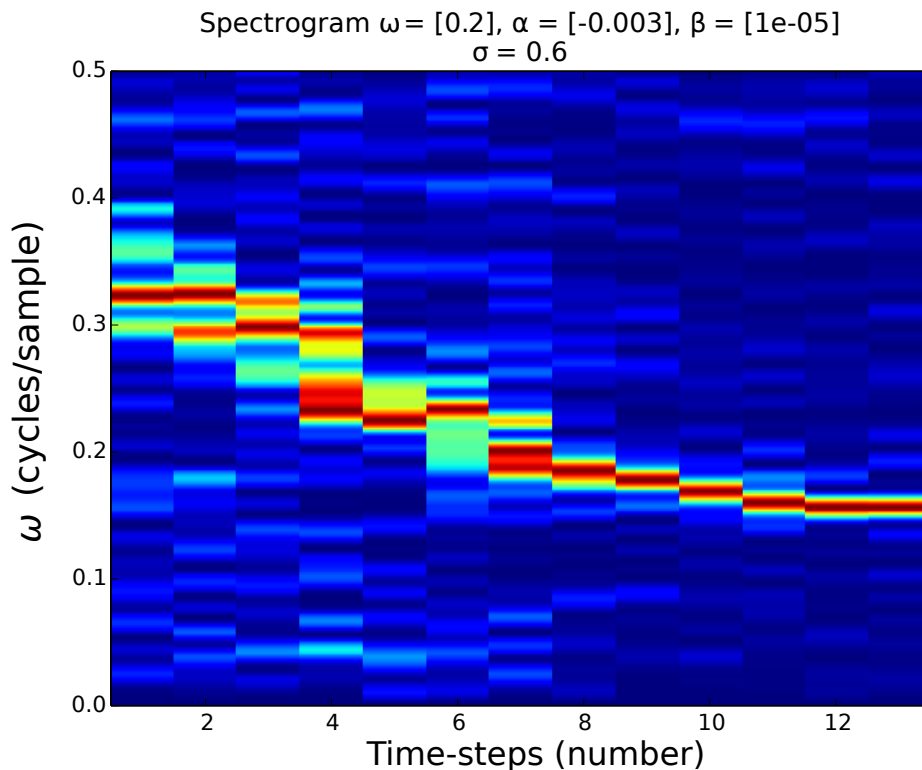


Figure 4.5: Spectrogram of a single sinusoid with quadratic chirp. It is not immediately apparent from the spectrogram that there is only one frequency component present. BSA, however, selects the correct model for this signal with near 100% probability from nine possible models.

4.5 Doppler Radar Signal Noise and “Noiseless” Model

As discussed in Chapter 3, BSA has a concept of signal noise unlike the DFT or periodogram. Knowledge about the signal noise can be explicitly incorporated in the BSA models, even if the only knowledge is an uninformative prior such as a Jeffrey’s probability distribution for the noise variance.

For the signal models from Section 4.2, most noise in the Doppler signal will be attributed to the mismatch between the signal model and the true signal behaviour. The noise is not necessarily thermal, measurement or other noise from the environment. It is possible that there are targets that we do not fully understand, or are not aware of, that will contribute to what our BSA models will then see as noise. This is likely due to the complex shapes of the objects which reflect the radar signal. As we do not have a good understanding of the noise, we use the noiseless model from Equation 3.5.7 for parameter estimation

and Equation 3.7.2 to approximate model selection in the following section:

$$P(\Theta|D, I) \propto \left[1 - \frac{m\bar{h}^2(\Theta)}{N\bar{d}^2} \right]^{\frac{m-N}{2}},$$

$$P(D|M_j, I) \approx \Gamma\left(\frac{m}{2}\right) \left(\frac{N-m}{2}\right) v_1^{-0.5} \dots v_r^{-0.5}$$

$$\times \left[\frac{m\bar{h}^2}{2} \right] \left[\frac{N\bar{d}^2 - m\bar{h}^2}{2} \right]^{\frac{m-N}{2}} \Bigg|_{\Theta}.$$

We used these same models for the simulated results and for the examples in Appendix C.

4.6 Long Segment BSA

There are two typical types of signals for our Doppler radar data: a golf swing with the club hitting a ball as in figure 4.6, and a golf swing that does not involve hitting a ball, such as in Figure 4.7. We break up the two types of signal into different signal segments that reflect our understanding of the underlying signal behaviour. Each segment contains the same typical target objects and target behaviour. In the long segment analyses, the selected models suggest that the underlying complexity of the signal is greater than our expected BSA hypotheses. The behaviour of the sinusoids and the signals contain non-linearities that add complexity to the signals. We will see in Section 4.7 that this also holds true for the shorter time-step analyses.

The following list includes the targets of interest that are present during the given segments of the signals.

- (A) The downward swing of the club.
Target objects: the club-head and sometimes low-velocity player body motion. Present in Figure 4.6 and Figure 4.7.
- (B) The ball post club swing.
Target objects: the ball. Present in Figure 4.6.
- (C) The ball and club post stick-flash.
Target objects: the ball and club-head. Present in Figure 4.6.
- (D) The club pre-impact.
Target objects: the club-head and stick-flash and sometimes low-velocity player body. Present in Figure 4.6.

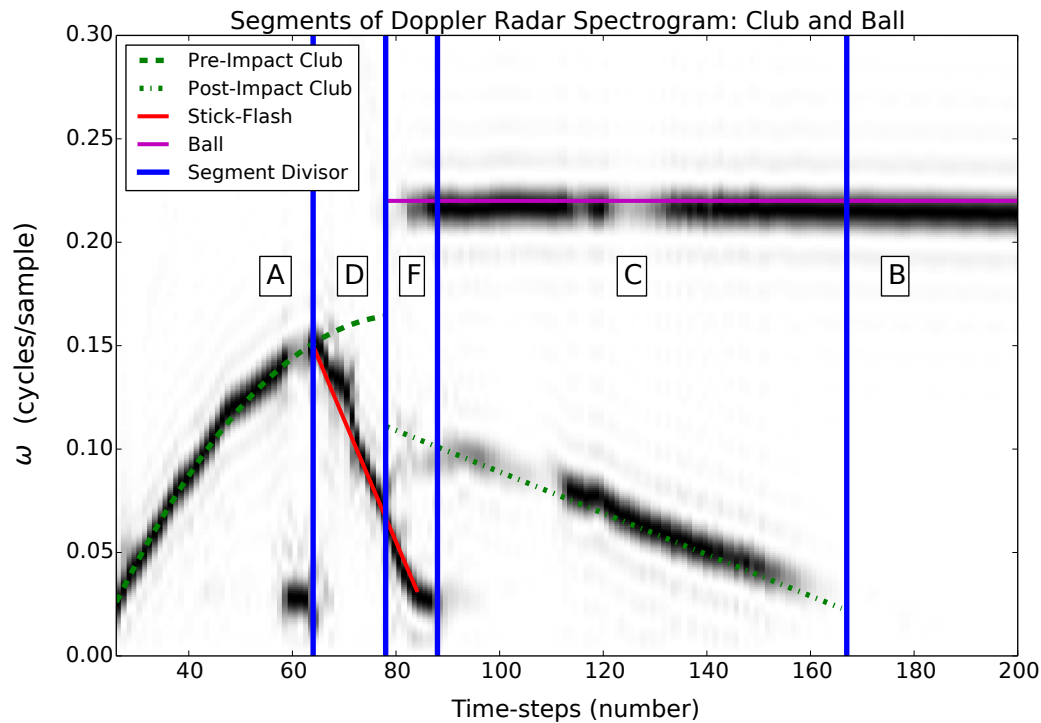


Figure 4.6: Spectrogram of a golf swing hitting a golf ball segmented into different parts: Window Size = 64, Step-size = 16. The frequency components of interest are caused by the club (before and after impact), the stick-flash, and the ball. The low-frequency component near time-step 60 is caused by the motion of the player's body.

(E) Stick-flash mid-swing.

Target objects: the club-head and stick-flash. Present in Figure 4.7.

(F) Post-impact ball and club.

Target objects: the ball, club-head and stick-flash. Present in Figure 4.6.

(G) Non-impact club post stick-flash.

Target objects: the club-head. Present in Figure 4.7.

The following subsections contain the typical findings of the analyses for each segment of the signals. We provide example results for Segment A, B, D and E and refer the reader to Appendix D for more examples of the signal parts. In order to evaluate the results we superimpose the instantaneous frequencies, calculated using the most likely BSA parameters, over the spectrogram. We display the most likely model parameters in a table with their respective standard deviations that were calculated by approximating the probability density functions as Gaussian functions.

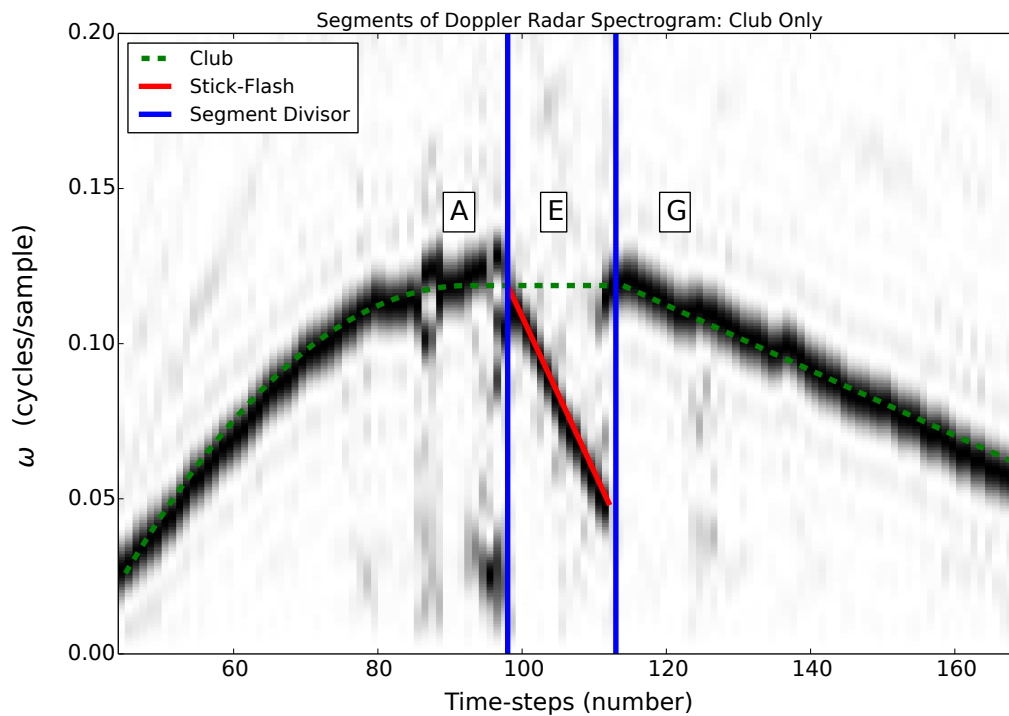


Figure 4.7: Spectrogram of a golf swing without hitting a ball segmented into different parts: Window Size = 64, Step-size = 16. The frequency components of interest are caused by the club and the stick-flash. The low-frequency component near time-step 100 is caused by the motion of the player's body.

4.6.1 Segment A: Club downward swing

Expected Model: Single Quadratic Chirped Sinusoid

Typical Results: For some signals Segment A had a single quadratic chirp as the most likely model, such as the example in Figure 4.8, where others had multiple quadratic chirps very close to each other. The multiple close signals suggest that the signal does not always follow a quadratic chirp path or that there is more complexity in the frequency range where the multiple close sinusoids are found.

Example: The example in Figure 4.8 has a quadratic chirp as the most likely model, but does not appear to match the spectrogram exactly. The chirpogram in Figure 4.9 of our example Segment A resembles the example chirpogram of a chirped simulated signal in Figure 4.1 and not the quadratic chirped sinusoid in Figure 4.2. This suggests that despite the quadratic chirp model being selected with high certainty above the chirp model, the frequency change is not approximated well by a quadratic polynomial. One can see

in Figure 4.8 how closely the chirpogram resembles the linear chirp model. Despite this, the quadratic chirp model was selected with near 100% certainty. Table 4.2 lists the most likely model's parameters for the Segment A example and their respective standard deviations. The standard deviations are small, which indicates that the model is very confident of the model parameters. The peak width of the DFT for the example would be $\frac{2}{320} = 6.250e-03$ for the 320 samples. In comparison the standard deviation for the centre instantaneous frequency, ω_0 , is $5.011e-05$. The peaks of the BSA probability distributions become sharper the more data is available, which in part explains the small standard deviations.

Table 4.2: *The most likely signal parameters of Segment A Example 1. The signal model is a single quadratic chirped sinusoid selected with near 100% probability.*

Model Probability	ω_0	α_0	β_0
1.00000e+00	1.01549e-01	3.65845e-04	-3.40525e-07
Std Dev	5.01091e-05	3.77107e-06	6.08204e-03

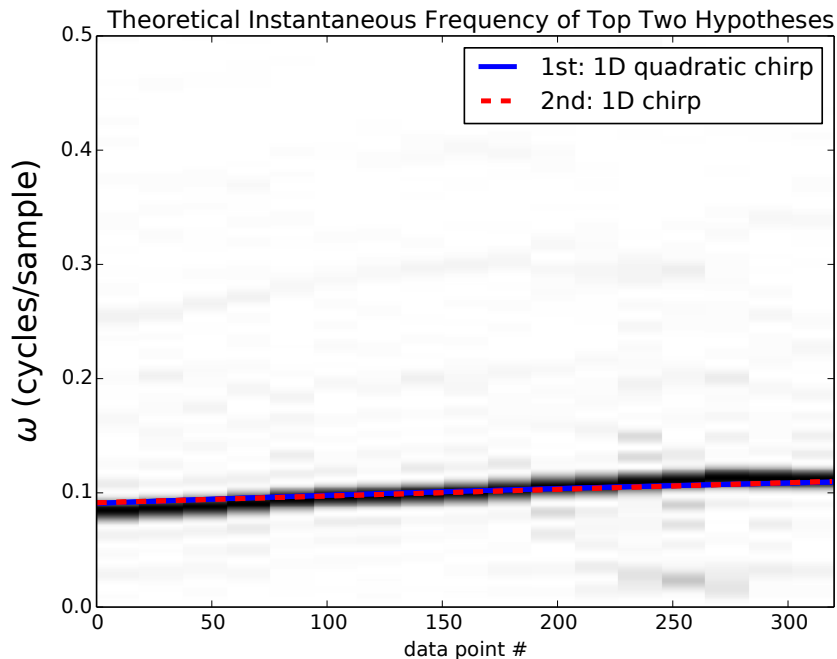


Figure 4.8: *Segment A (club downward swing) Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram. The two most likely models have near identical instantaneous frequencies which do not completely match the spectrogram.*

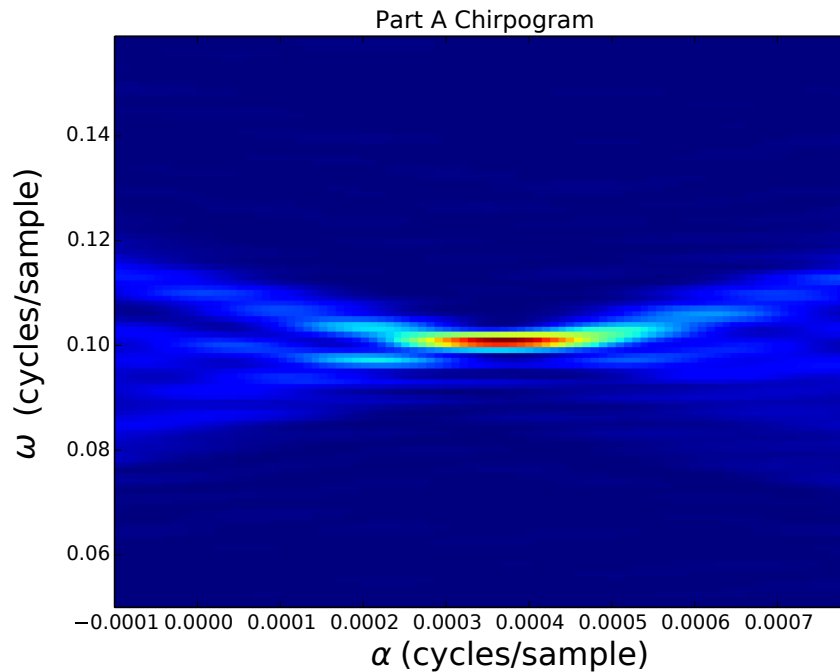


Figure 4.9: *The chirpogram of Segment A (club downward swing) Example 1. The probability distribution resembles a linear chirped sinusoid despite the quadratic chirped model being the most likely model.*

4.6.2 Segment B: Ball Post Club Swing

Expected Model: Single Stationary or Chirped Sinusoid

Typical Results: We expected this part of the signal to only contain the ball. Depending on the signal strength reflected from the ball there is sometimes a consistent low-frequency component present such as can be seen in Figure 4.10. Our initial impression was that this was caused by the spin of the ball; however, an investigation of the log spectrogram showed that there are consistently low-frequency effects throughout the golf swing, such as the example from Figure 2.2. These low-frequency signal components are often obscured by the stronger reflected signals. Some analyses of the ball also showed multiple, very close signals in the region of the ball's frequency component. This effect would appear to be similar to that seen with the club downward swing where the BSA sometimes indicates further complexity in a frequency range than the spectrogram suggests.

Example: The example in Figure 4.10 contains a low-frequency component and two very close frequency components where the ball Doppler shift is present. In Section 4.4 we had a simulated example of two very close frequency components. We hypothesise that the ball component itself contains

more frequency components than is apparent in the spectrogram. This could be explained by the fact that the ball is a three-dimensional object that reflects multiple close frequencies caused by the different Doppler shifts of the different spinning parts of the ball. The most likely parameters for the example are in Table 4.3, which also have very small standard deviations. The standard deviation of ω_1 has a comparable order of magnitude to ω_0 from Table 4.2 despite having fewer samples, namely 160 compared to 320. This is likely due to the relatively stationary nature of the frequency component near $\omega = 0.2408$ and the relatively little noise in comparison to the example in Segment A. In comparison, the other frequency components in Table 4.3 are less confident.

Table 4.3: *The most likely signal parameters of Segment B Example 1. The signal model is three stationary sinusoids selected with near 100% probability.*

Model Probability	ω_0	ω_1	ω_2
1.00000e+00	2.44803e-01	2.40793e-01	1.91549e-02
Std Dev	1.93046e-04	1.11478e-05	2.85377e-04

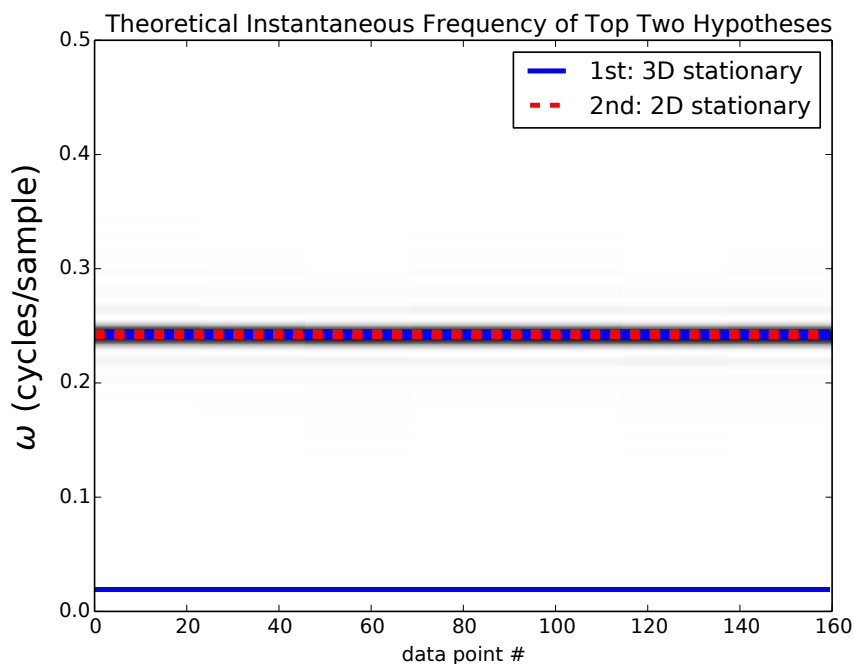


Figure 4.10: *Segment B (ball post club swing) Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram. The low-frequency component is likely caused by the low-frequency effects we saw in Figure 2.2. The two very close frequency components could be caused by two very close sinusoids.*

4.6.3 Segment C: Ball and Club Post Swing

Expected Model: Two Chirped Sinusoids

Typical Results: In this section we expect to start seeing signal segments with multiple targets of interest. Segment C contains the ball and the club-head after the ball has been hit. We expect two sinusoids to be the most likely model, one with a negative chirp (the club) and one with a near-zero negative chirp (the ball). Here we see the same effect of multiple sinusoids being present in the frequency area of the ball, such as in Segment B. The club can also be seen in the BSA. The club-head sometimes exhibits non-linear behaviour and “wobbles”, which cause the BSA to attribute multiple sinusoids to it. This is less surprising than the golf ball exhibiting multiple sinusoids and would suggest that the club-head is not always a linear chirped sinusoid.

4.6.4 Segment D: Pre-impact Club-head and Stick-flash

Expected Model: Two Chirped Sinusoids

Typical Results: We expected to see a positive chirped (club) and a negative chirped (stick-flash) sinusoid as the most likely model. What is noteworthy about this segment is that the stick-flash, described in Section 2.4.4, completely obscures the club-head frequency component pre-impact. The signal is further complicated by the fact that the stick-flash component is so strong that it tends to cause clipping of the measured signal, and the harmonic of the stick-flash obscures other signals even further. The effects of the stick-flash obscure the club-head even when removing the frequency areas where the stick-flash is present from analyses. Like Segment A and Segment B, the stick-flash also tends to contain multiple sinusoids in the BSA.

Example: The example in Figure 4.11 suggests that there is more complexity in the stick-flash than can be described by our BSA models. Three chirped sinusoids are attributed to the stick-flash and the club-head is not seen by BSA. The model parameters are shown in Table 4.4.

Table 4.4: *The most likely signal parameters of Segment D Example 1. The signal model is three chirped sinusoids selected with near 100% probability.*

Probability	ω_0	α_0	ω_1	α_1	ω_2	α_2
1.00e+00	1.41e-01	-2.47e-03	1.29e-01	-2.57e-03	1.14e-01	-2.40e-03
Std Dev	2.72e-04	5.02e-05	1.09e-08	1.04e-04	3.10e-04	4.01e-05

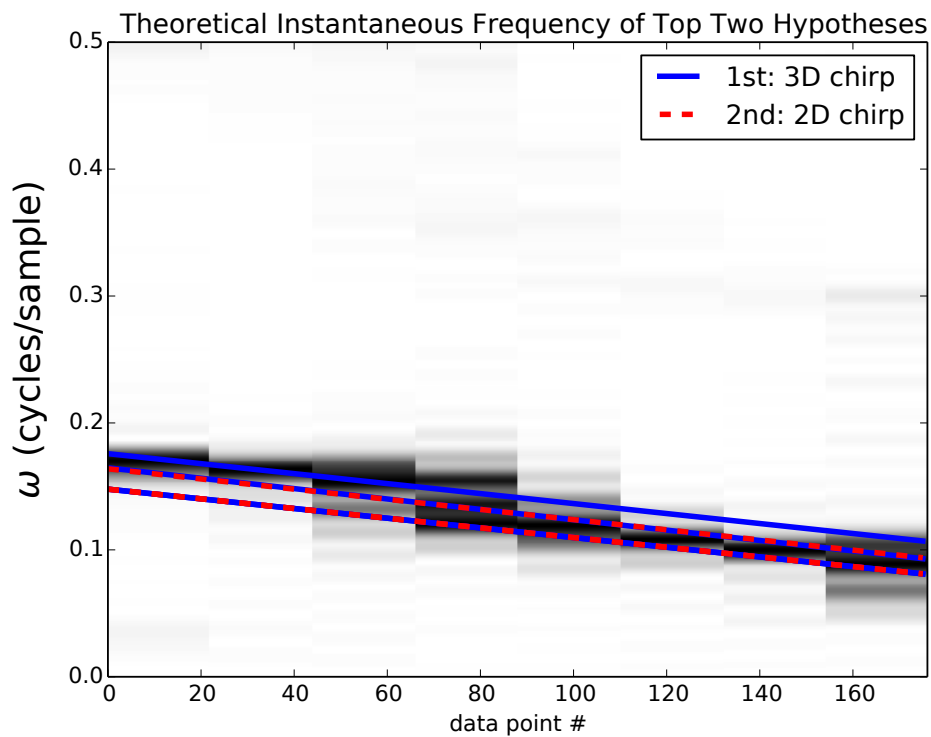


Figure 4.11: *Segment D (pre-impact club-head and stick-flash) Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram. Only the stick-flash was picked up by BSA with multiple frequency components being attributed to it.*

4.6.5 Segment E: No Impact Stick-Flash

Expected Model: Two Chirped Sinusoids

Typical Results: This is similar to Segment D where the targets of interest are the club-head and the stick-flash.

Example: The example in Figure 4.12 is noteworthy because the harmonic frequency caused by the clipping of the stick-flash component shown in the BSA results in a negatively chirped sinusoid. The club-head is also visible in the BSA result in this example, which is a rarer occurrence as it is usually obscured by the stick-flash. The model parameters are shown in Table 4.5.

Table 4.5: The most likely signal parameters of Segment E Example 1. The signal model is three chirped sinusoids selected with near 100% probability.

Probability	ω_0	α_0	ω_1	α_1	ω_2	α_2
1.00e+00	1.28e-01	-7.61e-05	8.18e-02	-1.25e-03	1.51e-01	-3.95e-03
Std Dev	2.91e-04	2.54e-05	6.88e-05	4.85e-06	4.10e-04	2.38e-05

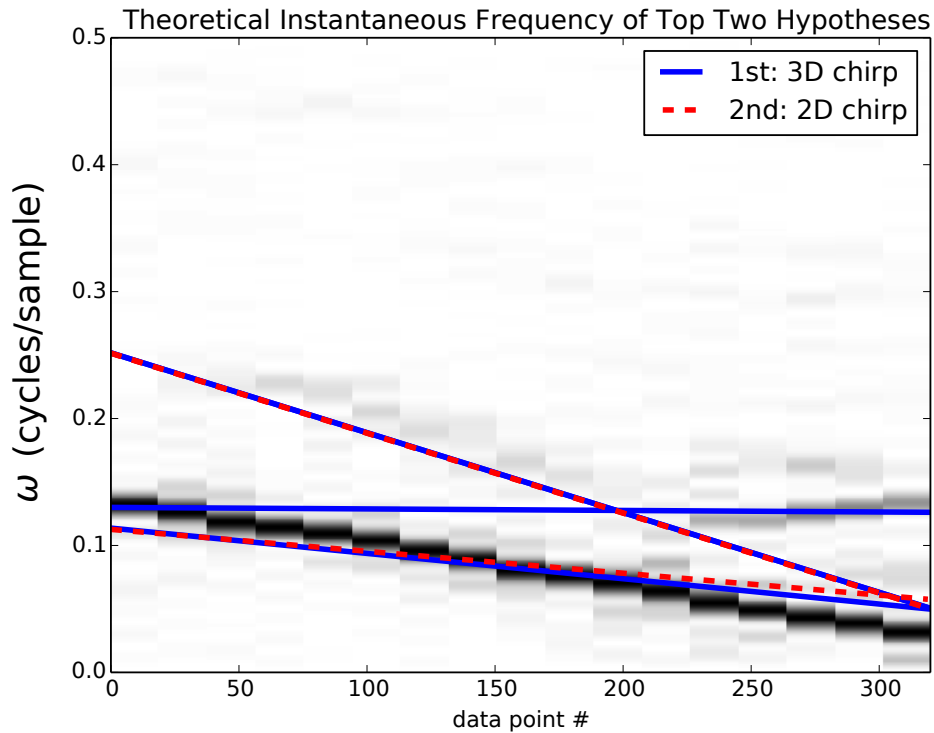


Figure 4.12: Segment E (no impact stick-flash) Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram. The club-head, stick-flash and a harmonic of the stick-flash was picked up by BSA.

4.6.6 Segment F: Post-impact Club-head, Ball, Stick-flash

Expected Model: Three Chirped Sinusoids

Typical Results: Depending on the signal strength of the club-head, these analyses either contain all three targets as expected, or BSA attributes two sinusoids to either the ball or the stick-flash similar to Segment B and Segment D.

4.6.7 Segment G: Club-head

Expected Model: Single Quadratic Chirped Sinusoid

Typical Results: This part resembles Segment C but without the ball being present.

4.6.8 Summary of Results

The results of the BSA models tend to be very confident in both parameter estimation and in model selection and do not always match our expectations given the spectrograms of the signals. In contrast, our analyses of simulated signals showed that BSA almost always picked the correct model. It is only when the simulated signals are particularly noisy and the spectrogram becomes uninterpretable that BSA struggles to select the correct model.

4.7 Short Time-step BSA

To reduce the effect of non-linearities on the signal of interest and to be able to use state-space estimation-like methods, we do a short time-step analysis of the Doppler radar signals. This approach is similar to a spectrogram where instead of a periodogram we do a BSA analysis for each signal model for a short window of the signal. Similarly to the spectrogram, we have a time-step that is shorter than the window length to allow overlap of the analyses and a higher time resolution. This will allow us to capture the non-linear effects of the transitions between signal segments. For each BSA model we find the peak of the probability density function for that model and its relative probability to the other models. We will use this analysis as the basis of our PGM models in later chapters.

4.7.1 Example Analysis

Figure 4.13 is an example of such an analysis. We superimpose a scatter-plot on the spectrogram of the signal. We group the models into the number of frequency components, namely one, two, or three frequency components. For each BSA model group the scatter-plot is located on the marginal of the most likely frequency parameters, ω . The size of each dot in the scatter-plot indicates the relative probability of the BSA model group and thus the relative likelihood of the number of frequency components present. This allows us to see at a glance the most likely number of signals and their most likely stationary frequency position. Figure 4.14 shows the relative probabilities of the number of signals present in a time-step.

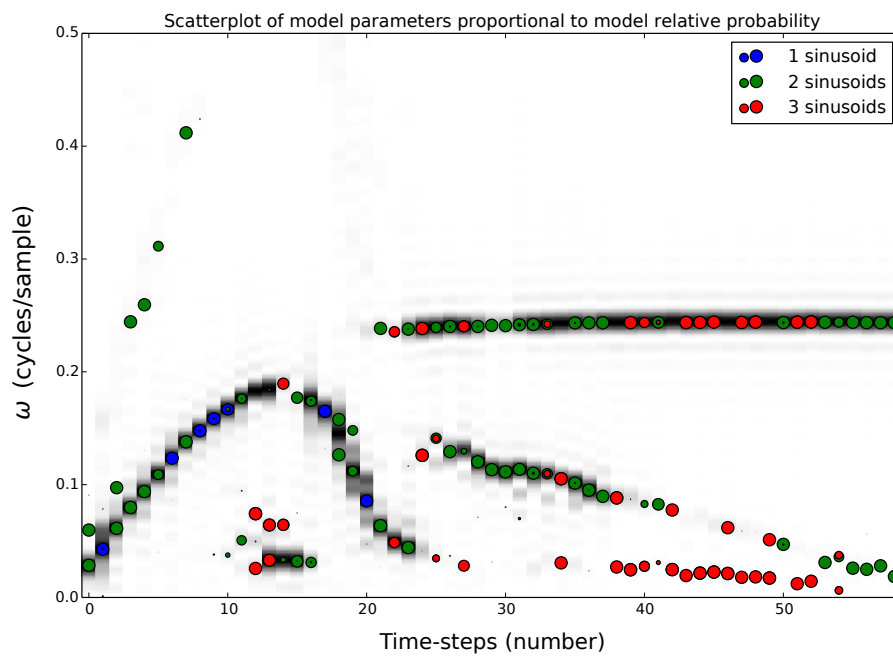


Figure 4.13: Scatter-plot of BSA stationary sinusoid frequency values with the unknown noise BSA model. The size of each marker indicates the probability of the specific BSA model. Note that some of the markers are very close. For example, at time-step 40 two of the markers for the three frequency components are close enough to overlap.

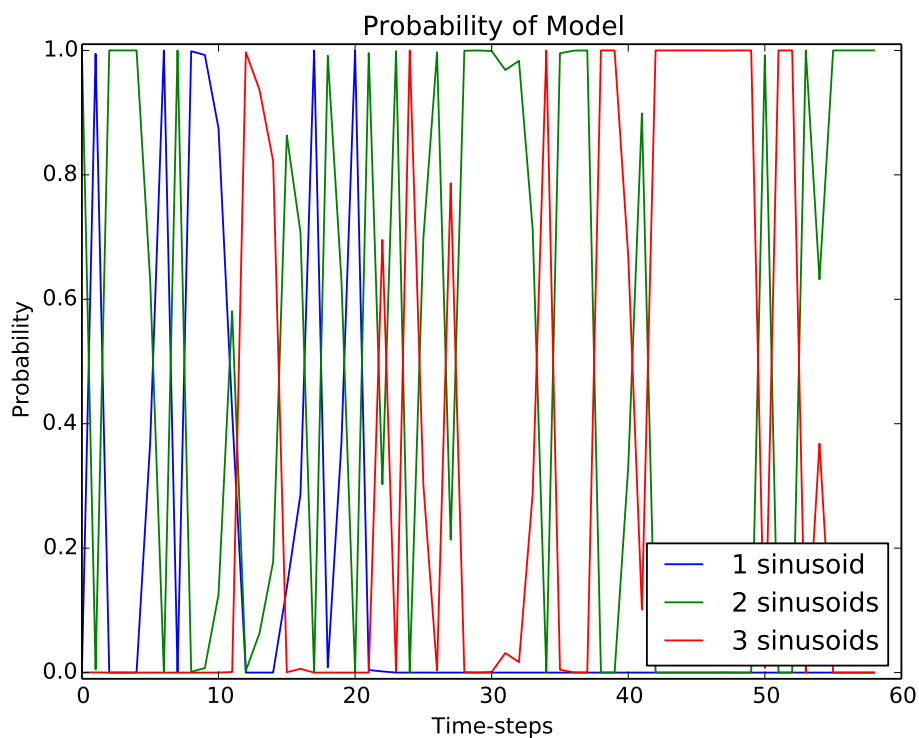


Figure 4.14: The BSA model probabilities from Figure 4.13

4.7.2 Short Time-step Parameter Estimation

Estimating parameters such as the linear and quadratic chirp parameters on short windows of the Doppler signal is problematic as the number of samples is often small enough that one gets contradictory information between time-steps. Despite the chirp parameters varying so unpredictably between time-steps, the frequency parameters, ω , are often much more reliable and more closely resemble inspection of the spectrogram as can be seen in Figure 4.13. Thus to avoid the unpredictability of the other spectral parameters and for the sake of simplicity we will only use the ω parameters in further analyses.

4.7.3 Increasing Model Noise

We can see in this particular example in Figure 4.13 that the more complex models are selected when there are seemingly less complex signals present. Even if we were to account for the background noise or other signals present in the spectrogram, the peaks are often grouped on the same frequency component such as with the long segment analyses. For example, time-step 22 most likely has three frequency components but two of them overlap on the same frequency component. In order to compensate for our models not capturing the signal complexity, we specify the noise component in the model. We simply do not marginalise the noise term σ . The value of the noise component was selected by inspection, but this can be done by way of regularization or machine-learning techniques. The result of choosing a noise component by inspection can be seen in Figure 4.15 and 4.16. The simpler frequency models are now selected more often in comparison.

4.8 Signal Complexity

In Chapter 2 we discussed the various effects present in the Doppler radar signal, namely the multi-path effect, the stick-flash phenomenon, the fact that objects reflect multiple or a range of sinusoids, and the fact that the signals measured are the radial speeds of the objects relative the Doppler radar unit. We did not expect our BSA models to be perfect fits for the signals, but the complex models were consistently selected as the most likely model and with very high certainty. Increasing the noise component in the BSA models mitigates the problem but only to a limited degree. Over longer periods the signal displays more complex behaviour despite being segmented into parts where there is seemingly little complexity in the spectrogram.

The more complex models being selected is not a failure of BSA but caused by insufficiently complex signal models for the real-world signal. As can be seen in Appendix C, simulated signals can easily be modelled and will give accurate results. We could continue to define increasingly more complex signal

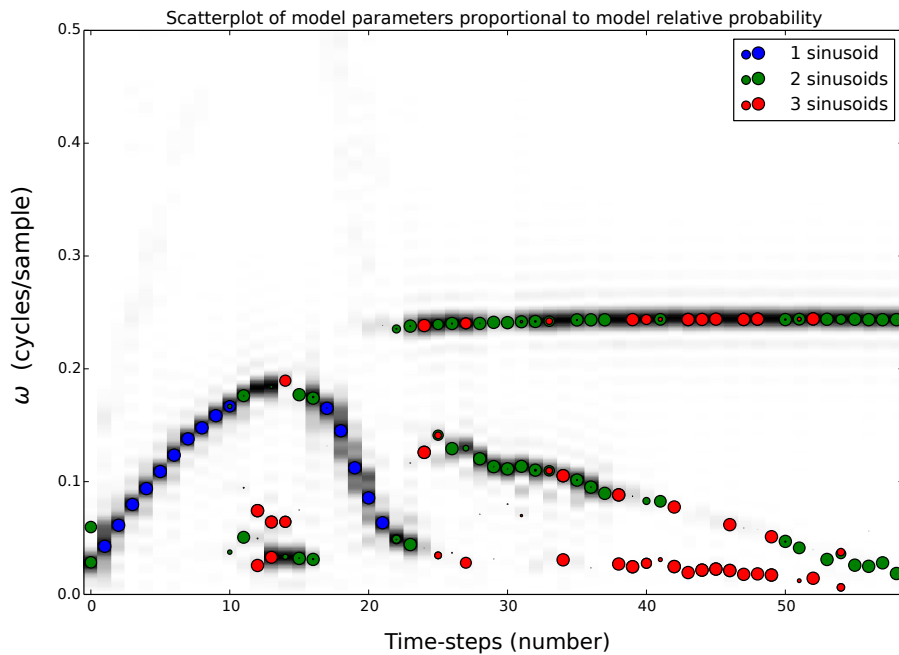


Figure 4.15: Scatter-plot of BSA stationary sinusoid frequency values with the specified noise BSA model. The size of each marker indicates the probability of the specific BSA model. Note that the less complex models are now selected more often compared to Figure 4.13.

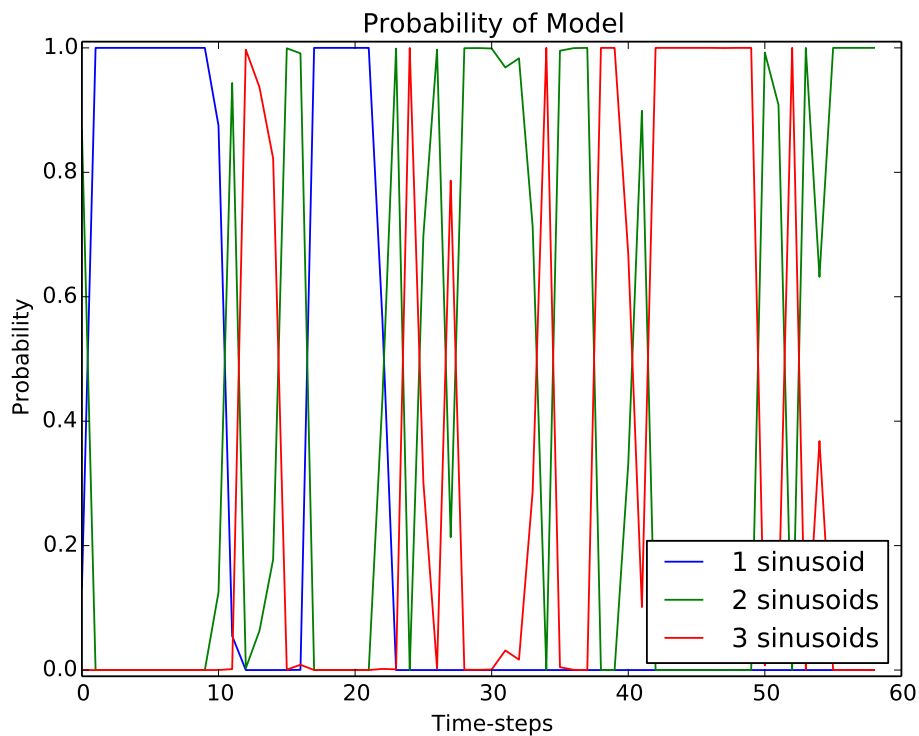


Figure 4.16: The BSA model probabilities from Figure 4.15

models until the return on investment for the model selection improvements has decreased due to the Occam factor. We might then find a model that would describe the complex signal behaviour more completely. However, without knowing more about the underlying physical signal, this might prove to be a fruitless exercise. An alternative to more complex models would be to have a better idea of the signal noise component, the part of the signal that we do not attribute to the signal of interest.

4.9 Conclusion

We analysed the measured Doppler radar signal using nine different BSA models. We used the models to calculate the most likely model parameters and to find the probability distribution of the BSA models in order to determine the most likely BSA model. Both the long and short signal analyses suggest that the Doppler radar data contain more complexity than our models are able to explain or that there are possibly multiple very close signals. The most likely models and parameters tend to be more complex than a cursory inspection of the spectrogram would suggest. In contrast the BSA models work very well on simulated data. We hypothesise that the Doppler radar signals are more complex than the current BSA models that we use. Explicitly specifying the amount of noise in the signal within our models and increasing it causes the simpler signals to be selected more often, which supports our hypothesis. This avoids having to model the higher complexity in the signal by effectively including a baseline noise as an acceptable level of error for our models. It will allow us to do better model selection in later chapters.

Despite the limitations we encountered by using our simple BSA models, they still provide us more information than a simple spectrogram would, even with the stationary frequency models. The stationary frequency models give us a probability density function over the frequency components in the form of a function that is the same dimension as the number of frequency components. This allows us to have a joint probability density function of our targets despite the measurement being a compound signal of all the target objects' frequency shifts.

Chapter 5

Bayesian State-Space Estimation

5.1 Introduction

In this chapter we will lay the groundwork for the estimation of the Doppler frequency parameters in a form to which PGMs can be applied. We discuss state-space estimation and briefly discuss a few different methods by which it can be represented. Typically, state-space estimation is done with recursive methods where, during each time-step, predictions are made for the next time-step using a stochastic transition process.

State-space estimation can be formulated as a Bayesian inference problem as shown by Stone et. al. [11]. Two examples of Bayesian approaches to well-known state-space estimation methods are for the Kalman filter in [27] and for HMMs in [28]. The difference in the Bayesian approach to traditional state-space estimation is that one needs to specify prior distributions for the model parameters explicitly, even if these priors denote one's ignorance of the random variable values. One should treat the starting state, other model parameters, and parameter transitions as prior information that is updated using likelihood functions.

The same principles of Bayesian inference discussed in Section 3.2 can now be applied with the end goal of estimating the states of the parameters of interest. Our goal is now re-framed as calculating the marginal posterior probability distributions of the parameters of interest. We will first state the single-target tracking formulation explaining the target state-space, prior information, likelihood functions, and assumptions. Where necessary we will then extend the formulation to multiple targets.

5.2 Single-Target Tracking

A target can be any object or measured phenomenon the behaviour of which we wish to track over time. For target tracking the motion of the object is modelled in continuous time, but the observations are received at discrete times. The following formulation of single-target tracking closely follows the formulation by Stone et al. from [11].

5.2.1 Prior Information

Let X_t be the unknown target state at time t starting at $t_0 = 0$. The prior information about the target is represented by a stochastic process where sample paths of this process correspond to the possible target paths through the state-space \mathbf{P} . The state-space can be discrete, continuous, or a mixture of the two.

The state of the target is typically a vector of components, for example position or velocity. In our application on Doppler radar data, it will be the BSA frequency parameters and their dynamics. In order to simplify state-space estimation problems, the assumption is made that the state-space parameters of each time-step form a *Markov chain* [11]. A Markov chain is a collection of random variables $\mathbf{X}_{t+1} = \{X_0 \dots X_{t+1}\}$ possessing the property that, given the parameters of a time-step, the parameters of the following time-step are conditionally independent of all other time-steps, that is to say

$$p(X_{t+1}|X_t X_{t-1} \dots X_0) = p(X_{t+1}|X_t).$$

We will see in Chapter 6 that this approach is complementary to PGMs that work well when a problem is broken up into local interactions between random variables. The state-space must also then be chosen so that the parameter dynamics of the target are Markovian. This implies that there must be enough parameters within the scope of the state-space to be able to give a good “snapshot” of the target behaviour during each time-step.

5.2.2 Likelihood Functions

Typically the random variables of interest, \mathbf{X}_T , are hidden and not observed directly. One needs likelihood functions that depend only on the state of the target at the time of the observation. The random variables that we observe form part of our data \mathbf{D}_T . In a state-space estimation problem, there would typically be sensors or some method of measuring the state or subset of the state of the target of interest where the measurements are made at discrete time-steps. One needs the probability of the data given the model parameters. In the general case we have a set of observations for each discrete time-step:

$$\mathbf{D}_T = \{D_0, \dots D_t, \dots D_T\}, \quad 0 \leq t \leq T,$$

namely the data observed from time-step $t_0 = 0$ to $t = T$. The values that our states at each time-step can assume are described by the random variables:

$$\mathbf{X}_T = \{X_0, \dots, X_t, \dots, X_T\}, \quad 0 \leq t \leq T.$$

Then the likelihood function for the k th time-step can be written as

$$L(X_k) = p(D_k|X_k),$$

which assumes that the observations for each time-step are independent of one another and depend only on the target state for that time-step [11].

5.2.3 Posterior

In this section we use our prior information and assumptions to calculate the posterior probability density of our hidden random variables. Since our observations at time-step k depend only on the target state at time-step k , we have [11]

$$p(\mathbf{D}_T|\mathbf{X}_T) = \prod_{k=0}^T p(D_k|X_k).$$

From the Markov chain assumption of the state transition functions we also have [11]

$$p(\mathbf{X}_T) = \prod_{k=1}^T p(X_k|X_{k-1})p(X_0).$$

The joint posterior probability density function of the data and the hidden variables is thus

$$p(\mathbf{X}_T, \mathbf{D}_T) = p(X_0)p(D_0|X_0) \prod_{k=1}^T p(D_k|X_k)p(X_k|X_{k-1}). \quad (5.2.1)$$

To calculate the marginal posterior of our states for each time-step we can use the sum and product rules from Chapter 3:

$$p(X_k|\mathbf{D}_t) = \frac{\int dX_0, \dots, dX_{k-1} p(\mathbf{D}_t|\mathbf{X}_t)p(\mathbf{X}_t)}{p(\mathbf{D}_t)}, \quad (5.2.2)$$

which is known as *filtering* [11]. In contrast, solving the posterior of the states given all the data observed:

$$p(X_k|\mathbf{D}_T) = \frac{\int dX_0, \dots, \cancel{dX_k}, \dots, dX_T p(\mathbf{D}_T|\mathbf{X}_T)p(\mathbf{X}_T)}{p(\mathbf{D}_T)}, \quad (5.2.3)$$

is known as *smoothing* [11].

5.2.4 Bayes Network Representation

While we will not implement directed graphs in this thesis, we will, in order to exercise our intuitions, represent the above prior probability and likelihood distributions in the form of a Bayes network. Most of the relationships between random variables in state-space estimation are causal and we found Bayes networks useful for thinking about and representing these causal relationships. In Figure 5.1 there is a Bayes network representation of state-space estimation which could be, for example, a Kalman filter or an HMM. The nodes are random variables and the directed lines indicate that the random variables form a parent-child relationship, which is described by a conditional probability density. The directed line from node X_t to D_t indicates the probability $p(D_t|X_t)$. The filled nodes have been observed.

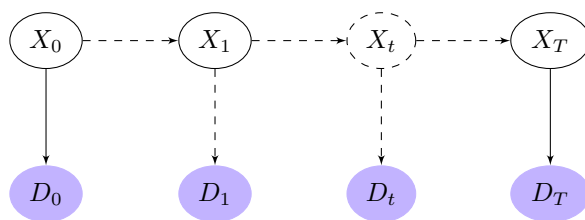


Figure 5.1: *Bayes network representation of state-space estimation.*

The Bayes network enables one to visualise the probabilistic link between the model's hidden parameters, X , and the observed random variables, D . Any nodes that do not have a parent node, such as X_0 , need a prior probability density for the graph to factorise a valid probability density. The graph in Figure 5.1 *factorises* the probability density $p(\mathbf{X}_T, \mathbf{D}_T)$. This means that multiplying the conditional probability densities together will give one the joint probability density. The directed lines between the nodes $X_0 \rightarrow X_1 \rightarrow X_t$ indicate that X_t is conditionally independent of X_0 given X_1 as per our Markov chain assumption.

5.3 Multitarget Tracking

The multitarget tracking generalisation also closely follows the formulation from [11]. For multitarget tracking, we may have an unknown number of targets bounded by a maximum number \mathcal{J} . The state-space now contains the states for each target $\mathbf{P} = \{P^0, P^1, \dots, P^{\mathcal{J}}\}$ where previously there was only one target state per time-step. We need to add an additional value, ϕ , to each sub-state-space P to denote when a target is not present in the region of interest. For example, if we have a maximum of two targets where we track their positions $\{x_0, x_1\}$ and only one target is present, we could have

$\{x_0 = 1, x_1 = \phi\}$ [11].

The extended sub-state-space is $P^+ = P \cup \{\phi\}$. The state of the system at time t would now be $X_t = \{x_{t,1}, \dots, x_{t,\mathcal{J}}\}$ where $x_{t,j} \in P^+$ is the state of target j at time t and the entire joint state-space would be $X_t \in \mathbf{P}^+ = P^+ \times \dots \times P^+$. Unlike the single-target case, the likelihood functions can now contain the joint probability of all the targets. The same calculations from the single target case apply for the posterior distribution of the target state-space.

5.4 Relation to Kalman Filters

In Kalman filters, one represents the stochastic transitions between states (where we ignore constant offsets for the sake of simplicity) as

$$X_t = \Lambda X_{t-1} + \mathbf{w}_t, \quad (5.4.1)$$

and the observations as

$$D_t = \mathcal{H}X_t + \mathbf{u}_t,$$

where \mathbf{w}_t and \mathbf{u}_t are zero-mean multivariate Gaussian noise. The matrices Λ and \mathcal{H} describe the linear relationship between the parameters.

According to Stone et. al. [11], Kalman filtering is the result of viewing tracking as a least squares problem. Recursive methods are used to solve the problem with two phases for each iteration, namely update and predict. In the update phase, the data D_t for the current time-step is incorporated. In the predict phase, the parameters of the next time-step are predicted given the updated state and the linear dynamics. When using Gaussian probability densities to describe X_t and D_t in the above single-target tracking Bayesian state-space estimation, the mean and covariance of the posterior Gaussians for X_t are identical to the mean and covariance of the least squares solution produced by the Kalman filter [11].

5.5 Discrete Markov Chains

Discrete state-space transitions have discrete probabilities that can be represented in a directed graph (not to be confused with Bayes networks). We will represent these state-space graphs using nodes with double lines to avoid mistaking them for Bayes networks. These directed graphs can be interpreted as finite state-machine transitions where the nodes are possible states that the model can assume during a time-step, and the directed edges indicate the possible state transitions.

5.5.1 Uninformative Transition Structure

In an uninformative discrete Markov chain, any state can go to any state with equal probability, including itself. An example model can be seen in Figure 5.2 with three equally likely states. If the transition probabilities were uniform, the model would be uninformative and equivalent to a uniform prior for the state values.

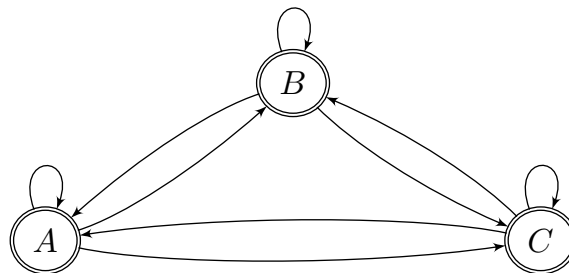


Figure 5.2: *An uninformative discrete Markov chain. The model is only uninformative if the three edges leaving each node are equally likely.*

5.5.2 Left-to-right Transition Structure

An example use-case for left-to-right models are HMMs in speech-recognition, where the phonemes in a word might follow a distinct order. Each state must be visited once and only once and in a strict sequence. For each time-step the transition options are to stay in the current state or to move to the next state. The system therefore only has a finite number of state switches to fit the model. Two *null* states are added in order to force the system to start and stop at the correct states, one before the first time-step and one after the last time-step.

An example left-to-right model can be seen in Figure 5.3 with the transition probabilities indicated above the edges. The same transitions are described in Table 5.1. The three states $A, B,$ and C occur in that order. In Figure 5.4 the prior for the state probabilities, without any data being observed, is graphed. This example has 30 time-steps and includes no other information other than the transition probabilities from Table 5.1.

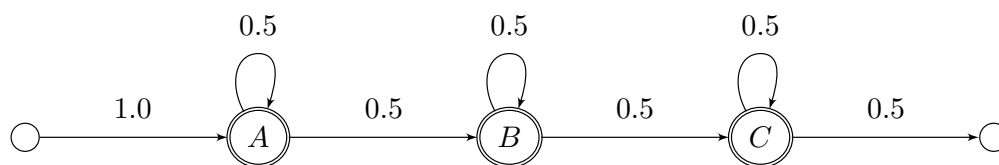


Figure 5.3: A left-to-right discrete Markov chain. States can progress in a strict order.

Table 5.1: Transition table describing the probabilities of the left-to-right discrete Markov chain from Figure 5.3.

X_t	X_{t+1}	$p(X_{t+1} X_t)$
Begin	A	1.0
A	A	0.5
A	B	0.5
B	B	0.5
B	C	0.5
C	C	0.5
C	End	0.5
Elsewhere		0.0

5.5.3 Explicit Duration Transition Structure

Explicit duration models are also employed in HMMs such as the Ferguson model from [29]. Here one can have multiple states that produce the same output and are conceptually sub-states of the same state. The transition probabilities can be fine-tuned to give one an arbitrary duration model for each state depending on the number of sub-states and their probabilities. We did not make use of explicit duration models in this thesis but believe that the models that we use can benefit from their use. We refer the reader to Appendix E for more information on explicit duration models.

5.6 Overview of Other Target Tracking Works

In this section we briefly explore different approaches to multiple target tracking, but do not go into detail about their implementation.

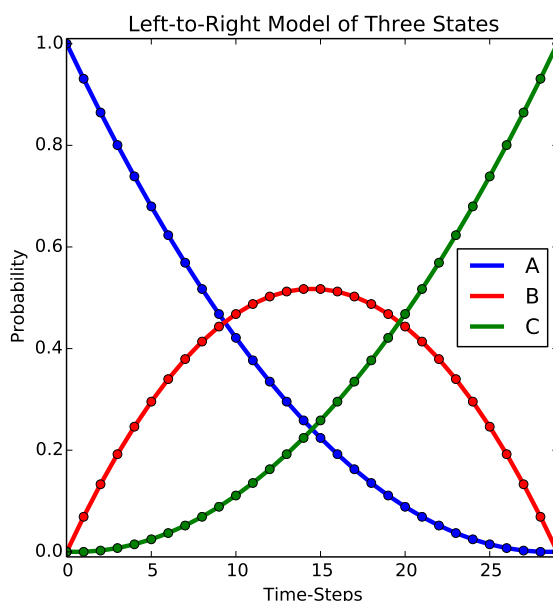


Figure 5.4: An example of the left-to-right model probabilities using the transitions in Table 5.1 for 30 time-steps. We plot the probabilities where no other information is known about the states other than the prior transitions.

5.6.1 Switching State-space Estimation

Switching state-space estimation models (SSSEMs) [18], also known as switching linear dynamical systems, combine discrete and continuous parameters. An example Bayes network representation of an SSSEM can be seen in Figure 5.5. The continuous parameters, X_t and D_t , are now conditioned on the discrete parameters A_t . These discrete random variables are typically used either for *data association*, to determine which object an observation is allocated to for multitarget tracking, or *dynamic switching*, to determine which set of dynamic behaviour should describe the transitions of the targets [18]. There are many different recursive methods used to perform multitarget tracking using SSSEMs, such as multiple-hypothesis tracking (MHT) [11]. Some methods of SSSEM do not consider all the different hypotheses at each time-step; instead they only track the most likely hypothesis and switch between them if the observations warrant it [14].

5.6.2 Track-oriented Multiple Hypothesis Tracking

One variation of MHT using data association is track-oriented multiple hypothesis tracking (TOMHT) [12]. This method keeps a pruned set of potential “tracks” where each track is a target data association hypothesis. Over time the tracks form a track tree of the different hypotheses, which contains dif-

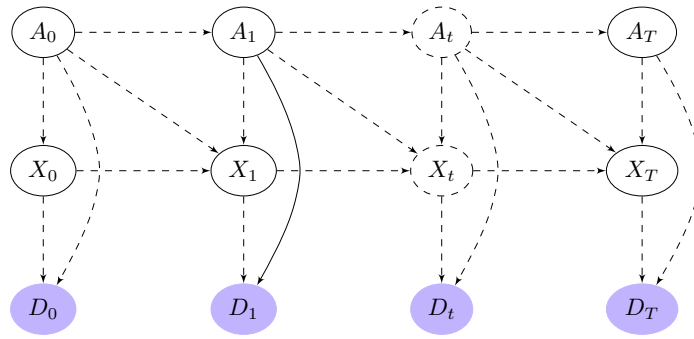


Figure 5.5: *Switching state-space estimation model*

ferent explanations for the observed data. For each time-step TOMHT solves a constrained optimisation problem in order to calculate the hypothesis that most likely fits the data. For each track there is a single target-tracking algorithm which tracks a target based on each of the different data association hypotheses. Over time these trees are then pruned by keeping only the most likely tracks for reasons of computational tractability. Target “births” and “deaths” are also tracked as hypotheses. A target birth is the addition of another target to track, and a target death is a target that has left the region under consideration. In [12] these track trees are implemented using factor graphs.

5.6.3 Joint Multitarget Probability Density Tracking

An alternative to assigning SSSEM or to have different tracks is to describe the joint probability density of the targets at each time-step, such as in joint multitarget probability density tracking (JMPDT) [19]. The targets have separate measurements, but do not have identities associated with them as in data association switching. Targets are tracked in the joint probability of all the targets under consideration. Target deaths and births are modelled within the transitions between time-steps. JMPDT [19] uses a particle filter [30] to make forward predictions of what the joint density of the targets’ parameters are, and updates the prediction using observations of the target positions. The Bayes network representation of the JMPDT would be the same as Figure 5.1. The method performs inference for the marginal probability densities at each time-step, effectively calculating Equation 5.2.2 similar to the Kalman filter.

5.7 Model Developed in Thesis

In this section we briefly contextualise the model we developed in this thesis. Our “targets” are the frequency components identified using BSA. We will track their behaviour and dynamics over time using state-space estimation

techniques. We will discuss this approach in further detail in Chapter 7 after discussing the necessary PGM theory in the next chapter. In Figure 5.6 we have extended the traditional state-space estimation's Bayes network from Figure 5.1 with an additional discrete Markov chain (indicated in red). We will use the additional Markov chain to enforce a left-to-right transition of the signal segments discussed in Section 4.6. We segmented the signal into different parts that represent the different number of frequency components and different signal dynamics. Although our model in Figure 5.6 appears

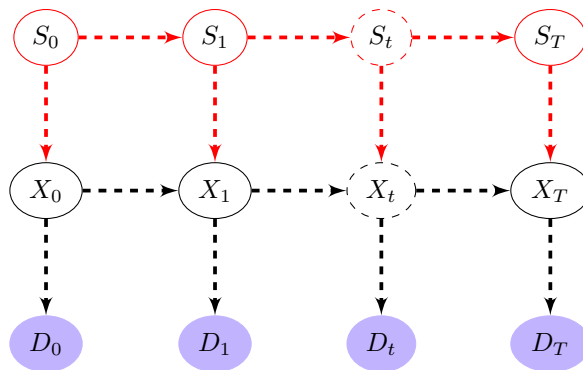


Figure 5.6: *Bayes network as a conceptual representation of the model developed in this thesis. An additional discrete Markov chain (indicated in red) is added to incorporate additional temporal information.*

similar to the SSSEM from Figure 5.5, it has subtle differences. Note that the data is not conditioned on the discrete nodes as with SSSEM. The SSSEM is used to perform dynamic switching or data association. We will not need data association as we have the joint probability density of all the targets of interest. Dynamic switching can be handled in the lower half of Figure 5.6 without branching out the discrete random variables as is done by SSSEM. In our model the discrete Markov chain will be an additional level of “switching” that is conceptually a level above the multitarget dynamics. It will allow us to add additional temporal information to our analyses and describe the evolution of the Doppler radar signal over time.

5.8 Conclusion

In this chapter we demonstrated the formulation of state-space estimation as a Bayesian inference problem. The important components necessary to describe one's models are the prior information which includes the Markov chain assumption of state transition, and the form of the likelihood functions which allow one to incorporate data such as in Bayes' theorem. We explored the

problem of single target tracking and extended it to multitarget tracking. We demonstrated a few discrete state-transition structures such as the left-to-right model, which will be useful for fixing the order of signal segment progression for our Doppler radar data. Formulating state-space estimation as Bayesian inference will allow us to solve state-space estimation problems using Bayesian inference techniques such as PGMs, where we will solve Equation 5.2.3. The model that we develop in Chapter 7 was contextualised with reference to other works in the literature.

Chapter 6

Probabilistic Graphical Models

6.1 Introduction

Probabilistic graphical models (PGMs) are declarative representations of statistical models. The graphs can compactly represent statistical relations between random variables by encoding their independence assumptions and factorising their joint probability density function. PGMs can be used for representation, inference, and learning of statistical models [31].

Inference is done in PGMs by algorithms that calculate the marginal probability densities of the joint probability. This is beneficial when only the marginal probabilities are of interest or when calculating the joint probability is intractable. An example of a marginal probability density of interest would be Equation 5.2.3:

$$p(X_k|\mathbf{D}_T) = \frac{\int dX_0, \dots, dX_k, \dots, dX_T p(\mathbf{D}_T|\mathbf{X}_T)p(\mathbf{X}_T)}{p(\mathbf{D}_T)}.$$

The graphs in PGMs can be directed, undirected, or a mixture of the two. In both graph types, nodes correspond to random variables, and edges correspond to the interactions between the random variables. We will examine the representation and inference of undirected graphs (specifically clique trees and cluster graphs) in more detail in this chapter.

6.2 Conditional Independence

PGMs factorise a joint probability density, thereby encoding a probability density function over a set of random variables [32]. Conditional independence can reduce the number of interactions between random variables inside the PGMs. In many applications, the random variables of interest interact with only a few other random variables directly. This can allow for computational tractability

and compact representations within PGMs. In Chapter 5 we discussed state-space estimation, which makes use of the fact that it is often easier to encode the local interactions as opposed to the joint probability density. These local interactions can be in the form of conditional probability densities (CPDs) that encode the interactions and conditional independences between random variables.

One can use the product rule from Equation 3.2.1, $p(A, B) = p(A|B)p(B)$, to factorise joint probability densities into a product of CPDs. An example of conditional independence is two random variables A and B that are conditionally independent given the random variable C so that

$$p(A|B, C) = p(A|C).$$

If each random variable could assume N discrete values, the CPD $p(A|B, C)$ would have N^3 different probabilities. The CPD $p(A|C)$, however, only has N^2 different probabilities. The CPDs in the joint probability $p(A, B, C)$ can be described more compactly using conditional independence:

$$p(A|B, C)p(B|C)p(C) = p(A|C)p(B|C)p(C).$$

6.3 Clique Trees and Cluster Graphs

We briefly introduced a type of directed graph, Bayes networks, in Chapter 5 that was used to exercise our intuition about state-space estimation models. In this section we introduce the concepts necessary for the representation and inference in the undirected graphs, clique trees and cluster graphs.

6.3.1 Nodes and Edges as Factors and Sepsets

In clique trees the relationships between random variables are described by more generalised CPDs called *factors*. These factors factorise the graph but can also represent the intermediate, unnormalised steps in our probability calculations. The different factors we use in this thesis are discrete, continuous, and hybrid factors. The nodes in our graphs, called *cliques*, contain one or more factors and are connected to each other with undirected edges containing a non-empty subset of random variables called separator sets or *sepsets*, \mathcal{S} [14]. The sepsets allow the knowledge about the random variables they contain to be shared between cliques. In Section 6.3.5 we discuss how each clique, \mathcal{C} , can pass a *message*, δ , to a neighbour along its edges under the right circumstances, consisting of their own and their other neighbours' beliefs. These messages are also encoded as factors.

6.3.2 Querying Graph Beliefs

The inference algorithms performed on the graphs change each clique's *belief* \mathcal{B} about its random variables [14]. In Section 6.3.9 we will discuss how these beliefs become congruent with the other cliques' beliefs. The graph is then said to have *converged*. Beliefs are also represented by factors and when normalised they are CPDs. Calculating the marginal probability density of a cluster's belief after the PGM has converged is known as *querying* the PGM. Querying the converged PGM results in the marginal of the joint probability density of the random variables factorised by the PGM. One can only query a PGM on the random variables, or a subset thereof, that are contained within each of the clusters.

6.3.3 Graph Structure and Definition

Clique trees have the restriction of being acyclic and cannot contain loops. The graphs in Figure 6.1a and Figure 6.1b factorise the same probability density. The first contains no loops and the second contains a single loop. The loopy generalisation of a clique tree is called a *cluster graph*. Instead of cliques, the nodes are called *clusters*.

The more formal definition of a cluster graph paraphrased from [14] is the following:

Cluster Graph: A cluster graph, \mathcal{G} , for a set of factors, $\{\Phi_0, \dots, \Phi_n\}$, over the set of random variables, \mathbf{X} , is an undirected graph where each node, which is a cluster \mathcal{C}_i , is associated with a subset of \mathbf{X} .



Figure 6.1: Both PGMs factorise the same joint probability density, $p(A, B, C, D)$, which is the normalised product of the factors: $\Phi_0, \Phi_1, \Phi_2, \Phi_3$.

In this thesis we will refer to both clique trees and cluster graphs as cluster graphs unless explicitly emphasising a graph's tree structure.

6.3.4 Family Preservation Property

The family-preservation property of cluster graphs requires that each factor Φ must be associated with a cluster \mathcal{C} such that the scope of the factor is a subset of the cluster, namely $Scope[\Phi] \subseteq \mathcal{C}_i$ [14]. In this thesis the clusters will consist of a single factor each.

6.3.5 Belief Propagation

Pearl [33] introduced belief propagation (BP), which is an iterative algorithm. It consists of a set of local message-passing rounds that continue until all nodes agree on their shared belief. The algorithm is also known as sum-product message passing [14]. In tree-structured graphs such as clique trees, the BP message-based marginals converge to the exact marginals.

The message $\delta_{i \rightarrow j}$ from cluster \mathcal{C}_i to \mathcal{C}_j is the cluster \mathcal{C}_i multiplied by its incoming messages, excluding the message that \mathcal{C}_j sends to \mathcal{C}_i . The message is marginalised to the scope of the sepset, $\mathcal{S}_{i,j}$, which is equal to, or a subset of, the intersection between the clusters:

$$\mathcal{S}_{i,j} \subseteq \mathcal{C}_i \cap \mathcal{C}_j. \quad (6.3.1)$$

We can formally define a message passed using BP as:

$$\delta_{i \rightarrow j} = \sum_{\mathcal{C}_i \setminus \mathcal{S}_{i,j}} \left(\Phi_i \times \prod_{k \in (Nb_i \setminus \{j\})} \delta_{k \rightarrow i} \right). \quad (6.3.2)$$

In the above equation we use the notation “ \setminus ” for the set difference and “ Nb ” for neighbouring factors. The scope of the random variables that are marginalised, $\mathcal{C}_i \setminus \mathcal{S}_{i,j}$, is thus the random variables in \mathcal{C}_i excluding the random variables in $\mathcal{S}_{i,j}$. The values that k assume in the product are the indices of the clusters neighbouring the cluster \mathcal{C}_i , Nb_i , excluding the index j . An outgoing message from cluster i to j is the product of the factors in the cluster and all incoming messages, excluding the message from j to i , marginalised to the scope of the sepset. An outgoing message therefore does not contain the information from the cluster that it is sending its message to.

An example of a message being passed can be seen in Figure 6.2, where a message $\delta_{3 \rightarrow 4}$ is being sent from \mathcal{C}_3 to \mathcal{C}_4 . The message can only be sent if the incoming messages $\delta_{0 \rightarrow 3}$, $\delta_{1 \rightarrow 3}$, and $\delta_{2 \rightarrow 3}$ are defined. The outgoing message can be calculated as the product of the factor in \mathcal{C}_3 with the incoming

messages from the left. Any random variables not in the scope of the sepset are marginalised, giving us:

$$\delta_{3 \rightarrow 4} = \sum_{B,C} (\Phi_3 \times \delta_{0 \rightarrow 3} \times \delta_{1 \rightarrow 3} \times \delta_{2 \rightarrow 3}).$$

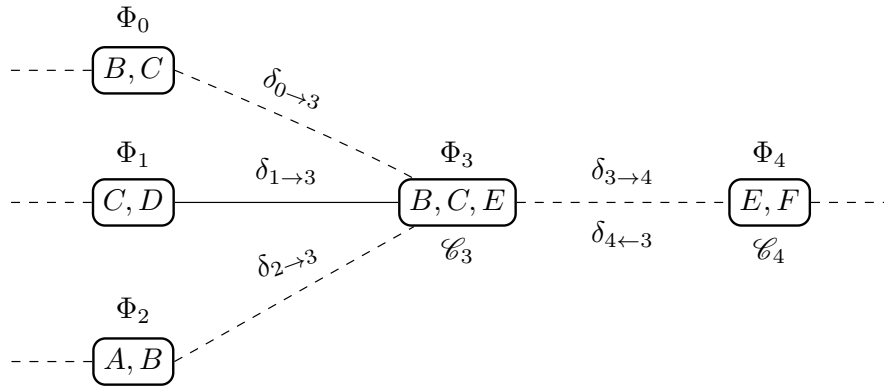


Figure 6.2: An illustration of the message-passing algorithm used in BP. The cluster \mathcal{C}_3 sends a message to the cluster to its right using the incoming messages from the left.

In clique tree graphs one starts passing messages from the leaf nodes (the outer clusters with only a single edge) of the graph, and then messages propagate inwards towards the root cluster. A cluster can only pass a message when all other incoming messages have been received. Once the root cluster has the messages needed to propagate its own message back, the flow of messages returns in the opposite direction towards the leaf nodes. In general, any node can be selected as the root cluster.

6.3.6 Belief Update

In the belief update (BU) algorithm (also known as the Lauritzen-Spiegelhalter algorithm) [34], there is a subtle yet important difference to BP when calculating the messages. For the message, $\delta_{i \rightarrow j}$ from cluster \mathcal{C}_i to \mathcal{C}_j , we multiply in all incoming messages and then divide out $\delta_{j \rightarrow i}$ instead of leaving it out of the calculation. Mathematically the two methods appear identical, as dividing out a factor which has been multiplied in should have no effect. BU messages are defined as:

$$\delta_{i \rightarrow j} = \left(\sum_{\mathcal{C}_i \setminus \mathcal{S}_{i,j}} \Phi_i \times \prod_{k \in (Nb_i)} \delta_{k \rightarrow i} \right) / \delta_{j \rightarrow i}. \tag{6.3.3}$$

Taking the same example from Section 6.3.5 for Figure 6.2 the message will be calculated as

$$\delta_{3 \rightarrow 4} = \left(\sum_{B,C} \Phi_3 \times \delta_{0 \rightarrow 3} \times \delta_{1 \rightarrow 3} \times \delta_{2 \rightarrow 3} \times \delta_{4 \rightarrow 3} \right) / \delta_{4 \rightarrow 3}.$$

According to [14], BP and BU are equivalent in the exact case – however, as we will discuss in Section 6.6, when making approximations in the factor operations they give different results.

6.3.7 Loopy Belief Propagation/Update

In cyclic graphs one can perform *loopy* belief propagation or update (LBP/LBU). Just as in BP/BU, messages are sent between clusters but now messages can be sent without necessarily requiring the other incoming messages to do so. To be able to send a message from a cluster which is not a leaf node and which has not received any incoming messages, all messages are initialised to be uninformative or *vacuous* before message passing begins. Vacuous messages are equivalent to multiplying a factor by 1. This method of message passing does not necessarily compute exact marginal probabilities and is a form of approximate inference [14]. For graphs that contain loops, there is no guarantee of convergence [14]. Despite the lack of guarantee it has been empirically shown, such as in [35], that these graphs still give good performance in many cases.

Any cluster can in theory pass a message to any other cluster in any order in LBP/LBU. However, scheduling the order in which messages are propagated can affect the speed of convergence and even the convergence point if conflicting messages are passed in a cyclical manner.

6.3.8 Running Intersection Property

The running intersection property (RIP) of cluster graphs requires precisely one direct path for each pair of clusters that information about any particular random variable can take. The cluster graph satisfies RIP if for each variable $X \in \mathbf{X}$, such that $X \in \mathcal{C}_i$ and $X \in \mathcal{C}_j$, there is exactly one path between \mathcal{C}_i and \mathcal{C}_j for which $X \in \mathcal{S}_{l,k}$ for all edges between clusters, $(\mathcal{C}_i, \dots, \mathcal{C}_l, \mathcal{C}_k, \dots, \mathcal{C}_j)$, in the path [14].

RIP implies that all edges associated with X form a tree that spans all the clusters that contain X . The fact that some path of information about X must exist between the clusters containing it, forces information to flow between these clusters and they must agree on the marginal of X when the graph is converged. This also mitigates information cycling in endless loops due to contradictory cyclic messages. Variables that are strongly correlated can still

allow strong cyclic arguments to occur in loopy graphs despite adhering to RIP. An example is if in Figure 6.1b the random variable pairs $A - B$, $B - C$, $C - D$ are strongly positively correlated but $A - D$ are strongly negatively correlated.

An illustration of RIP can be seen in Figure 6.3 where we indicate the sepsets between the clusters with sharp-edged rectangles. For example, examining the random variable B , if there is a path from the cluster $\mathcal{C}_{A,B}$ to $\mathcal{C}_{B,C,F}$ (indicated in blue) there may not be another path (indicated in red) without violating RIP.

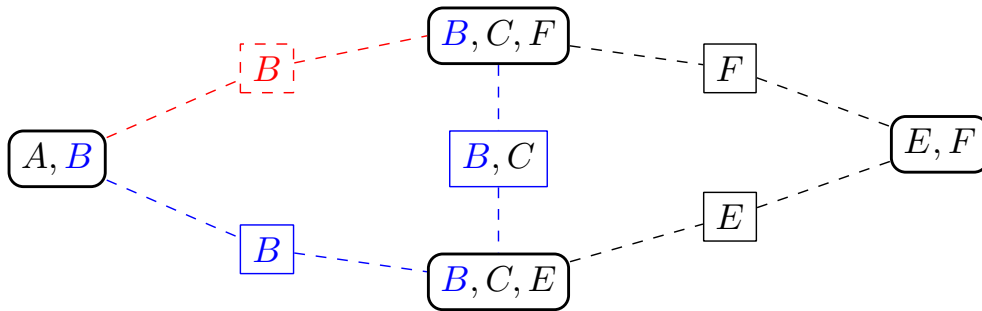


Figure 6.3: An illustration of the RIP. If the sepset (indicated in blue) is present, the sepset (indicated in red) will violate RIP if added.

6.3.9 PGM Beliefs and Convergence

The belief \mathcal{B}_i of a cluster \mathcal{C}_i is defined as the product of the initial factor and all incoming messages from each neighbouring cluster \mathcal{C}_k :

$$\mathcal{B}_i = \Phi_i \times \prod_{k \in Nb_i} \delta_{i \rightarrow k}. \tag{6.3.4}$$

As the messages are passed in a PGM using LBP/LBU, the beliefs of each cluster will change over time. A cluster graph is converged if for each edge connecting the clusters we have:

$$\sum_{\mathcal{C}_i \setminus \mathcal{S}_{i,j}} \mathcal{B}_i = \sum_{\mathcal{C}_j \setminus \mathcal{S}_{i,j}} \mathcal{B}_j. \tag{6.3.5}$$

If the cluster graph satisfies RIP, the marginal belief of any random variable is identical in all clusters that contain it. Querying a cluster will give \mathcal{B}_i or some subset of the belief [32].

6.3.10 Message Scheduling

For clique trees using BP/BU, the order in which messages are sent is constrained by which clusters are able to send a message. In contrast, in cluster graphs using LBP/LBU any cluster can generally send a message at any time. There are many different methods for scheduling the message-passing order that will have an effect on whether and how quickly a graph converges. Synchronous BP passes messages from all clusters simultaneously for each iteration of message passing [14]. Asynchronous message passing sends a single message per iteration depending on the message scheduling. One scheduling method is passing messages via *round-robin* scheduling where the messages are updated, one after the other, in the same arbitrary order for each iteration of message passing [33]. Another method of scheduling is to propagate a message as soon as one of the input messages has changed [35]. Residual belief propagation (RBP) in Elidan et al. [35] updates the message with the highest *message residual*, the distance between the current value of the message, and the value of the message after it is updated.

Elidan [35] shows that RBP converges significantly more reliably than synchronous and round-robin message passing and has shorter convergence times. The intuitive argument behind RBP is that not all messages are equally useful toward achieving convergence. If a message has not changed its value significantly since the previous iteration, it would be redundant to pass the message again, while sending a message that has a greater difference to its previous value is more likely to be informative to the cluster it is sending a message to [35]. Although the word propagation is in the method name, it can be applied to both LBP and LBU.

6.3.11 Message Damping

Message damping is a process that smooths messages between message-passing iterations [14]. The message being sent, $\delta_{i \rightarrow j}$, is the average between the previous message, $\delta_{i \rightarrow j}^{\text{old}}$, and the new message, $\delta_{i \rightarrow j}^{\text{new}}$, giving us

$$\delta_{i \rightarrow j} = \lambda \delta_{i \rightarrow j}^{\text{new}} + (1 - \lambda) \delta_{i \rightarrow j}^{\text{old}}.$$

The damping severity is regulated by choosing $0 < \lambda \leq 1$. When $\lambda = 1$ there is no damping, but for other values of λ the messages are partially updated. Damping messages reduces the fluctuations in messages being passed between clusters in graphs with many loops [14].

6.4 Discrete Table Factors

Representing a discrete factor can be done in a table which lists the probabilities of each of the combinations of possible values that the random variables

can assume. Despite the computational limitations, discrete table representations are still very useful if the number of random variables and possible values is kept low.

6.4.1 Representation

With n variables, and $\mathfrak{K} = |\text{dom}(X_i)|$ possible values for each, this requires specifying \mathfrak{K}^n numbers which can cause the representation of a joint probability to become very large. By not representing zero probabilities in the table, one has a sparser representation. Encoding conditional independences can further help reduce the number of probabilities to be specified, but the factors can still be too large to perform inference in a reasonable time.

Table 6.1 is an example of a discrete factor representing the CPD $p(A, B|C)$. For it to be a valid CPD, the probabilities for each value of C need to sum to 1 for the factor to be normalised. The discrete factor can also represent the intermediate, unnormalised steps where the factor operations are being performed.

Table 6.1: *An example discrete factor representing $p(A, B|C)$.*

A	B	C	$p(A, B C)$
0	0	0	0.25
1	0	0	0.45
1	1	0	0.30
0	0	1	0.15
0	1	1	0.25
1	1	1	0.60
Elsewhere			0.00

6.4.2 Operations

In order to perform inference on PGMs with the discrete factors we need to define the following operations.

Multiplication Multiplication is performed by calculating the product of the probabilities where the values of the shared random variables match. Random variables that are not shared between the factors create a Cartesian product of the new combinations.

Division This is similar to multiplication where random variables of the same scope's probabilities are divided. Typically one does not divide by random variables outside the shared scope. The operation is not defined if the denominator is zero and the numerator is not.

Marginalisation Marginalising a random variable in a discrete factor is done by summing all the different entries for the random variable to one entry for each combination of the remaining random variables.

Normalisation To normalise a discrete factor, divide all probabilities in the table by the sum of the table so that the table then sums to 1.

Message Residual The software used for the discrete factors calculates the distance between factors using Kullback-Leibler divergence [36]. For two discrete factors Φ_0 and Φ_1 , from [25], the Kullback-Leibler divergence of Φ_1 from Φ_0 is defined to be

$$d_{\text{KL}}(\Phi_0||\Phi_1) = \sum_i \Phi_0(i) \log \left(\frac{\Phi_0(i)}{\Phi_1(i)} \right).$$

The above equation is a measurement of the departure of the density described by Φ_1 from Φ_0 and is non-symmetric [25].

6.5 Continuous Gaussian Factors

In contrast with discrete factors where we need to specify any non-zero probability, we can represent probabilities using parameterised continuous functions. If the probability densities of interest are not parameterised, a trade-off can be made between approximating them with parameterised functions or using discretising techniques. The Gaussian probability density function is commonly used to approximate uni-modal probability density functions. The other advantage is that a Gaussian function is described by only two parameters, the mean and covariance matrix.

6.5.1 Representation

For a Gaussian probability density function over \mathbf{X} we have

$$p(\mathbf{X}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{X} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) \right].$$

The probability density function is parameterised by the mean, $\boldsymbol{\mu}$, and the covariance matrix, Σ , and is represented in the covariance form as

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma).$$

An alternative representation is the *canonical* form:

$$\mathcal{N}(\boldsymbol{\mu}, \Sigma) = \mathcal{C}(\mathbf{X}; K, \mathbf{h}, g).$$

The canonical parameters are calculated by taking the log of the probability density function $p(\mathbf{X})$ and are:

$$\begin{aligned} K &= \Sigma^{-1}, \\ \mathbf{h} &= \Sigma^{-1}\boldsymbol{\mu}, \\ g &= -\frac{1}{2}\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu} - \log((2\pi)^{n/2}|\Sigma|^{1/2}), \end{aligned} \quad (6.5.1)$$

where n is the number of random variables. The different representations are useful for different operations performed on the Gaussian probability density function.

The vacuous canonical form, which is analogous to an improper probability density function of 1 over all random variables, is defined as

$$\begin{aligned} K &= \text{all } 0, \\ \mathbf{h} &= \text{all } 0, \\ g &= 0. \end{aligned}$$

When a factor is multiplied with or divided by a vacuous factor it has no effect.

6.5.2 Operations

Some operations on Gaussian functions are easier to perform in canonical form, and some in covariance form. For the purposes of this thesis we perform all operations in canonical form to avoid performing matrix inversions on large matrices. The derivations for the following operations can be found in [37].

Multiplication

$$\mathcal{C}(\mathbf{X}; K_1, \mathbf{h}_1, g_1) \times \mathcal{C}(\mathbf{X}; K_2, \mathbf{h}_2, g_2) = \mathcal{C}(\mathbf{X}; K_1 + K_2, \mathbf{h}_1 + \mathbf{h}_2, g_1 + g_2)$$

Factors of different scopes are extended by adding zeros to the K matrices and \mathbf{h} vectors, and then performing the above.

Division

$$\frac{\mathcal{C}(\mathbf{X}; K_1, \mathbf{h}_1, g_1)}{\mathcal{C}(\mathbf{X}; K_2, \mathbf{h}_2, g_2)} = \mathcal{C}(\mathbf{X}; K_1 - K_2, \mathbf{h}_1 - \mathbf{h}_2, g_1 - g_2)$$

Division is done similarly to multiplication but is only defined if the numerator's scope is equal to or larger than the scope of the denominator.

Normalisation To normalise a Gaussian function in canonical form, g is recalculated using Equation 6.5.1.

Message Residual For the message residual, one can use the Kullback-Leibler divergence such as for the discrete factor. The software we developed for this thesis did not yet have the Kullback-Leibler divergence functionality completed for the Gaussian factors. Instead we used the sum of the squared element-wise differences between the K , \mathbf{h} , and g parameters of the two factors:

$$d = \sum |K_0 - K_1|^2 + \sum |\mathbf{h}_0 - \mathbf{h}_1|^2 + |g_0 - g_1|^2.$$

Marginalisation Marginalisation is easily done in covariance form as one simply reduces the scope of the Σ and $\boldsymbol{\mu}$ to the new subset of random variables:

$$\mathbf{X}, \mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma).$$

$$\int d\mathbf{Y} \mathcal{N}(\boldsymbol{\mu}, \Sigma) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}}, \Sigma_{\mathbf{X}\mathbf{X}}),$$

where we have

$$\Sigma = \begin{bmatrix} \Sigma_{\mathbf{X}\mathbf{X}} & \Sigma_{\mathbf{X}\mathbf{Y}} \\ \Sigma_{\mathbf{Y}\mathbf{X}} & \Sigma_{\mathbf{Y}\mathbf{Y}} \end{bmatrix}; \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{Y}} \end{pmatrix}.$$

One can also perform marginalisation in the canonical form:

$$\int d\mathbf{Y} \mathcal{C}(\mathbf{X}, \mathbf{Y} : K, \mathbf{h}, g) = \mathcal{C}(\mathbf{X}; K', \mathbf{h}', g'),$$

with

$$\begin{aligned} K' &= K_{\mathbf{X}\mathbf{X}} - K_{\mathbf{X}\mathbf{Y}} K_{\mathbf{Y}\mathbf{Y}}^{-1} K_{\mathbf{Y}\mathbf{X}}, \\ \mathbf{h}' &= \mathbf{h}_{\mathbf{X}} - K_{\mathbf{X}\mathbf{Y}} K_{\mathbf{Y}\mathbf{Y}}^{-1} \mathbf{h}_{\mathbf{Y}}, \\ g' &= g \frac{1}{2} (\log |2\pi K_{\mathbf{Y}\mathbf{Y}}^{-1}| + \mathbf{h}_{\mathbf{Y}}^T K_{\mathbf{Y}\mathbf{Y}}^{-1} \mathbf{h}_{\mathbf{Y}}). \end{aligned} \quad (6.5.2)$$

where, the matrices, $K_{\mathbf{Y}\mathbf{Y}}$, $K_{\mathbf{X}\mathbf{Y}}$ are submatrices of the K matrix – namely, it is reduced to the scope of the random variables in the subscripts:

$$K = \begin{bmatrix} K_{\mathbf{X}\mathbf{X}} & K_{\mathbf{X}\mathbf{Y}} \\ K_{\mathbf{Y}\mathbf{X}} & K_{\mathbf{Y}\mathbf{Y}} \end{bmatrix}; \quad \mathbf{h} = \begin{pmatrix} \mathbf{h}_{\mathbf{X}} \\ \mathbf{h}_{\mathbf{Y}} \end{pmatrix}. \quad (6.5.3)$$

6.5.3 Linear Gaussian

Gaussian factors would not be as useful for state-space estimation as they are if we could not capture the conditional dependence between random variables. By using *linear Gaussians* one can capture the linear dependence between random variables. For the two multivariate Gaussian random variables \mathbf{Y} and \mathbf{X} , with lengths m and n respectively, one can capture the linear relationship between them as

$$\mathbf{X} = \Lambda \times \mathbf{Y} + \mathbf{b} + \mathbf{w}.$$

The matrix Λ ($m \times n$) and vector \mathbf{b} (n) describes the linear relationship between \mathbf{X} and \mathbf{Y} . The vector \mathbf{w} is the zero mean *process noise* that has a covariance matrix $\Sigma_{\mathbf{w}\mathbf{w}}$. Without the process noise, the relationship between \mathbf{Y} and \mathbf{X} is no longer stochastic and causes the density function to be singular. This is how a Kalman filter represents the linear relationships between parameters such as those discussed in Section 5.4.

Writing \mathbf{Y} as

$$p(\mathbf{Y}) = \mathcal{N}(\mathbf{Y}; \boldsymbol{\mu}_{\mathbf{Y}}, \Sigma_{\mathbf{Y}\mathbf{Y}})$$

and writing \mathbf{X} conditioned on \mathbf{Y} we have:

$$p(\mathbf{X}|\mathbf{Y}) = \mathcal{N}(\mathbf{X}; \boldsymbol{\mu}_{\mathbf{X}|\mathbf{Y}}, \Sigma_{\mathbf{X}|\mathbf{Y}})$$

or

$$p(\mathbf{X}|\mathbf{Y}) = \mathcal{N}(\mathbf{X}; \Lambda\mathbf{Y} + \mathbf{b}, \Sigma_{\mathbf{w}\mathbf{w}}).$$

From [37] we can write their joint probability density function $p(\mathbf{Y}, \mathbf{X})$ as:

$$p(\mathbf{Y}, \mathbf{X}) = \mathcal{N}(\mathbf{Y}, \mathbf{X}; \boldsymbol{\mu}_{\mathbf{Y}\mathbf{X}}, \Sigma_{\mathbf{Y}\mathbf{X}})$$

with

$$\boldsymbol{\mu}_{\mathbf{Y}\mathbf{X}} = \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{Y}} \\ \Lambda\boldsymbol{\mu}_{\mathbf{Y}} + \mathbf{b} \end{bmatrix}$$

and

$$\Sigma_{\mathbf{Y}\mathbf{X}} = \begin{bmatrix} \Sigma_{\mathbf{Y}\mathbf{Y}} & \Sigma_{\mathbf{Y}\mathbf{Y}}\Lambda^T \\ \Lambda\Sigma_{\mathbf{Y}\mathbf{Y}} & \Sigma_{\mathbf{w}\mathbf{w}} + \Lambda\Sigma_{\mathbf{Y}\mathbf{Y}}\Lambda^T \end{bmatrix} = \begin{bmatrix} \Lambda^T\Sigma_{\mathbf{w}\mathbf{w}}^{-1}\Lambda + \Sigma_{\mathbf{Y}\mathbf{Y}}^{-1} & -\Lambda^T\Sigma_{\mathbf{w}\mathbf{w}}^{-1} \\ -\Sigma_{\mathbf{w}\mathbf{w}}^{-1}\Lambda & \Sigma_{\mathbf{w}\mathbf{w}}^{-1} \end{bmatrix}^{-1}$$

For simplicity we assume the offset of the linear relationship \mathbf{b} to be zero. From [38] we can then write the Linear Gaussian in canonical form using Bayes' theorem

$$p(\mathbf{X}|\mathbf{Y}) = \frac{p(\mathbf{Y}, \mathbf{X})}{p(\mathbf{Y})},$$

using division of Gaussian densities defined in Section 6.5.2 where

$$K_{\mathbf{X}|\mathbf{Y}} = K_{\mathbf{Y},\mathbf{X}} - K_{\mathbf{Y},\mathbf{Y}} = K_{\mathbf{X},\mathbf{Y}} - \Sigma_{\mathbf{Y},\mathbf{Y}}^{-1},$$

and given that the $\mathbf{h}_{\mathbf{X}|\mathbf{Y}}$ vector is vacuous as we have no prior yet we have the following:

$$K_{\mathbf{X}|\mathbf{Y}} = \begin{bmatrix} \Lambda^T\Sigma_{\mathbf{w}\mathbf{w}}^{-1}\Lambda & -\Lambda^T\Sigma_{\mathbf{w}\mathbf{w}}^{-1} \\ -\Sigma_{\mathbf{w}\mathbf{w}}^{-1}\Lambda & \Sigma_{\mathbf{w}\mathbf{w}}^{-1} \end{bmatrix},$$

$$\mathbf{h}_{\mathbf{X}|\mathbf{Y}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$g_{\mathbf{X}|\mathbf{Y}} = -\log((2\pi)^{n/2}|\Sigma|^{1/2}).$$

Message Scheduling of Linear Gaussians The canonical form of a linear Gaussian such as above is a potential function and not a Gaussian probability density. One cannot normalise or marginalise the canonical linear Gaussian without first multiplying it with a Gaussian function. This means that factors which are linear Gaussian functions cannot send messages until they have received messages which allow them to be marginalised [14]. Initialising all messages as vacuous will therefore not allow the factors to send messages such as with the discrete factors. These factors will therefore have to wait for incoming messages to be able to send messages.

In [14] Koller and Friedman suggest using an order-constrained message-passing algorithm where the messages are carefully scheduled to allow the factors to become normalisable and able to send messages. We discuss our approach in Section 8.3.1, where we allow unnormalisable factors only to send vacuous messages.

6.6 Hybrid Conditional Linear Gaussian

The conditional linear Gaussian (CLG) is a hybrid factor that combines continuous and discrete factors to represent both types of random variables in a single factor. It is represented using a table with each entry of the table being a Gaussian. Multiplication and division are done similarly to the discrete tables, but instead of multiplying and dividing probabilities, one multiplies and divides Gaussian factors instead. Marginalising out continuous variables is done as with the Gaussian factors where the random variables are integrated out for each Gaussian factor in the table. The challenge is when one wants to marginalise out a discrete variables, as summing the Gaussian factors together gives one a mixture Gaussian function consisting of weighted Gaussian functions summed together.

Instead of representing the mixture Gaussian factors, an approximation can be made by collapsing the mixture Gaussian factors with some projection into a single Gaussian that minimises the error in some manner. M-Projection [39, 14] does so by minimising the Kullback-Leibler distance [25], $d_{\text{KL}}(p||\hat{p})$, between the mixture Gaussian, p , and the Gaussian \hat{p} that approximates it. This form of marginalisation is known as *weak marginalisation*. From [14], the new collapsed mean and covariance matrix, given the weight of each Gaussian, w_i , where $\sum_{i=1}^k w_i = 1$ is

$$\begin{aligned}\boldsymbol{\mu} &= \sum_{i=1}^k w_i \boldsymbol{\mu}_i, \\ \boldsymbol{\Sigma} &= \sum_{i=1}^k w_i \boldsymbol{\Sigma}_i + \sum_{i=1}^k w_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T.\end{aligned}\tag{6.6.1}$$

Taking into account that the integral of the new Gaussian factor, m_{mass} , is the sum of each integral of the Gaussian factors in the mixture, one can recalculate g , similarly to Equation 6.5.1, as:

$$g = -\frac{1}{2}\boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} - \log((2\pi)^{n/2} |\Sigma|^{1/2}) + \log(m_{\text{mass}}).$$

6.6.1 Approximate Messages in BP and BU

Performing weak marginalisation will cause the messages sent using either BP or BU to be approximated messages. The effects on the messages are different for BP and BU. In BU the incoming message, which is cancelled out, is used to determine the approximation as it is multiplied in before the weak marginalisation is performed. We discuss the impact this has on our graphs in more detail in Section 8.3.2.

Approximate Message-passing Convergence For approximate message passing we will need to have an alternative definition for convergence than from Section 6.3.9. With the approximated messages the clusters will not necessarily share the same marginal beliefs. We can define convergence as the state in which the messages being passed differ from their previous iteration with a sufficiently small margin [14].

6.7 Comparison of Discretised and Hybrid Computational Tractability

An alternative to describing continuous random variables by using parameterised functions is to discretise them. As a rule of thumb, discretised factors require greater computational processing than continuous or hybrid factors [14]. As discussed in Section 6.4.1, a discrete factor contains \mathfrak{K}^n probabilities for n variables, and \mathfrak{K} possible values. The exact number of numerical operations for the factor operations (multiplication, marginalisation, etc.) are based on the sparsity of the discrete factors. A Gaussian factor, on the other hand, is represented using a K matrix with n^2 entries, a vector of length n and a single g value. A CLG has a table of Gaussian factors and thus contains \mathfrak{K}^m Gaussian factors for m discrete variables that have \mathfrak{K} possible values. Our CLG representations tend to be much more compact than the exponentially sized discretised factors since they only have two discrete random variables per CLG cluster. These random variables also have less possible values than the discretised continuous random variables. Based on the above considerations, and despite the multiplications for CLG factors being more complex, the discrete factor operations tend to require many more numerical operations compared to those of the hybrid factor.

6.8 Conclusion

In this chapter we introduced the theory behind PGMs where we specifically looked at clique trees, undirected graphs with no loops, and cluster graphs, the loopy generalisation of clique trees. We showed how inference can be performed on these graphs using belief propagation or belief update and, more generally, loopy belief propagation or loopy belief update. Discrete random variables can be represented by a table of probabilities, and continuous random variables by parameterised functions such as Gaussian functions. We showed how one can have a graphical model with a mixture of discrete and continuous random variables, namely hybrid PGMs. The theory discussed in this chapter is used to implement the PGMs modelled in the next chapter.

Chapter 7

PGM Model Design

7.1 Introduction

In Chapter 4 we applied BSA to our Doppler radar data in both large signal segments and shorter time-steps. In this chapter we discuss the model we developed to smooth the shorter, spectrogram-like analyses while incorporating temporal information about the transitions of the longer segments. The concepts of state-space estimation from Chapter 5 will be combined with the PGM theory covered in Chapter 6.

In Section 5.7 we contextualised our model with respect to other multitarget approaches. Our targets of interest are the frequency components identified using BSA which we wish to smooth. In this chapter we discuss the modelling thereof in more detail. Using PGMs we model multitarget state-space estimation and add a discrete Markov chain, describing the signal segment transitions parallel to the traditional multitarget tracking. This additional Markov chain enforces a left-to-right progression of signal segments that determine the underlying state-space estimation dynamics and the active BSA model for each time-step. This differs from the traditional multitarget state-space estimation approach of modelling target birth and death rates within the target transitions. The extra temporal information allows the models to smooth over regions where the measurements are noisy or where a frequency component is missed. The transitions only allow the PGM to switch between signal segments in a specific strict order. To better enable the transition over areas where the targets do not appear in BSA, we implement a “missed target” model. We use both a fully discrete and a hybrid PGM that follow the same overarching principles but differ in implementation and how we construct the factors.

7.2 General Model Overview

This section contains the general approach to modelling the PGMs. The exact design of the factors differs between the discrete and a hybrid PGMs which we discuss in Section 7.3 and 7.4.

7.2.1 Model Assumptions

We make the same assumptions as in Section 5.2.1 which are summarised below:

1. The maximum number of targets is known.
2. The state-space transitions are a Markov chain.
3. Observations are made at discrete time-steps and are independent of the measurements of other time-steps.
4. The discrete state-space describing the number of targets include a “target not present” value, ϕ .

Assumption 3 implies uncorrelated noise between time-steps, which is not true for our spectrogram-like BSA approach. The unknown and unmodelled effects in the signal are considered to be noise in our BSA models and are not necessarily uncorrelated between time-steps. These assumptions simplify the PGMs by including conditional independences between the time-steps.

7.2.2 Graph Modelling Approach

Our PGM approach consists of three levels:

- lower-level competing BSA likelihood functions of the signal
- mid-level multitarget state-space estimation of those parameters with additional parameter dynamics
- high-level discrete signal segment transitions

Figure 7.1 is a Bayes network representation of our approach with green, blue, and red directed edges representing the lower-, mid-, and high-level transitions respectively. Once again we are only using the Bayes network to exercise our intuition about the approach.

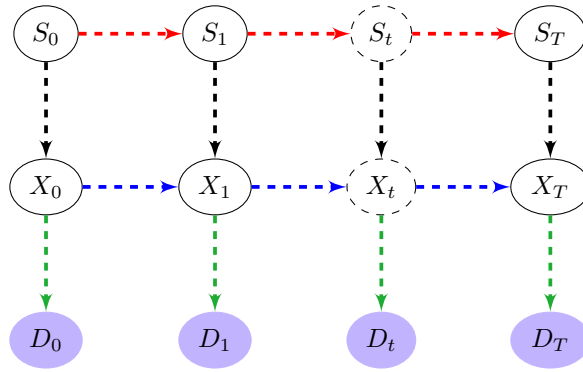


Figure 7.1: A Bayes network representation of the general model approach. Green: lower-level (likelihoods), Blue: mid-level (multitarget parameter dynamics), Red: high-level (Segment transitions).

The lower- and mid-level state-space transitions are similar to the approach discussed in Bayesian multitarget state-space estimation in Section 5.3. The Doppler radar data “observed” during each time-step are represented by the bottom nodes, $\mathbf{D}_T = \{D_0, \dots, D_T\}$. The data is broken into parts and used to calculate the BSA probability density functions for each time-step, as discussed in Section 4.7. The mixture of discrete and continuous random variables, $\mathbf{X}_T = \{X_0, \dots, X_T\}$, determine the signal frequency parameters and their dynamics, Ω_t as well as the active BSA model, M_t , for each time-step t so that

$$\begin{aligned} X_t &= \{\Omega_t, M_t\}, \\ \Omega_t &= \{\omega_t, \Delta\omega_t, \mathbf{a}_t, \Delta\mathbf{a}_t\}. \end{aligned} \quad (7.2.1)$$

These parameters, which are described in Table 7.1, will allow us to refine the likelihood functions that describe our stationary frequency parameters, $\omega_t = \{\omega_{t,0}, \omega_{t,1}, \omega_{t,2}\}$, given the active BSA model, $M_t \in \{1, 2, 3\}$.

Table 7.1: The state-space estimation parameters of interest describing the signal’s frequency dynamics

ω	Stationary frequency parameters
$\Delta\omega$	Difference in frequency parameters between time-steps
\mathbf{a}	Difference in $\Delta\omega$ between time-steps
$\Delta\mathbf{a}$	Difference in \mathbf{a} between time-steps

We add a discrete Markov chain, $\mathbf{S}_T = \{S_0, \dots, S_T\}$, parallel to the traditional switching state-space estimation as the high-level transitions for the Doppler radar signal segments from Section 4.6. The Markov chain is connected to each time-step $S_t \rightarrow X_t$ which is necessary to propagate information

about the segments to the signal dynamics.

In our PGMs, the three levels of the model correspond loosely to the three different sources of knowledge we used to construct the model in Figure 7.1:

- The different BSA estimations of the parameters at each time-step, including the frequency estimates and the probability of each BSA model which is described by their relative BSA lower-level likelihood functions (indicated in green).
- The fact that each time-step's parameters are dependent on the previous time-step in a way that is limited by the behaviour of the physical targets' motion, which is described by the mid-level parameter dynamics (indicated in blue).
- The fact that each discrete signal segment occurs in a certain order, and other temporal information about signal segments which is described by the high-level segment transitions (indicated in red).

7.2.3 Prior Probability Density Functions

The three levels of knowledge we incorporate are built into the PGM as prior probability density functions in the following factors. Generally we do not have much prior information about the random variables other than the relationships between them and the possible values that they can assume. This means that for the discrete random variables we assume a uniform prior over the region of interest with the CPD correctly normalised.

7.2.3.1 High-level Signal Segment Estimation

We model the transition between the different long signal segments from Section 4.6 (downward club swing, ball only, etc.) with a discrete Markov chain. The transitions describe a strict left-to-right progression of signal segments as described in Section 5.5.2. This forces the model to decide on the ideal location to transition between signal segments as they are limited in number and can only occur in a certain order. The state transition factors are described by the following CPDs:

$$p(S_t|S_{t-1}), \quad 0 < t \leq T,$$

$$p(S_0), \quad t = 0.$$

An example of a segment transition containing three segments can be seen in Table 7.2.

Table 7.2: A discrete factor describing an example left-to-right segment transition of three segments.

S_t	S_{t-1}	$p(S_t S_{t-1})$
segment 1	Begin	1.0
segment 1	segment 1	0.5
segment 2	segment 1	0.5
segment 2	segment 2	0.5
segment 3	segment 2	0.5
segment 3	segment 3	0.5
End	segment 3	0.5
Elsewhere		0.0

Between the high- and mid-level factors there are factors which translate the segment random variables, \mathbf{S}_T , to the active BSA models:

$$p(M_t|S_t), \quad 0 \leq t \leq T.$$

Each BSA model may appear in multiple signal segments. An example of such a factor can be seen in Table 7.3 where the model, $M = 1$, occurs in both signal segment $S = 1$ and $S = 3$.

Table 7.3: A discrete factor describing an example relationship between signal segments and the active BSA model

M_t	S_t	$p(M_t S_t)$
model 1	segment 1	0.5
model 2	segment 2	1.0
model 1	segment 3	0.5
Elsewhere		0.0

7.2.3.2 Mid-level State-space Estimation

The mid-level state-space estimation describes the dynamics of our frequency parameters and is similar to the multitarget state-space estimation from Chapter 5. The parameter dynamics are described by the following CPDs:

$$p(\Omega_t, M_t|\Omega_{t-1}, M_{t-1}), \quad 0 < t \leq T,$$

$$p(\Omega_0|M_0)p(M_0), \quad t = 0.$$

The BSA models differ in the number of targets present and thus have a different number of parameters to estimate, and different signal dynamics. The dynamics are linear relationships such as used in Kalman filters. However, these dynamics will be conditioned on discrete random variables. For each

combination of the three BSA models we would have a linear relationship between the frequency parameters:

$$\Omega_t = \Lambda\Omega_{t-1} + \mathbf{w}_t$$

As an example, the linear relationship between a single-frequency component to a single-frequency component transition is as follows:

$$\begin{bmatrix} \omega_{t+1} \\ \Delta\omega_{t+1} \\ a_{t+1} \\ \Delta a_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_t \\ \Delta\omega_t \\ a_t \\ \Delta a_t \end{bmatrix} + \mathbf{w}.$$

The full list of frequency transitions and their linear relationships can be found in Section F.1. In Section 7.2.5 we discuss in more detail how we handle transitions between different numbers of frequency components.

7.2.3.3 Likelihood Functions

The lower-level likelihood functions are calculated similarly to the analysis of short-step parameter estimation from Section 4.7.2. We have the same approach of breaking the signal up into overlapping windows similar to a spectrogram with a window size of 64 samples and a step-size of 16 samples. We employ the nine BSA models used throughout Chapter 4 to calculate the competing hypotheses for a given window. For the sake of simplicity, and because the BSA models from Chapter 4 do not fully explain the data complexity, we only calculate the probability density functions for the stationary frequencies. This gives us more conservative estimates of the frequency parameters.

The likelihood functions have the following form:

$$p(D_t|\boldsymbol{\omega}_t, M_t), \quad 0 \leq t \leq T,$$

where $\boldsymbol{\omega}$ are the stationary frequency parameters and M indicates the active BSA model. In this case we have likelihood functions for only a subset of the state-space of interest so that $\boldsymbol{\omega} \subset \Omega$.

7.2.4 Factor and State-space Symmetry

Our state-space estimation models do not need to contain target identity, as we have the joint probability density function of the frequency components for each time-step. As discussed in Chapter 3, the BSA joint probability densities are symmetrical, as each of the frequency components present in the Doppler radar data can be associated with any of the frequency parameters. For a two-dimensional probability density function with a clearly defined maximum,

there are two identical peaks and symmetry around the $\omega_0 = \omega_1$ axis. A three-dimensional probability density function would have 6 (3!) identical peaks.

The discrete factors can easily encode the symmetry, but doing so increases the size of the factors by the number of symmetrical peaks. For the sake of computational tractability we remove the symmetry in the discrete factors and renormalise them to contain their original model probability ratios. The Gaussian factors cannot describe the symmetry as they contain a single peak. Therefore for both factor types the lack of symmetry creates implicit target identity, which must somehow be maintained. Each frequency parameter, $\{\omega_{t,0}, \omega_{t,1}, \omega_{t,2}\}$, is now associated with a specific frequency component in the Doppler radar signal. This subtly differs from the data association in Section 5.6.1, which requires one to associate measurements with targets. A data association problem might have multiple measurements that are associated with multiple targets. Our single “measurement” (BSA factor) is associated with up to an arbitrary number of targets (three in our case), which always gives us all of the hypotheses. The targets of interest do not cross paths in our Doppler radar signals since the frequency components tend to move away from each other, and so we can arbitrarily assign the frequency components’ identities from the highest to the lowest frequency.

7.2.5 Implicit Target Identity When Adding/Removing Targets

As discussed in the previous section, we have implicit target identity due to the removal (discrete model) or lack (hybrid model) of symmetry in the factors. Although the model does not explicitly designate a target to a frequency component, this can be handled implicitly by the frequency component transitions. For example, we could have two frequency components at time-step t , $\{\omega_{t,0}, \omega_{t,1}\}$, and the top frequency component dies away. The transition factor would then model that the component of the one frequency model at time-step $t + 1$ would be the continuation of the lower frequency component from the previous time-step, $\omega_{t,1} \rightarrow \omega_{t+1,0}$. The linear transition of the above example then is the following:

$$\begin{bmatrix} \omega_{t+1,0} \\ \Delta\omega_{t+1,0} \\ a_{t+1,0} \\ \Delta a_{t+1,0} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{t,0} \\ \Delta\omega_{t,0} \\ a_{t,0} \\ \Delta a_{t,0} \\ \omega_{t,1} \\ \Delta\omega_{t,1} \\ a_{t,1} \\ \Delta a_{t,1} \end{bmatrix} + \mathbf{w}.$$

We handle the addition of a new frequency component by multiplying a transition factor with an uninformative prior for the new component. For the discrete models this is a uniform prior and in the case of the hybrid model this is a Gaussian with a large variance. One could add more prior information at this point where, for example, one could have a prior based on the starting frequency of the ball component. We do not do so, however, as the BSA peaks are so sharp that any prior that does not have a comparable variance has little effect.

7.2.6 Posterior

We could calculate the posterior probability of our model by multiplying all the factors together:

$$\begin{aligned} p(\mathbf{D}_T, \mathbf{X}_T, \mathbf{S}_T) &= p(\Omega_0|M_0)p(M_0|S_0)p(S_0) \\ &\times \prod_0^T p(D_t|\boldsymbol{\omega}_t, M_t)p(M_t|S_t) \\ &\times \prod_1^T p(\Omega_t, M_t|\Omega_{t-1}, M_{t-1})p(S_t|S_{t-1}). \end{aligned}$$

Calculating the joint probability density function $p(\mathbf{D}_T, \mathbf{X}_T, \mathbf{S}_T)$ and then marginalising to the random variables of interest is computationally intractable for a large number of total time-steps $T + 1$. However, as discussed in Chapter 6, we can use PGMs to calculate the marginal probability densities. The approximate marginal posterior probability density functions are calculated using LBU [14] with RBP scheduling [35] without having to multiply all the factors together. We query the relevant clusters to find the posterior probability densities of the active signal segment and the stationary frequency parameters. This is equivalent to calculating

$$p(X_k|\mathbf{D}_T) = \frac{\int dX_0 \dots \cancel{dX_k} \dots dX_T d\mathbf{S}_T p(\mathbf{D}_T, \mathbf{X}_T, \mathbf{S}_T)}{p(\mathbf{D}_T)} \quad (7.2.2)$$

for our parameters at each time-step t , where $X_k = \{\Omega_k, M_k\}$.

7.3 Discrete Model

Discretising the continuous parameters in our PGM may cause the inference to be computationally intractable, as the number of probabilities represented in the factors become prohibitively large. The Bayes network in Figure 7.1 is sparsely connected between the signal segments and the parameter dynamics which means that there are seemingly not many random variable interactions to describe in our PGM. However, some of the complexity is hidden in the X_t

node transitions that describe the parameter dynamics. The discrete factor sizes of the parameter dynamics are large even for a low resolution of sampling. To be able to perform inference within a reasonable time-frame while also maintaining an adequate discrete resolution, we will need to make full use of conditional independence between the random variables.

7.3.1 Likelihood Functions: Discretisation of BSA

We discretise the continuous random variables by:

- calculating the probability density functions at regular intervals.
- multiplying them by their volume element to make them probabilities.
- storing them in a table.

Multiplying each sample by a volume element approximates the integral of probability density function in that volume.

7.3.2 Discrete Prior Factors

We can reduce the connections between random variables by encoding more conditional independences. The frequency component values $\omega = \{\phi, 0.1\}$ can only occur in the single frequency model, implying that $M_t = 1$. We can thus write the likelihood function factors as $p(\omega)$. This is also true for the dynamics of the mid-level transitions, since we need to make use of place-holder values for those random variables as well.

We can reduce the factor size of our parameter dynamics by modelling conditional independence between the parameter from Table 7.1. As an example, the product of a time-step's parameter dynamics with a prior density on the previous time-step dynamics and the current BSA model:

$$p(\omega_{t+1}, \Delta\omega_{t+1}, \mathbf{a}_{t+1}, \Delta\mathbf{a}_{t+1} | \omega_t, \Delta\omega_t, \mathbf{a}_t, \Delta\mathbf{a}_t, M_{t+1}) \\ \times p(\omega_t, \Delta\omega_t, \mathbf{a}_t, \Delta\mathbf{a}_t, M_{t+1}),$$

can be separated into smaller CPDs since many of the random variables are conditionally independent of each other:

$$p(\omega_{t+1} | \omega_t, \Delta\omega_{t+1}) p(\Delta\omega_{t+1} | \Delta\omega_t, \mathbf{a}_{t+1}) p(\mathbf{a}_{t+1} | \mathbf{a}_t, \Delta\mathbf{a}_{t+1}) \\ \times p(\Delta\mathbf{a}_{t+1} | \Delta\mathbf{a}_t, M_t) p(\Delta\mathbf{a}_t | M_{t+1}) p(M_{t+1}).$$

The resulting PGM after including conditional independences can be seen in Figure 7.2. The graph now contains many more clusters, but the size of the factors needed to factorise the joint probability density has been greatly reduced. This also serves as a good example of how decomposing the problem

into smaller factors can allow computational tractability for PGM inference.

The segment transition factors are similar to the general approach as described in Section 7.2.3.1, however, the linear relationships for the mid-level dynamics are now split over multiple factors. We also do not model the process noise as with the linear Gaussians from Section 6.5.3 and instead use a uniform prior over the range of possible values. The discrete factors have linear relationships over a smaller subset of random variables, for example, for the CPD of $p(\omega_{t+1}|\omega_t, \Delta\omega_{t+1})$, the relationship can be described as

$$[\omega_{t+1}] = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_t \\ \Delta\omega_{t+1} \end{bmatrix}.$$

Similar factors are constructed for $p(\Delta\omega_{t+1}|\Delta\omega_t, \mathbf{a}_{t+1})$ and $p(\mathbf{a}_{t+1}|\mathbf{a}_t, \Delta\mathbf{a}_{t+1})$.

We need to use place-holder values in the discrete factors to cater for the transitions between the different dynamics of the different BSA models which are all contained within the same discrete table. An example of how we calculate the inner workings of the linear relationship described above for $p(\omega_{t+1}|\omega_t, \Delta\omega_{t+1})$ can be seen in Table 7.4. More examples of how we calculate the discrete factors can be found in Appendix F.2. The other transition factors have the same premise; we exhaustively list all possible values while making use of place-holder values and then normalise the factor as a CPD. The use of the place-holder values also ensure that the transitions are unique for each model transition.

Table 7.4: *This table describes how we calculate the possible values for the discrete factors of the form $p(\omega_{t+1}|\omega_t, \Delta\omega_{t+1})$. The random variables on the right of the conditioning bar (denoted with the double line), determine the possible values of the random variables on the left of the conditioning bar for the given model transition. The place-holder values in the factor are ϕ , \mathfrak{A} , \mathfrak{B} , \mathfrak{C} , and \mathfrak{D} . The ϕ value denotes that there is no target, \mathfrak{A} and \mathfrak{C} indicate a transition of the top target, and \mathfrak{B} and \mathfrak{D} of the bottom target.*

Model Transition	$\omega_{t+1,0}$	$\omega_{t+1,1}$	$\omega_{t,0}$	$\omega_{t,1}$	$\Delta\omega_{t+1,0}$	$\Delta\omega_{t+1,1}$
1 freq to 1	ϕ	$\omega_{t,1} + \Delta\omega_{t,1}$	ϕ	$\omega_{t,1}$	ϕ	$\Delta\omega_{t,1}$
2 freq to 2	$\omega_{t,0} + \Delta\omega_{t,0}$	$\omega_{t,1} + \Delta\omega_{t,1}$	$\omega_{t,0}$	$\omega_{t,1}$	$\Delta\omega_{t,0}$	$\Delta\omega_{t,1}$
1 freq to 2 \mathfrak{A}	$\omega_{t+1,0}$	$\omega_{t,1} + \Delta\omega_{t,1}$	ϕ	$\omega_{t,1}$	\mathfrak{A}	$\Delta\omega_{t,1}$
1 freq to 2 \mathfrak{B}	$\omega_{t,1} + \Delta\omega_{t,1}$	$\omega_{t+1,1}$	ϕ	$\omega_{t,1}$	$\Delta\omega_{t,0}$	\mathfrak{B}
2 freq to 1 \mathfrak{C}	ϕ	$\omega_{t,1} + \Delta\omega_{t,1}$	$\omega_{t,0}$	$\omega_{t,1}$	\mathfrak{C}	$\Delta\omega_{t,1}$
2 freq to 1 \mathfrak{D}	ϕ	$\omega_{t,0} + \Delta\omega_{t,0}$	$\omega_{t,0}$	$\omega_{t,1}$	$\Delta\omega_{t,0}$	\mathfrak{D}

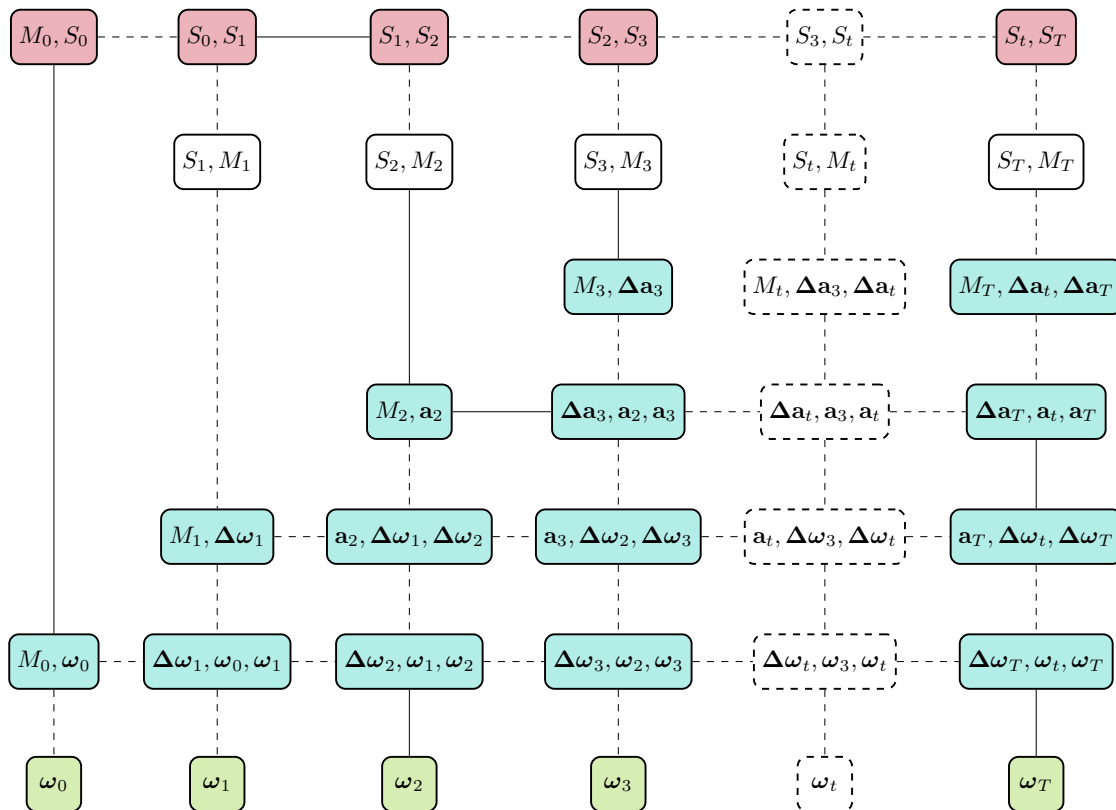


Figure 7.2: *Discrete Doppler radar PGM. The red, blue, and green clusters indicate the high-level segment transitions, mid-level parameter dynamics, and lower-level likelihood functions respectively.*

7.4 Hybrid Model

The hybrid discrete-continuous model is constructed using CLGs where the continuous random variables are approximated with Gaussian densities. Since the parametric representation of the frequency dynamic is more compact (see Section 6.7), we do not need to make as much use of conditional independence.

7.4.1 Likelihood Functions: Continuous BSA Factors

We find the peak of the BSA using the Nelder-Mead algorithm as discussed in Section 4.3. We approximate the covariance matrix for each BSA model using Laplace's method [25] (see Appendix B.2) and their relative probabilities are encoded in the g term of the Gaussian factors.

7.4.2 Hybrid Prior Factors

Using the linear relationships in Section 7.2.3.2, we can describe the factors using linear Gaussians for each of the possible BSA model transitions. The high-level transitions from Section 7.2.3.1 are discrete factors which can pass messages to CLG factors by summing the log probability of the discrete factor with the g value in the relevant Gaussian factor. In a similar way the CLG factors can pass messages to the discrete factors by marginalising out all the continuous random variables and calculating the probability from the g value for each Gaussian factor. The resulting hybrid cluster graph can be seen in Figure 7.3. Note that there are much less clusters than the discrete cluster graph, since we can more compactly represent the CPDs.

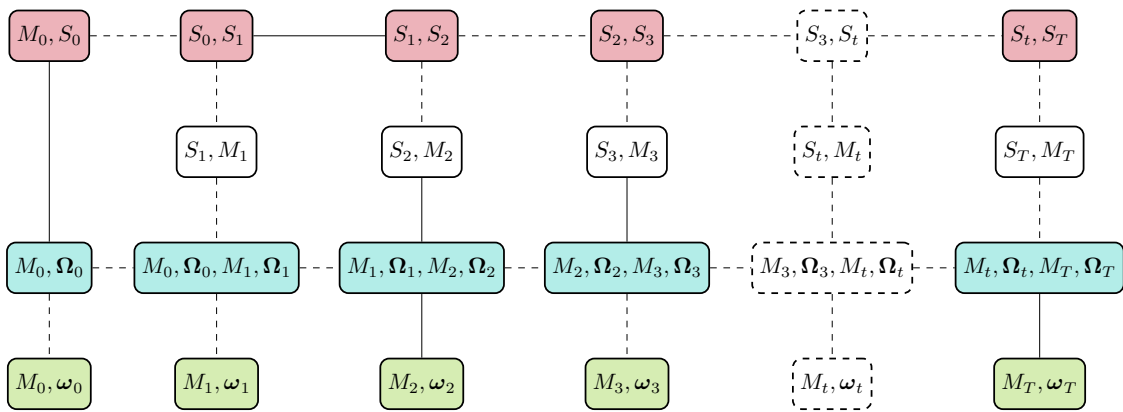


Figure 7.3: Hybrid Doppler radar PGM. The red, blue, and green clusters indicate the high-level segment transitions, mid-level parameter dynamics, and lower-level likelihood functions respectively.

7.4.3 False Alarms and the Missed-target Hypothesis

In state-space estimation methods, the target birth and death rates are coupled with the concepts of false alarms and missed targets. False alarms are when a target is detected but is not actually present, and missed targets are when a target is present but not detected. Within our own models, the fact that we are using competing hypotheses with different numbers of targets serves the function of a false-alarm model. The PGMs can always default back to a BSA hypothesis with fewer targets and handle a false alarm if the surrounding observations and signal segment transitions do not support the presence of a target.

While we found that our models could perform without the concept of a missed-target model, the hybrid model greatly benefits from implementing it. We take the results of a BSA and multiply it with a Gaussian prior probability

density function in order to create the missed-target model, which competes with the other BSA models. The variance of the undetected target is very large in order to allow the other time-steps' BSA factors to influence the system's knowledge of the undetected target. This will allow the system to extrapolate a path for the target given the signal dynamics without having any input for that target from BSA. For example, we can take a single frequency BSA and multiply it with a univariate Gaussian prior to create a two-frequency hypothesis where one of the targets is a missed target.

7.5 Conclusion

In this chapter we discussed our approach to modelling PGMs which we apply to Doppler radar data in Chapter 9. The concepts from Chapter 3, 5, and 6 come together in order to create a multitarget tracking PGM that includes temporal information about signal segment transitions. Our overall approach was conceptually similar for the two graph types but differed in how we constructed the discrete and hybrid PGMs' factors. The discrete factors required more conditional independence to be built into the graph. We discussed how concepts such as process noise, false-alarms, and missed targets applied to our application within the context of PGMs.

Chapter 8

Implementation Details and Challenges

8.1 Introduction

In this chapter we discuss the implementation details, constraints, and associated challenges of the discrete and hybrid models. We need to consider the fact that discretising our BSA factors has effects on inter-BSA model probabilities. There are subtleties in the CLG factors such as constraints on message-passing scheduling, message approximations, and choosing the linear Gaussian process noise. To perform inference on our models, we use LBU with RBP. We choose LBU, as the software is further developed and potentially performs better for the hybrid model which we discuss in Section 8.3.2.

8.2 Discretisation

One of the challenges we face with the discrete factors is the computational intractability of the PGMs when the resolution of the discretisation is high. As discussed in Section 6.4.1, the number of probabilities that need to be represented in a discrete factor increase exponentially with the increase in random variables. This consequently increases the number of mathematical operations that need to be performed in order to calculate factor multiplication, cancellation, normalisation, and marginalisation. However, reducing the resolution degrades the accuracy of the discrete approximation as it can fail to capture the shape of the original probability density. It also affects the relative probabilities between BSA models ($p(M)$). The sharp peak of the most likely point in the likelihood functions may be missed when calculating the probability density function values at regular intervals as discussed in Section 7.3.

We mitigate some of these effects by binning a finer-resolution discretisation into a lower resolution. Our approach to creating the discrete BSA factors is

the following:

1. Calculate the probability density function values at regular intervals at a higher resolution (256 intervals per dimension).
2. Bin and average the values to a lower resolution (32 intervals per dimension).
3. Multiply the averaged probability density with a volume element to create a probability for the discrete factor.

While the averages more closely resemble the probability in the region being approximated, a very sharp peak may still be missed using the approach above and thus skew the relative BSA model likelihoods. In order to mitigate this, we calculated the relative BSA model probabilities within a factor by using the model selection from Section 3.7. We adjust the factors so that the sum of the relative probabilities between the models is equal to the relative probabilities of the models:

$$\frac{\sum \Phi(M = 1)}{\sum \Phi(M = 2)} = \frac{p(M = 1|D)}{p(M = 2|D)},$$

where $\sum \Phi(M = 1)$ is the element-wise sum of the values in the factor Φ relating to the single stationary frequency model. This ensures that there is a more accurate portrayal of the marginal probability of the models so that:

$$\sum_{\omega} \Phi \approx \int d\omega p(D|\omega, M).$$

8.3 CLG Factors Implementation Subtleties

8.3.1 Message Passing Constrained by Matrix Inversion and Unnormalisable Factors

While it is simpler to marginalise a Gaussian density in covariance form such as discussed in Section 6.5.2, one would then need to convert the Gaussian factors between covariance and canonical form when passing messages. The limits on computational precision when calculating the matrix inversions in the CLG factor operations can negatively influence the accuracy of the results [14]. This is especially a problem if the process noise for the linear Gaussians from Section 6.5.3 is very small. We avoid inverting the matrix K by keeping the factors in the canonical form when marginalising, which only inverts a submatrix of K such as from Equation 6.5.2:

$$K' = K_{\mathbf{X}\mathbf{X}} - K_{\mathbf{X}\mathbf{Y}}K_{\mathbf{Y}\mathbf{Y}}^{-1}K_{\mathbf{Y}\mathbf{X}},$$

where K' is the new K matrix for the marginalised Gaussian and from Equation 6.5.3

$$K = \begin{bmatrix} K_{\mathbf{X}\mathbf{X}} & K_{\mathbf{X}\mathbf{Y}} \\ K_{\mathbf{Y}\mathbf{X}} & K_{\mathbf{Y}\mathbf{Y}} \end{bmatrix}.$$

However, this would still require that $K_{\mathbf{Y}\mathbf{Y}}$ be invertible for each Gaussian within a CLG factor in order to marginalise a continuous random variable.

When a model consists of only Gaussian factors, there is no need to invert K when one follows the approach above. However, in order to marginalise discrete random variables in the CLG factors, there are additional restrictions on the K matrices. If K is not invertible for a particular Gaussian, we cannot calculate the probability mass of that Gaussian. This also means that the Gaussian cannot be normalised, hence we call such Gaussians “unnormalisable”. CLG factors containing unnormalisable Gaussians cannot marginalise discrete random variables, as they require weak marginalisation from Equation 6.6.1:

$$K^{-1} = \Sigma = \sum_{i=1}^k w_i K_i^{-1} + \sum_{i=1}^k w_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T.$$

This can prevent such clusters from sending messages in our hybrid model. Figure 8.1 (which is a subset of the hybrid model from Figure 7.3) indicates the messages passed to a mid-level transition cluster. The subscript in each message indicates the scope of the random variables in the message. The cluster \mathcal{C}_t will only be able to send a message to \mathcal{C}_{t+1} or \mathcal{C}_M if it has received a non-vacuous message from \mathcal{C}_{t-1} . Canonical-form linear Gaussians such as those contained in \mathcal{C}_t are not normalisable unless multiplied with a factor containing the continuous random variables they are conditioned on [14]. The mid-level clusters will therefore have to wait for specific incoming messages to be able to send their messages. The cluster \mathcal{C}_t is thus *order constrained* [14] by \mathcal{C}_{t-1} since it can only send a message once it has received a message from \mathcal{C}_{t-1} .

Scheduling the CLG message passing requires some method of identifying which clusters are order-constrained by which other clusters [14]. For the sake of simplicity, we instead use the following approach when attempting to send a message from a CLG cluster:

1. Attempt to marginalise a CLG factor in a cluster for an outgoing message.
2. If it succeeds, create the message.
3. If it fails due to an unnormalisable Gaussian factor, create a vacuous message instead.

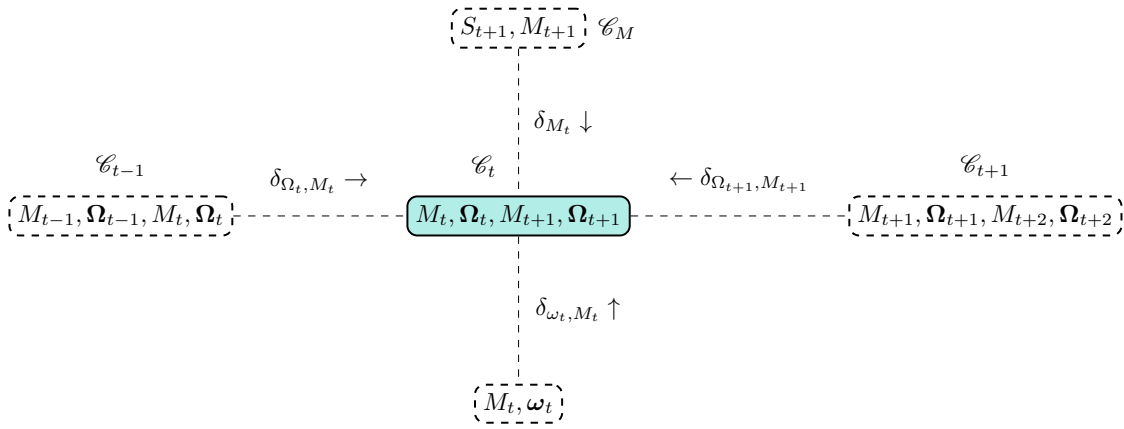


Figure 8.1: Subset of Figure 7.3, the hybrid Doppler radar PGM. It indicates the messages passed to a mid-level transition cluster containing a CLG factor.

4. RBP accordingly schedules the messages according to the message residuals.

In this way we can employ the same LBU algorithm used for the discrete factors and avoid writing a custom message-passing algorithm for the CLG graphs. RBP will automatically move the perpetually vacuous messages to the bottom of the message schedule until their clusters have received the necessary inputs from their neighbours. Once marginalisation is possible for a CLG factor, its message residuals will change accordingly and the cluster will also start sending non-vacuous messages. In Figure 8.1 this means that \mathcal{C}_t requires a non-vacuous message from \mathcal{C}_{t-1} .

8.3.2 LBU and LBP Approximated Messages

The weak marginalisation we perform means that we will be sending approximated messages when performing inference on the hybrid model. As discussed in Section 6.6, weak marginalisation is performing M-Projection, which minimises the Kullback-Leibler distance between the projection and the true density function. The approximation process makes a trade-off in that when certain regions of the probability space are better approximated, other regions will have a worse approximation [14]. However, there is a difference between LBU and LBP when approximating messages. For LBP, all messages but one are multiplied in when weak marginalisation takes place, whereas for LBU, *all* incoming messages are multiplied in first. For example, in Figure 8.1, when sending a message from \mathcal{C}_t to \mathcal{C}_{t+1} , the approximation will include $\delta_{\Omega_{t+1}, M_{t+1}}$ in LBU but not LBP. Thus LBU, in contrast with LBP, takes into account the current approximation $\delta_{\Omega_{t+1}, M_{t+1}}$, sent from \mathcal{C}_{t+1} , potentially focusing the approximation on “more relevant” parts of the probability space [14]. This is

similar to the principle used in *expectation propagation* [40] in which messages are approximated by using all incoming messages.

We do not use damping (from Section 6.3.11) for our hybrid as it also creates a mixture of Gaussian densities such as those found in weak marginalisation. We wish to avoid approximating the message as much as possible.

8.3.3 Process Noise

The process noise in conjunction with the measurement noise determine how at each time-step an “observation” (in our case the BSA factors) is weighed against the messages coming from the neighbouring time-steps. For example, in Figure 8.1 the variances of the Gaussians in the hybrid message δ_{ω_t} is determined by the measurement noise (in our case, the BSA confidence in the parameters selected). In turn, the variances of the Gaussians in the hybrid messages δ_{Ω_t, M_t} and $\delta_{\Omega_{t+1}, M_{t+1}}$ are determined by the process noise \mathbf{w} .

A very small measurement noise will require a small process noise to allow for observations to be smoothed by the model dynamics – otherwise the observed data will carry too much relative weight in determining the frequency parameters. On the other hand, if the model transitions do not match the true signal behaviour, a small process noise can be detrimental to the results and the graph will discount the information from the BSA factors. Ideally one would find a middle ground which reflects the accuracy of the processes in the parameter dynamics relative to the signal noise.

The covariance of the process noise can be calculated as the expected value of the signal dynamics [41]:

$$\mathcal{Q} = E[\Omega \Omega^T].$$

In our case we only smooth the signal and have no input into the signal processes. This makes it difficult to determine the process noise since we do not have a clear understanding of $E[\Omega \Omega^T]$. As we could not find an easily implemented way to automatically select the process noise, we do so by inspecting the results and adjusting the process noise. This resembles a method described in [41]; however, in our case we have the advantage of being able to use the spectrogram as a starting point:

1. Generate a spectrogram of the signal of interest.
2. Superimpose a linear model over it by inspection and look for the maximum deviation from our linear model.
3. Use the maximum deviation as two times the standard deviation of the process noise (which will contain $\approx 98\%$ of the values) as a rule of thumb.

4. Manually adjust and fine-tune the process noise to obtain better results.

8.4 Finding the BSA Peak

In order for us to be reasonably certain that our use of the Nelder-Mead optimisation method [26], discussed in Section 4.3, finds the peak in the BSA probability density function, we choose multiple good starting points for the optimisation technique. The starting points for the algorithm are the top four local maxima in the zero-padded (1024) periodogram. We use a very simple method, namely looking at the nearest neighbours in the periodogram to find the local maxima.

8.5 Non-convergence of PGMs

In both the discrete and the hybrid model, the PGMs occasionally do not converge within a reasonable time. This was more likely to happen in the particularly loopy discrete graphs. It would occur in both models if there were parts where the BSA did not reflect the parameter dynamics used to describe the frequency components. This was mitigated by the use of the missed target model from Section 7.4.3 for the hybrid PGMs.

8.6 Conclusion

We discussed the challenges we faced in implementing the models in Chapter 7. We experienced challenges such as selecting the process noise, the effects of discretisation, the message-passing scheduling, non-convergence of graphs, and challenges surrounding the computational tractability of our models. In the next chapter we will demonstrate the experimental results of the implementation of the models.

Chapter 9

PGM Experimental Results

9.1 Introduction

In this chapter we apply the PGMs modelled in Chapter 7 to the Doppler radar data. Our PGMs calculate the posterior probabilities of the random variables in each cluster. We query our graphs for the frequency parameters and the BSA model for each time-step, namely $p(\boldsymbol{\omega}_t, M_t)$. To be able to visualise the frequency parameters, we use an approach where we represent the different frequency components on one axis using different colours for each BSA model. The result is a weighted frequency representation much like a “probabilistic” spectrogram.

We demonstrate the use of the two PGM models from Chapter 7, the discrete model, and the hybrid model. Both models smooth the frequency parameters of the short time-step BSA from Section 4.7.2. We had moderate success with the Doppler radar data, and demonstrate some of the methods we used to mitigate the uncertainty surrounding the underlying signal behaviour.

Even though we do not have all the puzzle pieces to decisively solve the multitarget tracking problem, the Bayesian approach does have the advantage that it is easily extendible with more prior information about the signal. It improves upon the conventional multitarget tracking.

9.1.1 Representation of Results

We wish to estimate the posterior marginal probabilities of the frequency components and the active BSA model such as in Equation 7.2.2:

$$p(X_k | \mathbf{D}_T) = \frac{\int dX_0 \dots \cancel{dX_k} \dots dX_T d\mathbf{S}_T p(\mathbf{D}_T, \mathbf{X}_T, \mathbf{S}_T)}{p(\mathbf{D}_T)},$$

where we defined X_k in Equation 7.2.1 as the set containing our frequency dynamics parameters and the active BSA model:

$$X_k = \{\Omega_k, M_k\}.$$

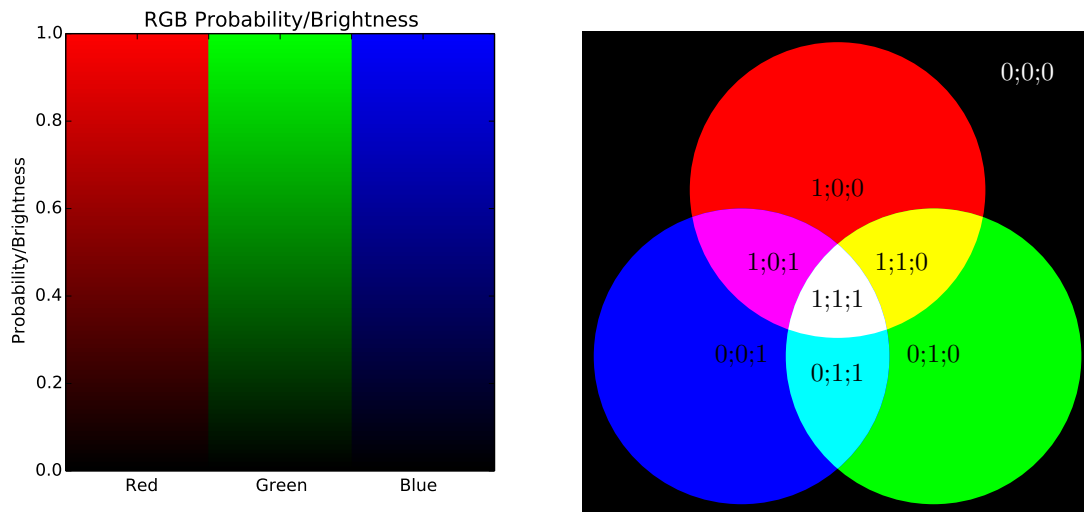
As discussed in Section 6.3.2, the converged PGMs can be queried on the random variables contained within a cluster. We are specifically interested in a subset of X_k , which represents the smoothed stationary-frequency parameters of our BSA factors, and the active BSA model, namely ω_k and M_k . However, it is difficult to visualise the resulting multidimensional probability density function. To achieve this goal, and since the frequency components and their relation to one another change over time, we use a spectrogram-like representation that combines the three BSA models. In order to still clearly discern the different hypotheses, we display the results as a red/green/blue (RGB) image. The intensity of each of the colours is proportional to the relative probability density of each of the three possible BSA models. The colour black would then indicate the absence of frequency components.

An example of how different combinations of RGB values create different colours can be seen in Figure 9.1. The intensity of each of the RGB values is shown in Figure 9.1a, and the combinations of full-intensity RGB values are shown in Figure 9.1b. We use the properties of red/green/blue images to display multiple BSA hypotheses on top of one another in the same image.

We do the following in order to represent the multidimensional $p(\omega_t, M_t)$ for each time-step t :

1. Query $p(\omega_{t,l}, M_t)$ for each frequency parameter l and each BSA model k .
2. For each of the models $M_t = k$, sum the l marginal frequency components together. This produces “composite” probability densities for each of the three BSA models.
3. Weigh each of the three composite probability densities by the marginal probability of that BSA model $p(M_t = k)$.
4. Assign a colour to each of the models: red for one frequency component, blue for two frequency components, and green for three frequency components.
5. Normalise each of the three weighed composite probability densities by dividing them by the maximum value between all three probability densities. This highlights the parts in which there are frequency components.

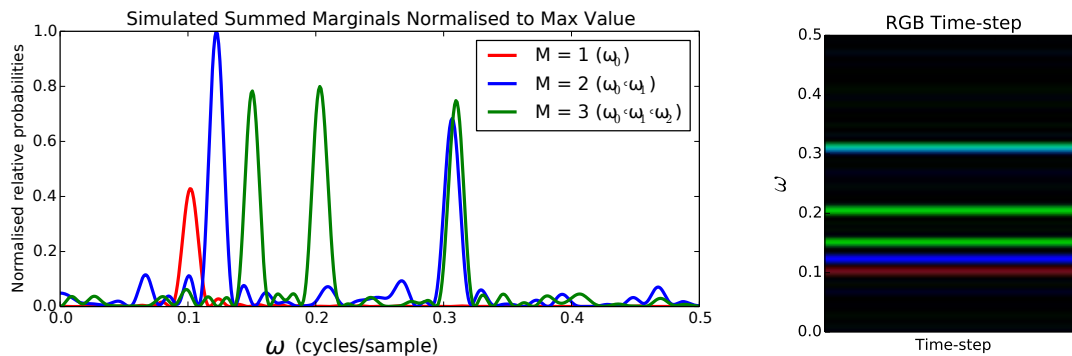
The above approach allows us to easily compare the relative probability densities of the different BSA models and how they change over time. An



(a) Example brightness of red, green, and blue respectively.

(b) RGB example colour combinations with full brightness.

Figure 9.1: Images can be represented using varying brightness in red green and blue. Figure (a) shows red, green, and blue of varying brightness corresponding to the relative probability density of a frequency component. Figure (b) displays different full intensity colour combinations of red, green, and blue with the intensity indicated in the format: $\{R;G;B\}$



(a) Simulated summed marginal probability density functions that are normalised to the maximum value.

(b) The RGB representation of Figure (a).

Figure 9.2: In Figure (a) we see simulated examples of summed marginal probability density functions for a given time-step. Each colour represents a BSA model. Using the values from a graph such as Figure (a), we can represent the graph as an RGB image for each time-step in order to visualise the graphs as they change over time. Figure (b) is an example RGB time-step of Figure (a).

example result of the steps above can be seen in Figure 9.2a. In order to add an extra dimension, time, we display the representation of each time-step as a column in an RGB image, much like a spectrogram does with the periodogram. The example time-step in Figure 9.2a would then appear as in Figure 9.2b. The disadvantage of this representation, however, is that when the probability density peaks are particularly sharp or spread out, they become difficult to see. In order to mitigate the effect, we enlarge the regions of interest when this occurs.

9.2 Discretise Model Results

We limit the discrete analyses to two dimensions, as the factors become very large if we add a third dimension. In the case of the discrete factors that do not contain M_t , the probability of the single sinusoid is still easy to calculate since ω_t uniquely determines the model. The probability of the placeholder ϕ , which indicates that a target is not present, is the probability of the single frequency component model:

$$p(\omega_0 = \phi) = p(M_t = 1).$$

9.2.1 Discrete: Segment C-B

In this example we demonstrate the effect that restricting the frequency components' dynamics can have on the segment transitions. Segment C contains the ball and the club-head after the ball has been hit, and Segment B contains only the ball post club swing. In the spectrogram in Figure 9.3 this corresponds to two sinusoids transitioning to only a single sinusoid near time-step 35. Our minds automatically complete the path of the lower-frequency component backwards to time-step 0. However, at that time-step there is a higher peak at 0.04 Hz/sample, which is likely caused by motion of the golf player. The analyses below will demonstrate the effects of an additional target that is not modelled by the dynamics. In this example the ball contained a reflective material that causes the “wobble” seen in the top frequency component. The discrete factors are, however, too low in resolution to pick this effect up.

The analysis in Figure 9.4 correctly switches from the two-frequency model to the one-frequency model. Some low-frequency noise picked up by BSA causes a “jump” in the frequency component near time-step five. The signal dynamics attempt to smooth the transition between the noise (the unexpected frequency component) and the true position of the club head.

In Figure 9.5 we show the results of changing the dynamics of the analysis to only allow a negative frequency chirp for the bottom frequency component.

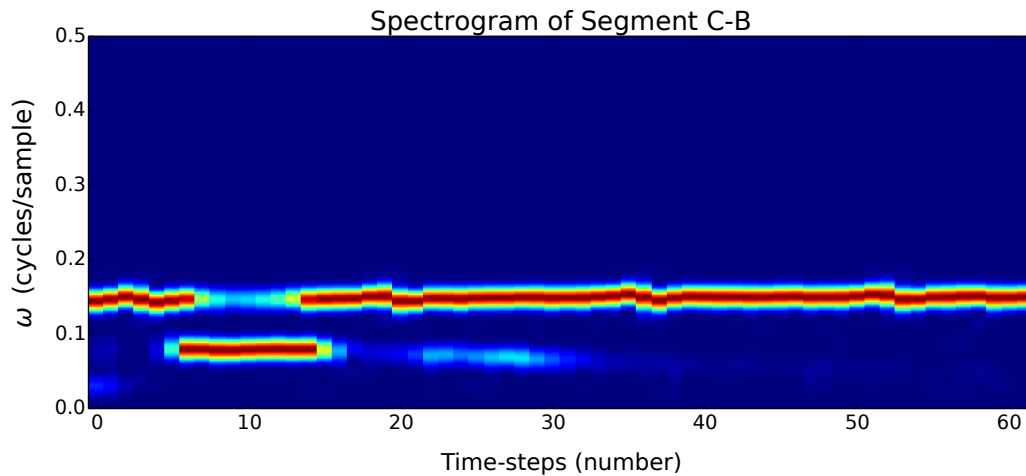


Figure 9.3: *The spectrogram of Segment C and B of Doppler radar data. There are two strong sinusoids visible, one stationary (the ball) and one with a negative chirp (the club head). Note the presence of low-frequency components, which are stronger than the club head at time-step 0. There is also a slight “wobble” in the frequency component of the ball caused by a reflective material that was placed in the ball.*

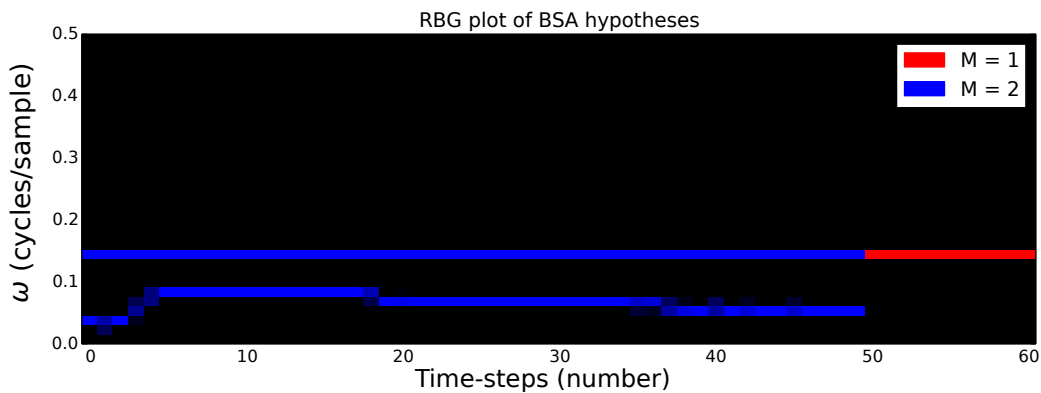


Figure 9.4: *PGM analysis of Segment C and B of Doppler radar data. The segment transitions are correctly estimated. However, the lower-frequency noise is incorrectly seen as part of the club head between time-step 0 and 5.*

This is done by setting the probabilities of positive values of $\Delta\omega_1$ to zero. The analysis now correctly extrapolates the frequency parts that are missing from the BSA factors. If we know something about the nature of the signal dynamics, restricting it to a subset of behaviour can be beneficial to smoothing the signal.

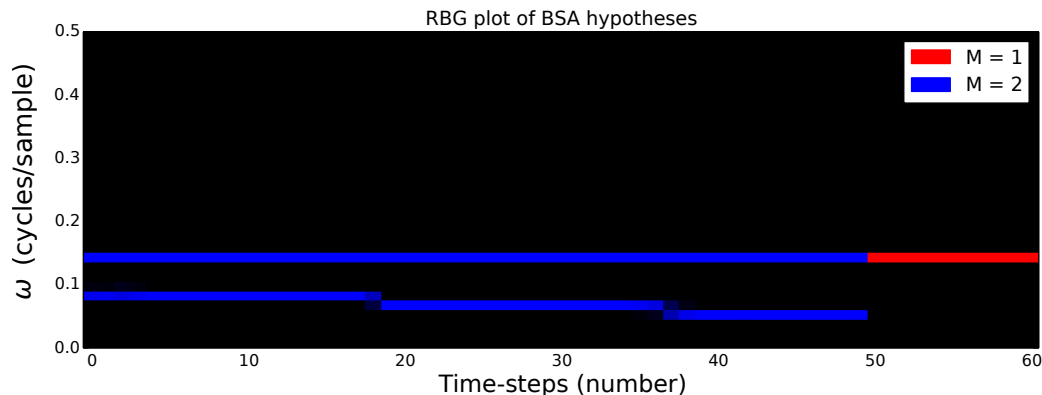


Figure 9.5: PGM analysis of Segment C and B of Doppler radar data with a restriction of a negative chirp on the lower-frequency component. The segment transitions are correctly estimated. Since the lower-frequency component is restricted to having a negative chirp, the model can now correctly smooth the club head's frequency backwards in time.

9.2.2 Discrete: Segment A-E-G

Figure 9.6 is the spectrogram of a golf swing without a ball containing Segments A, E and G. To simplify the analyses, we manually removed the low-frequency component near time-step 35 from consideration in the BSA factors by setting them to zero. In this example there are three segments but only two possible BSA models. The transitions are from a one-frequency component to two components and back to one component. Modelling this left-to-right behaviour would not be possible without our parallel Markov chain, which describes the segment transitions. The same underlying BSA model is used by both Segment A and Segment G. The example from Section 9.2.1 could, however, be modelled without a Markov chain if the two-frequency model was restricted to transitioning to a one-frequency model.

Figure 9.7 contains the analysis of the signal without any left-to-right segment transitions enforced. The analysis incorrectly shows that there is only a single frequency component present throughout most of the signal. This is despite the harmonics that can be seen in Figure 9.6. Remember from Section 4.7.3 that we need to manually adjust the noise component in order for our PGMs to make a reasonable estimation of the underlying BSA model. Without this adjustment, BSA would select two very close frequency components for the club-head. Note the magenta (purplish-red) parts of the analyses at time-step 0 and near time-step 40. The magenta colour indicates that the analysis considered both the one-frequency and two-frequency models as likely hypotheses for those time-steps. These time-steps contrast with the results in the previous section in Figure 9.4 and Figure 9.5 where the PGM was very certain of the BSA model.

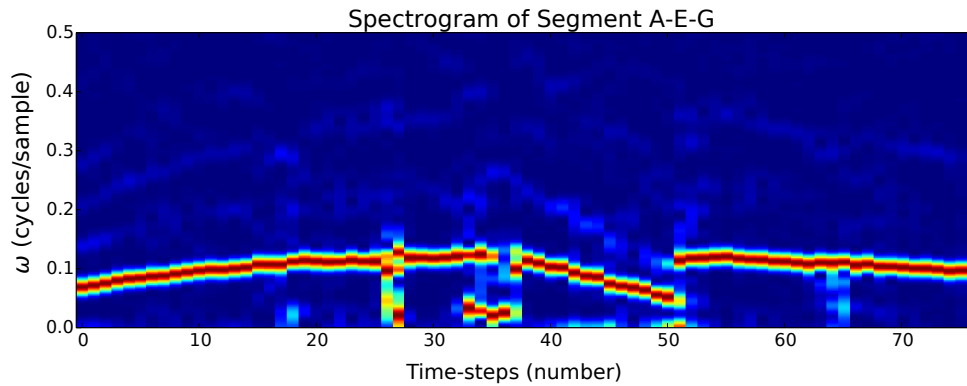


Figure 9.6: The spectrogram of Segments A, E and G of Doppler radar data. There is one frequency component (club head), then two (club head and stick-flash) and then one again (club head). (We ignore the low-frequency component near time-step 35.) However, the club head is obscured during the stick-flash. There are also harmonics present in the spectrogram as there was clipping present in the signal.

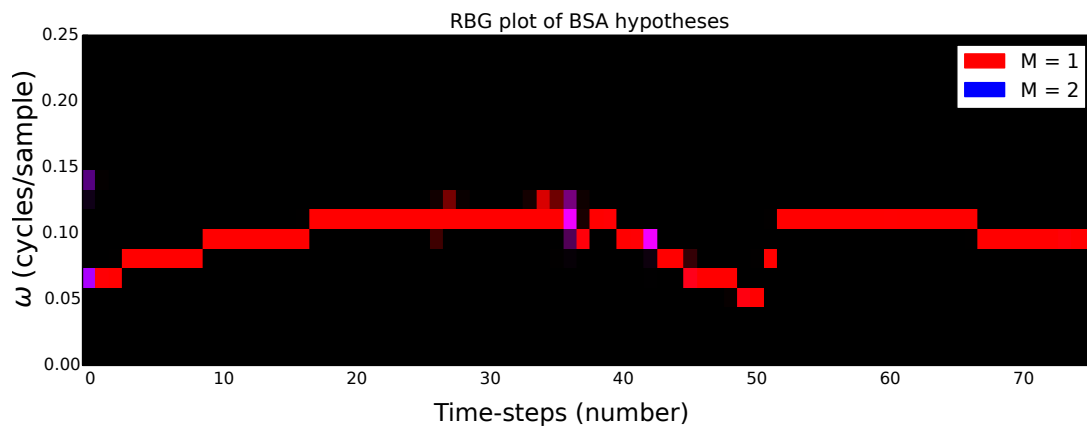


Figure 9.7: PGM analysis of Segment A, E and G of Doppler radar data without left-to-right segment transitions. The segment transitions are not correctly estimated, with the single-frequency model being chosen as the most likely for the most of the signal. Note the magenta parts of the analysis, which indicate a mixture of single- and two-frequency models near time-step 0 and 40.

Including the left-to-right segment transitions improves the analysis as can be seen in the following example. In contrast to the model without left-to-right transitions in Figure 9.7, the analysis in Figure 9.8 correctly shows that there are two signals present during the stick-flash, and smooths over the missing BSA by extrapolating the results from either side. Note that the same BSA

model occurs in Segment A and G but are related to different segments in the analysis.

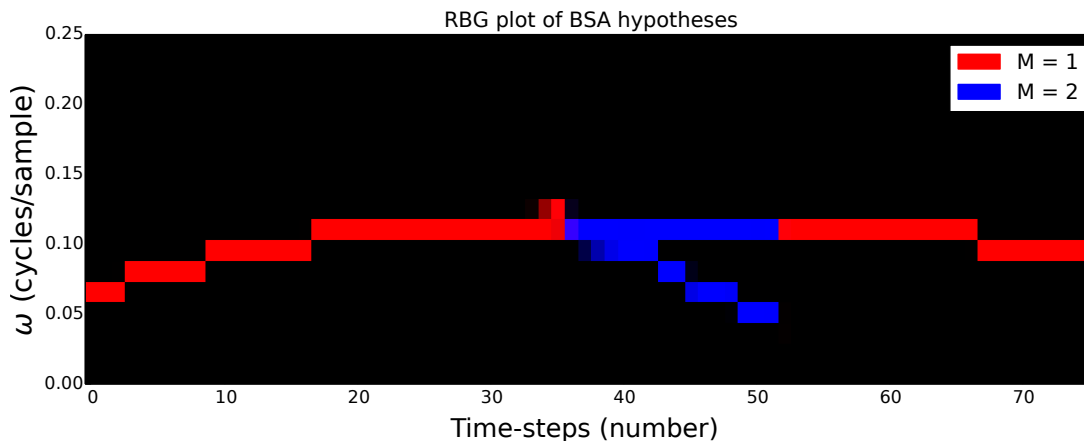


Figure 9.8: PGM analysis of Segment A, E and G of Doppler radar data which includes the left-to-right segment transitions. The segment transitions are correctly estimated with high certainty. The analysis smooths over the part where the club head was not visible in the spectrogram.

9.3 Hybrid Model Results

In order to calculate the results for visualisation, we calculate the Gaussian probability density functions using the mean and covariance matrices of the marginal probability densities. The results of the hybrid model often have very sharp peaks, since the underlying BSA factors are very sharply peaked as we saw in Chapter 4. The resulting sharp BSA probability densities are not present in the discrete analyses, which have a very low resolution due to binning.

9.3.1 Hybrid: Segment C-B

In this section we discuss the results for a signal with the same segment transitions as in Section 9.2.1. Figure 9.9 contains the spectrogram of a signal with very similar behaviour to the signal from Figure 9.3. In this example the ball did not contain a reflective material.

Figure 9.10 contains the PGM analysis of the signal using the hybrid model. It correctly transitions from the two-frequency model to the one-frequency model. The peaks are so sharp that they are barely visible. The hybrid model analysis is similar to the discrete analysis in Figure 9.4 in the sense that it also saw the low-frequency player motion frequency component at 0.04 Hz/sample

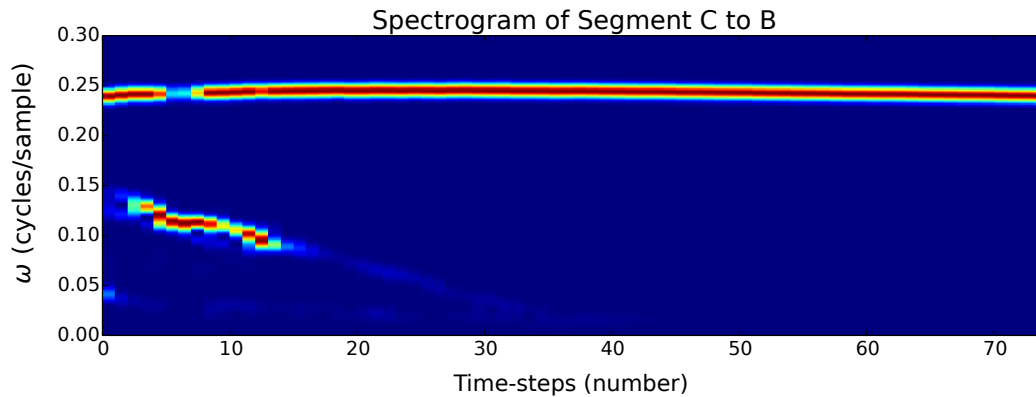


Figure 9.9: *The spectrogram of Segment C and B of Doppler radar data. This signal is similar to Figure 9.3. There are two strong sinusoids visible, one stationary (the ball) and one with a negative chirp (the club head). Note the presence of low-frequency components, which are stronger than the club head at one point.*

as part of the club head. In this case, however, we are not able to restrict the signal dynamics of the lower-frequency component to a negative chirp, as Gaussian probability densities cannot be restricted to only positive or negative values. Since the peaks are so sharp and difficult to see, we enlarge the top and bottom parts of the analysis in Figure 9.11 and Figure 9.12 respectively. Decreasing the process noise will not work in this example, since the PGMs start to encounter issues with computational problems when inverting submatrices.

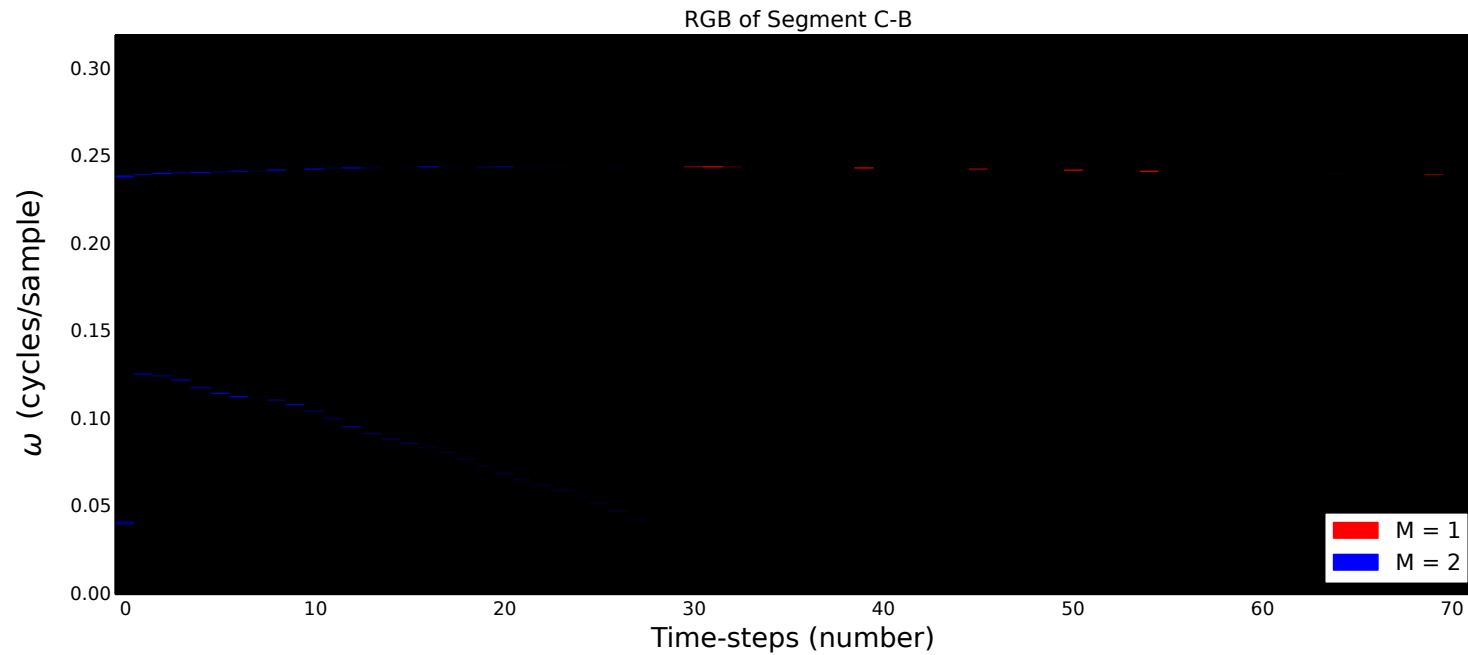


Figure 9.10: PGM analysis of Segment C and B of Doppler radar data. The segment transitions are correctly estimated. However, the lower-frequency noise is incorrectly seen as part of the club head at time-step 0. This is similar to the discrete model analysis in Figure 9.4. For the hybrid model, the peaks are very sharp and almost not visible as a result. This is caused by the sharp BSA factor peaks.

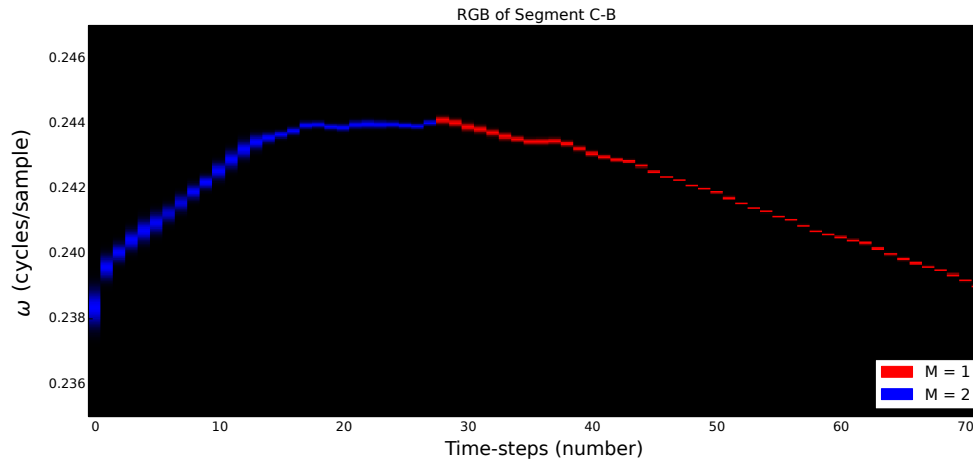


Figure 9.11: An enlargement of the top region of Figure 9.10. Note the sharp peaks.

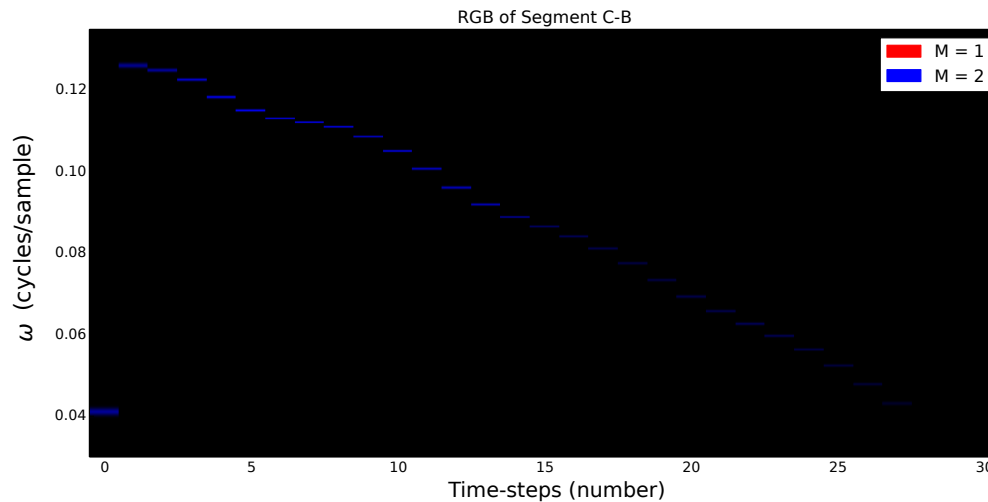


Figure 9.12: An enlargement of the bottom region of Figure 9.10. Note the sharp peaks and the “jump” of the frequency component between time-step 0 and 1.

As a second experiment, we use artificially created covariance matrices with a variance of 0.01 but still use the the original BSA means. The resulting analysis is shown in Figure 9.13. The result is less sharply peaked and follows the linear dynamics of the PGM more closely. In this case we are able to only adjust the ratio between the process noise and measurement noise by artificially increasing the measurement noise.

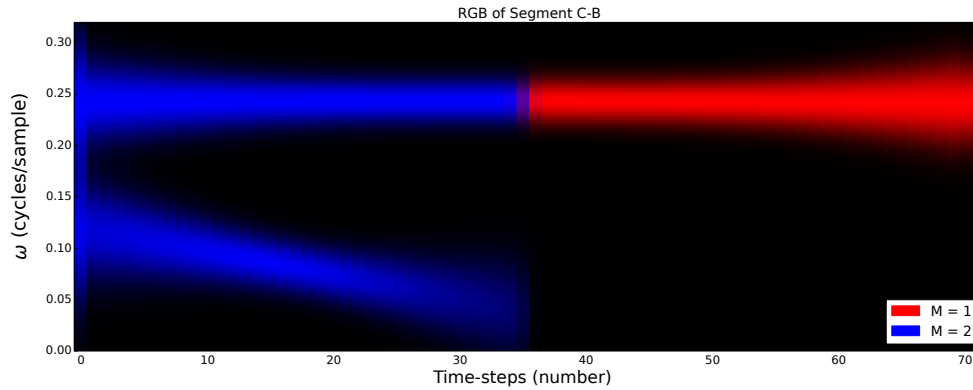


Figure 9.13: PGM analysis of Segment C and B of Doppler radar data using the BSA means but with artificially created covariances. The segment transitions are correctly estimated. The peaks are much less sharp than Figure 9.10 and the signal dynamics much more closely resemble the linear Gaussian dynamics.

9.3.2 Hybrid: Segment F-C-B

In this section we include the use of the three-frequency model, which we could not do for the discrete model. Figure 9.14 is the spectrogram of the signal containing Segment F, C, and B. The spectrogram contains three frequency components (from the top, the ball, club head, and stick-flash), then two (the ball and club head), and then only one (the ball). Note that the club-head frequency component is relatively faint between time-step 0 and 30.

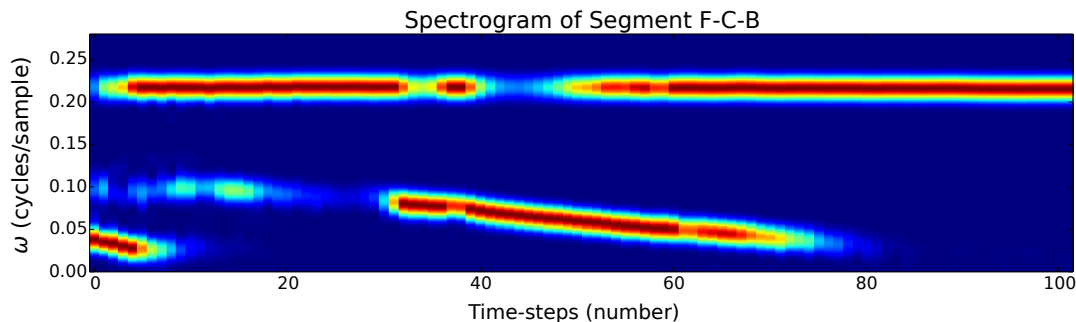


Figure 9.14: The spectrogram of Segment F, C, and B of Doppler radar data. There are three strong sinusoids visible, one stationary (the ball) and two with a negative chirp (the club head in the middle and stick-flash at the bottom). Note the middle frequency component is relatively weak from time-step 0 to 30.

Figure 9.17 contains the hybrid model analysis without the left-to-right segment transitions and Figure 9.15 contains the enlargement of its top region. Note that there are sharp peaks, which are similar to the peaks in Figure 9.10.

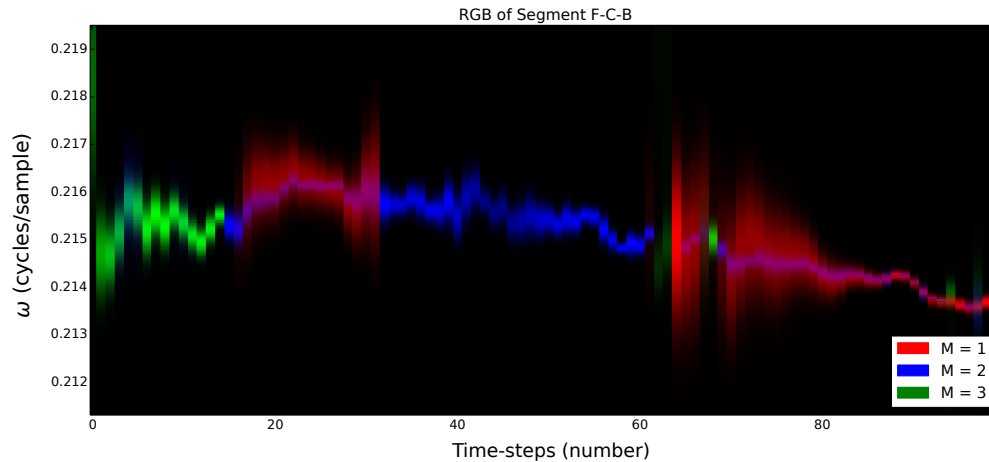


Figure 9.15: *This is an enlargement of the top region of Figure 9.17.*

The magenta and cyan (light-blue) indicate a mixture of possible BSA models similar to that which occurred for the discrete model in Figure 9.7. Where the magenta indicates a mixture of one-frequency and two-frequency models, the cyan indicates a mixture of two-frequency and three-frequency models.

When we include the left-to-right segment transitions in the analysis, we have the result in Figure 9.18, with the top region enlarged in Figure 9.16. In contrast to the previous example, the segment transitions are now correctly estimated.

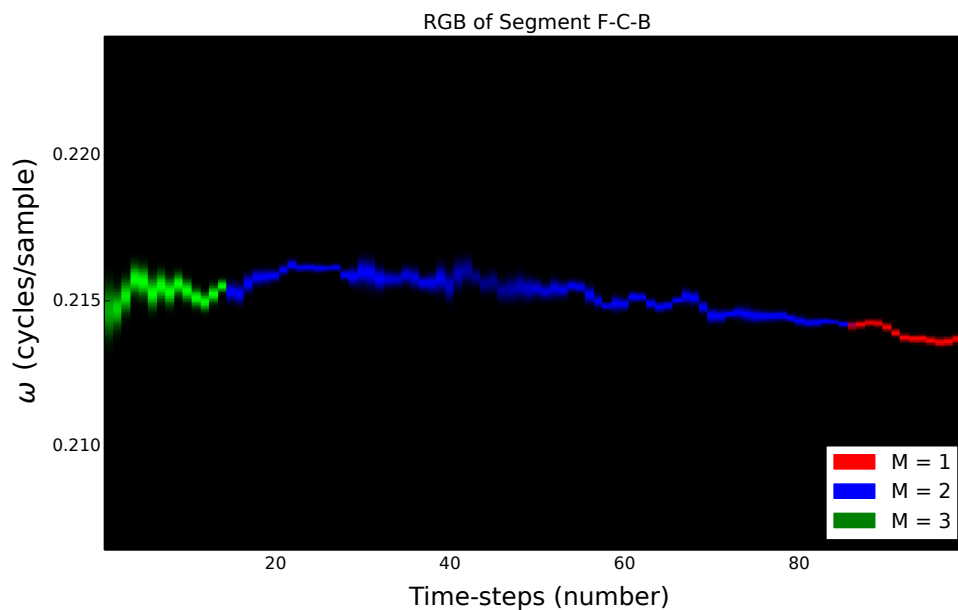


Figure 9.16: *This is an enlargement of the top region of Figure 9.18. Note the sharp peaks*

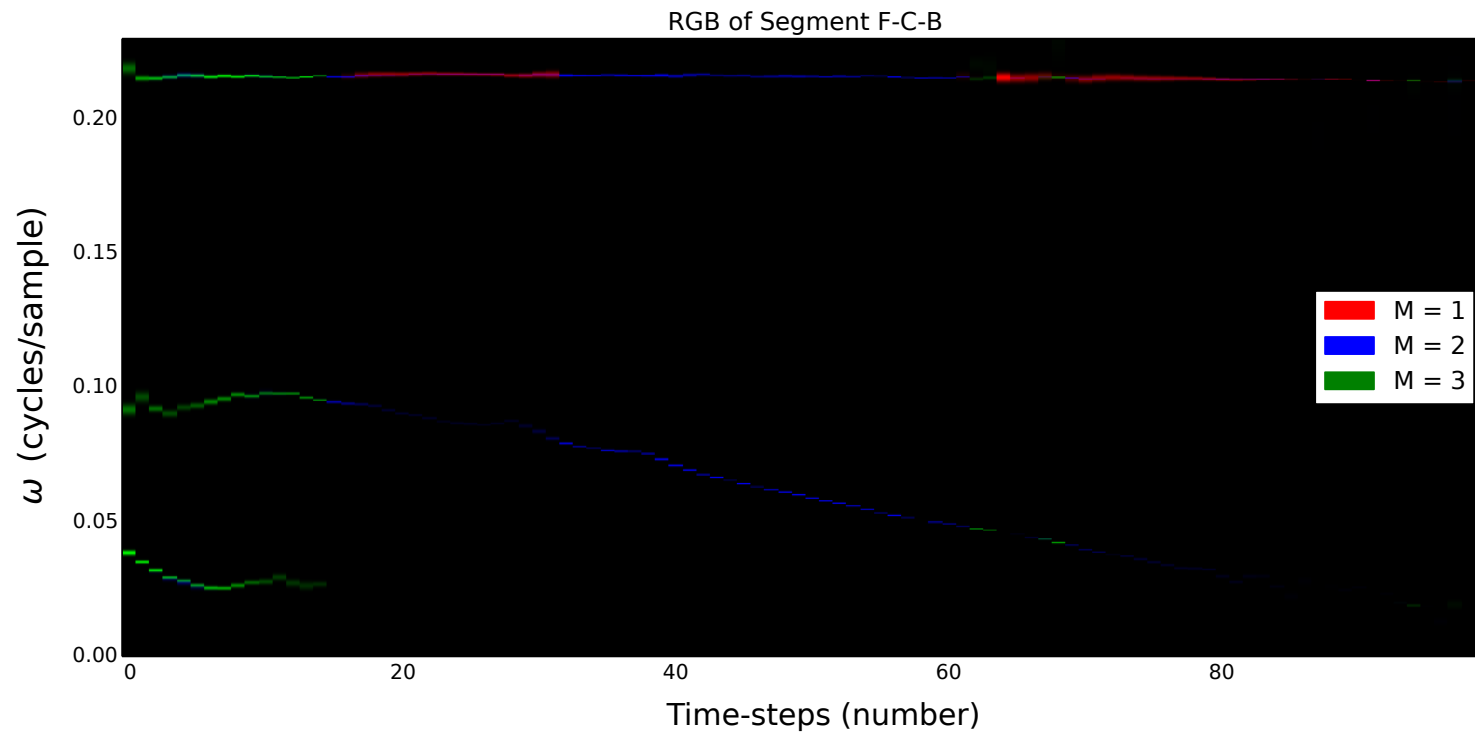


Figure 9.17: PGM analysis of Segment F, C and B of Doppler radar data without left-to-right segment transitions. For the hybrid model, the peaks are very sharp and almost not visible as a result. This is caused by the sharp BSA factor peaks. Note the uncertainty between the active BSA model indicated by the cyan and magenta colours.

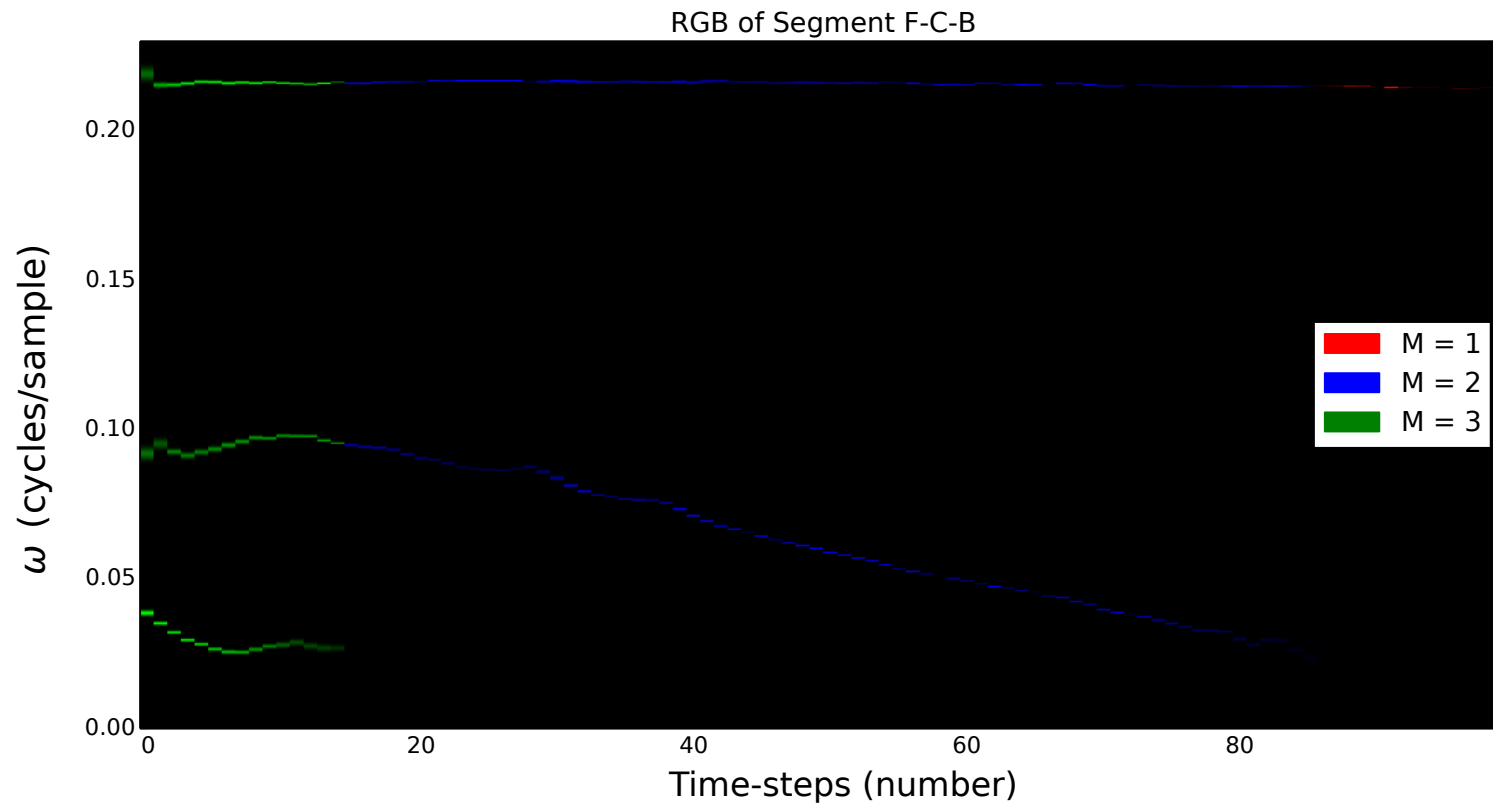


Figure 9.18: PGM analysis of Segment F, C and B of Doppler radar data. The segment transitions are correctly estimated. For the hybrid model, the peaks are very sharp and almost not visible as a result. This is caused by the sharp BSA factor peaks.

9.3.3 Hybrid: Segment A-E-G

For the following analyses we manually specified the time-steps of the segment transitions in order to demonstrate the missed-target model. The signal dynamics are not enough on their own to help the PGM determine the segment transitions without the help of the BSA factors' model probabilities. However, in the case where the frequency is missed, the BSA model probability is wrong. If we did not specify the segment transitions, the result would be much like in Figure 9.7 where the single-frequency model dominates even with left-to-right segment transitions.

Figure 9.19 contains the spectrogram of the signal part under analysis, which is the same signal in Section 9.2.2 but with slightly fewer time-steps. In this case we once again ignore the low-frequency component created by the player's body near time-step 24 for the sake of simplicity. We do so, by disallowing the Nelder-Mead optimisation technique from Section 4.3 to search for a peak in that region.

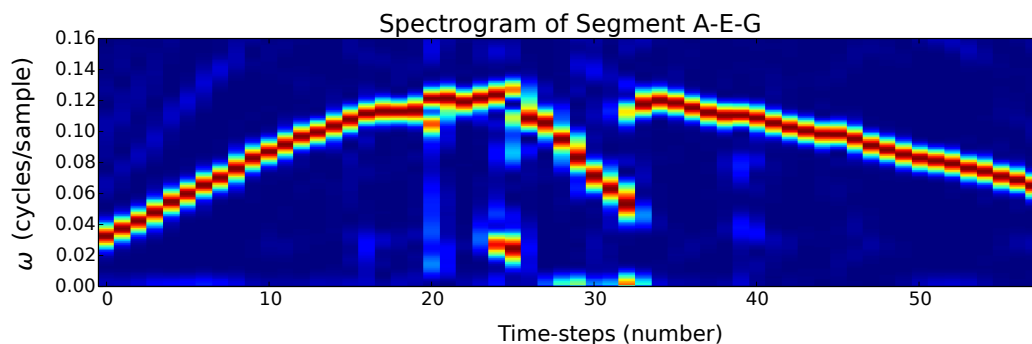


Figure 9.19: *The spectrogram of Segments A, E and G of Doppler radar data. This spectrogram is the same signal as Figure 9.6. There is one frequency component (club head), then two (club head and stick-flash) and then one again (club head). (This is ignoring the low-frequency components between time-step 23 and 33.) During the stick-flash, the club head is obscured.*

Figure 9.20 contains the left-to-right analysis of the signal with the segment transitions specified. BSA did not detect the club head during the stick-flash (time-step 24 to 33) and rather hypothesises two very close signals which are then smoothed by the PGM. We use artificially created covariances such as was done in Figure 9.13 with the result shown in Figure 9.21; however, this still does not give us the desired result of smoothing over the empty region where we expect the club-head to be.

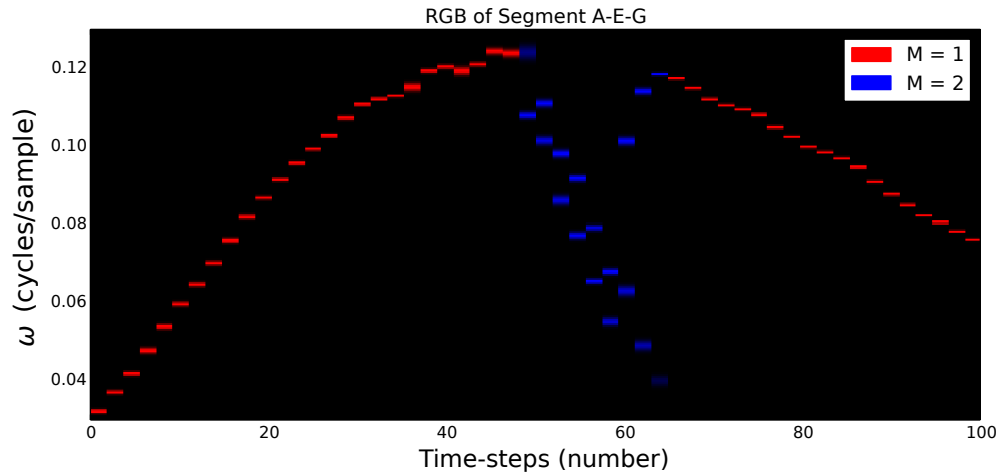


Figure 9.20: PGM analysis of Segment C and B of Doppler radar data. The segment transitions were not estimated and were instead specified in order to demonstrate the difference the missed target model makes in Figure 9.22.

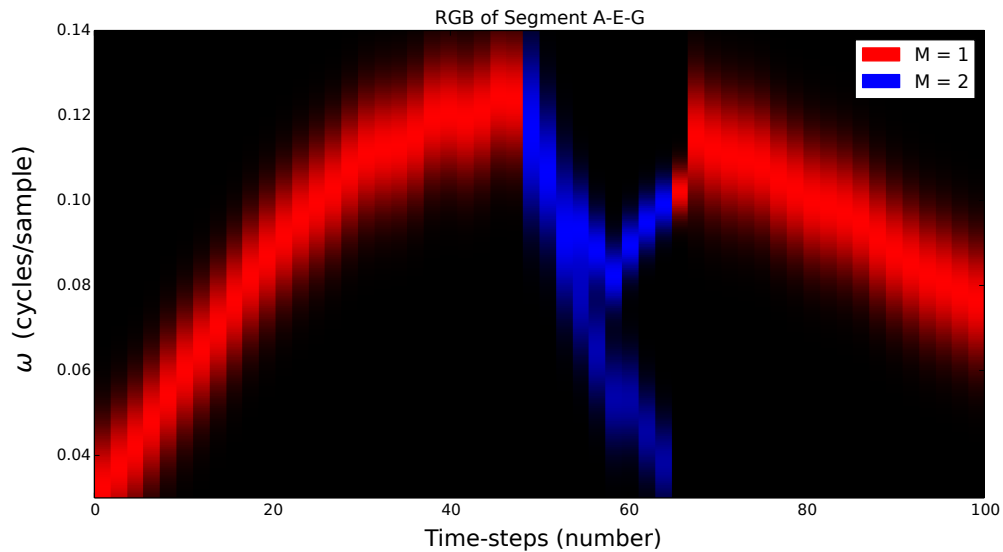


Figure 9.21: PGM analysis of Segment C and B of Doppler radar data using the BSA means but with artificially created covariances. The segment transitions were not estimated and were instead specified in order to demonstrate the difference the missed target model makes in Figure 9.22.

We include the missed target model from Section 7.4.3 for the top frequency component of the two-signal for the analysis in Figure 9.22. To create our two-frequency model (which contains a missed target), we take the lower frequency component, the stick-flash, from the single frequency BSA and we multiply it with a Gaussian with a variance of 1 and mean of 0.15 to create a very uncertain prior for the missed-target frequency component. In this case the PGM can

successfully smooth over the region where the club head was invisible to BSA. For the discrete model with a similar signal in Section 9.2.2, we did not need the missed-target model. We hypothesise that this is due to the noise floor being created by binning the probability density function. The binning from the discretisation creates a noise floor that is high enough for the PGM to find a path for the missed target. The peak of the BSA factor therefore does not have as large an impact if it is wrong in the discrete model. For the hybrid model, however, there is no such floor as the Gaussian function dies away from the peak very quickly. The Gaussian peak therefore has a bigger influence and “pulls” the probability of the signal dynamics towards it.

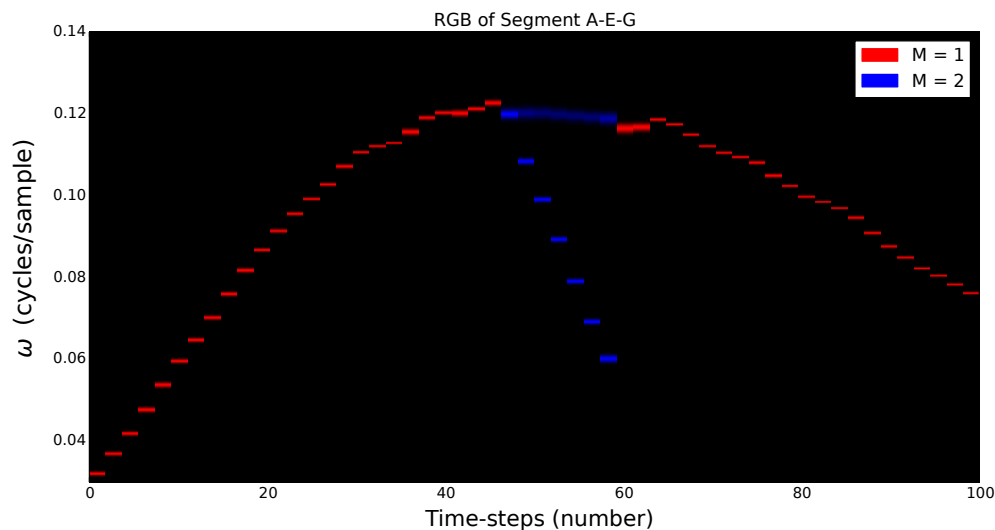


Figure 9.22: PGM analysis of Segment C and B of Doppler radar data. The missed target model is able to smooth the position of the club head where it was not visible to BSA.

9.3.4 Hybrid: Segment A-A'-D

In the next few analyses we include the low-frequency component in Segment A caused by the player’s movement, which had previously been excluded from consideration for the BSA. We call the part containing the low-frequency component Segment A’ to differentiate it from what we had previously only referred to as Section A. We once again make use of the missed-target model but this time it is for Segment D. Since the club head is obscured by the stick-flash, we do not know where the true frequency component should be. We thus need to make an educated guess of the behaviour of this frequency component (even if we do not know enough about it), as it is a requirement for using the missed-target model. We use the end of the player-movement frequency component to signify the start of our missed-target model and presuppose that the missed

target has the same $\Delta\omega$ from when the club head was still visible to BSA. To create our two-frequency model, we use the same approach as in Section 9.3.3.

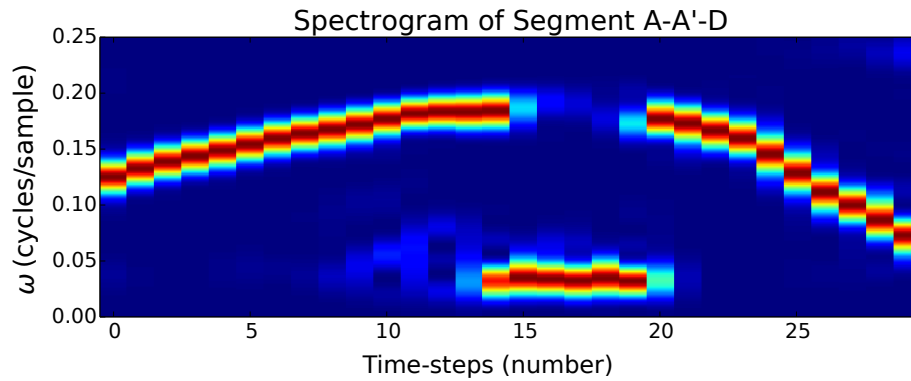


Figure 9.23: *The spectrogram of Segments A and D of Doppler radar data. There is one frequency component (club head), then two (club head and player motion), and then a different two (club head and stick-flash). In this example we distinguish between Segment A without player motion as A, and with player motion as A'. However, during the stick-flash, the club head is obscured.*

Figure 9.24 contains the analysis *without* using the missed-target model where the signal segment transitions are specified. BSA could not detect the club head during the stick-flash, and models two close frequency components.

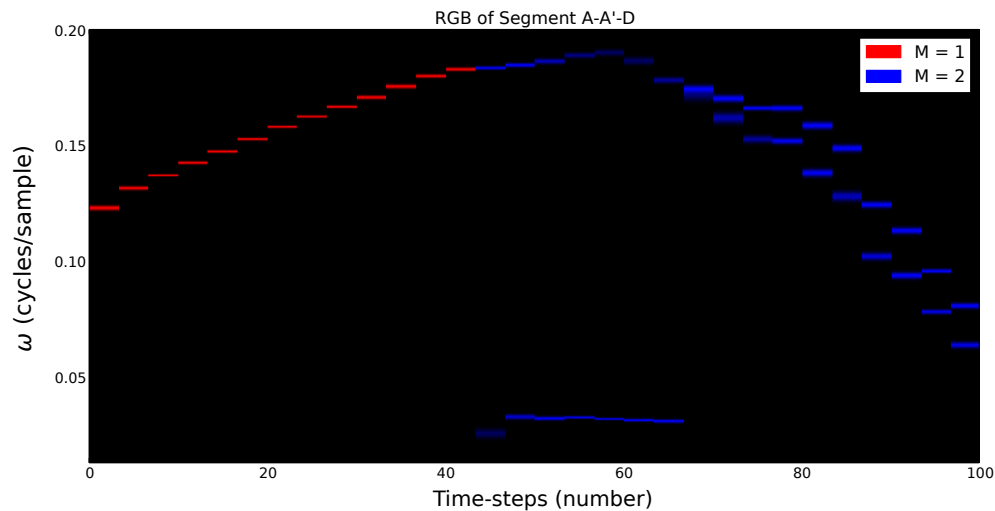


Figure 9.24: *PGM analysis of Segment A and D of Doppler radar data. The segment transitions were not estimated and specified in order to demonstrate the difference the missed-target model makes in Figure 9.25.*

Figure 9.25 contains the analysis with the missed-target model where it was unnecessary to specify the signal segments as the system automatically detected the correct signal segments. In this case the PGM had more information specified than in Section 9.3.3 in the form of the low-frequency component caused by the player's body. The disappearance of a frequency component allows the PGM to detect the segment transition. The missed-target model can now extrapolate the position of the club head during Segment D based on the behaviour of the club head during Segment A'. The colour becomes fainter over each time-step since there are no longer any BSA factors indicating the position of the club head and the effects of the process noise build up over time.

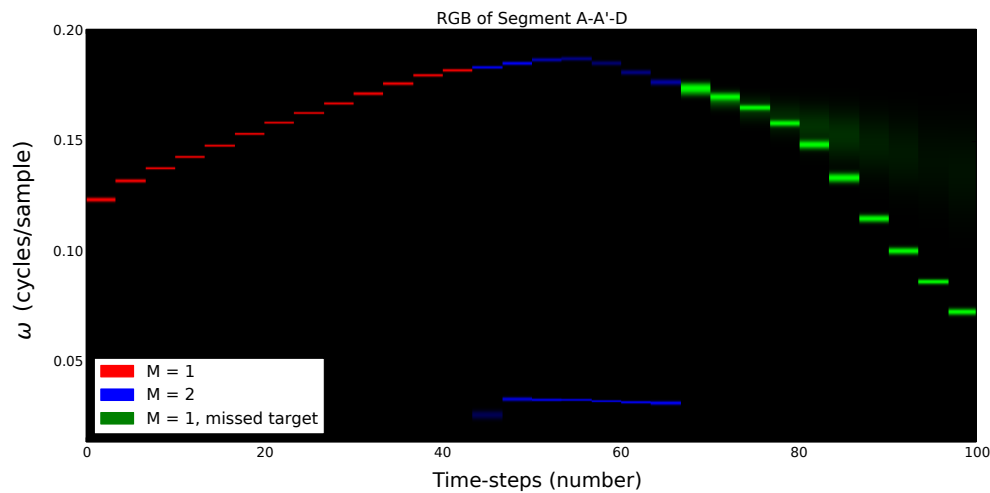


Figure 9.25: PGM analysis of Segment C and D of Doppler radar data. The missed-target model is able to smooth the position of the club head where it was not visible to BSA. Note the faint colour of the club head. It continues as a linear extension from the previous segment without receiving any information from the BSA factors.

We enlarge Segment D in Figure 9.26 and indicate the mean of the missed-target model. The frequency component continues its downwards trend over the time-steps, becoming ever fainter as the PGM becomes ever less certain of its position.

9.3.5 Hybrid: Segment A-A'-D-F-C-B

In this analysis we combined the analyses from Section 9.3.2 and Section 9.3.4. We still needed to specify the signal segments, but were able to smooth the frequency components and extrapolate the frequency component that was missing in Segment D. Figure 9.28 contains the analysis of the entire golf swing.

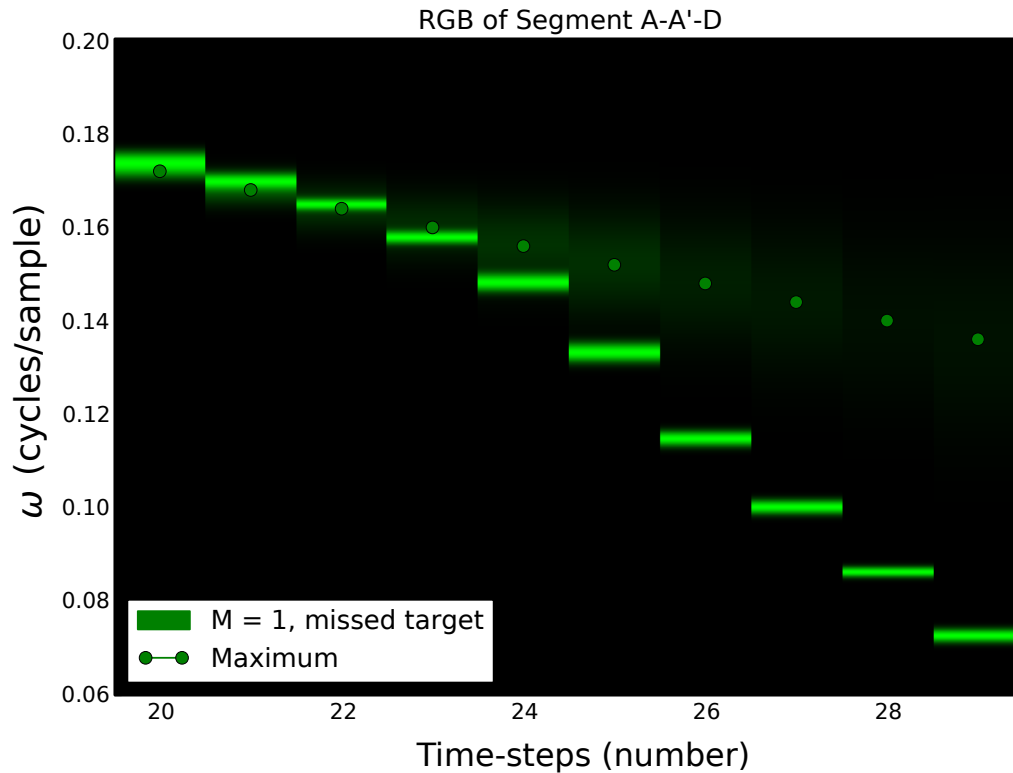


Figure 9.26: This is an enlargement of the top region of Figure 9.25. We include the maximum values of the club-head position. Note that the PGM becomes more uncertain of the club head position for each time-step.

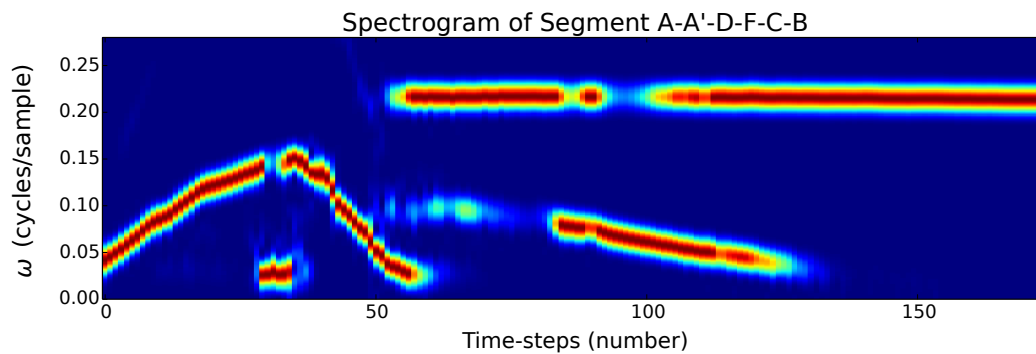


Figure 9.27: The spectrogram of Segments A, D, F, C, and B of Doppler radar data. This spectrogram contains the frequency components from Figure 9.23 and Figure 9.14.

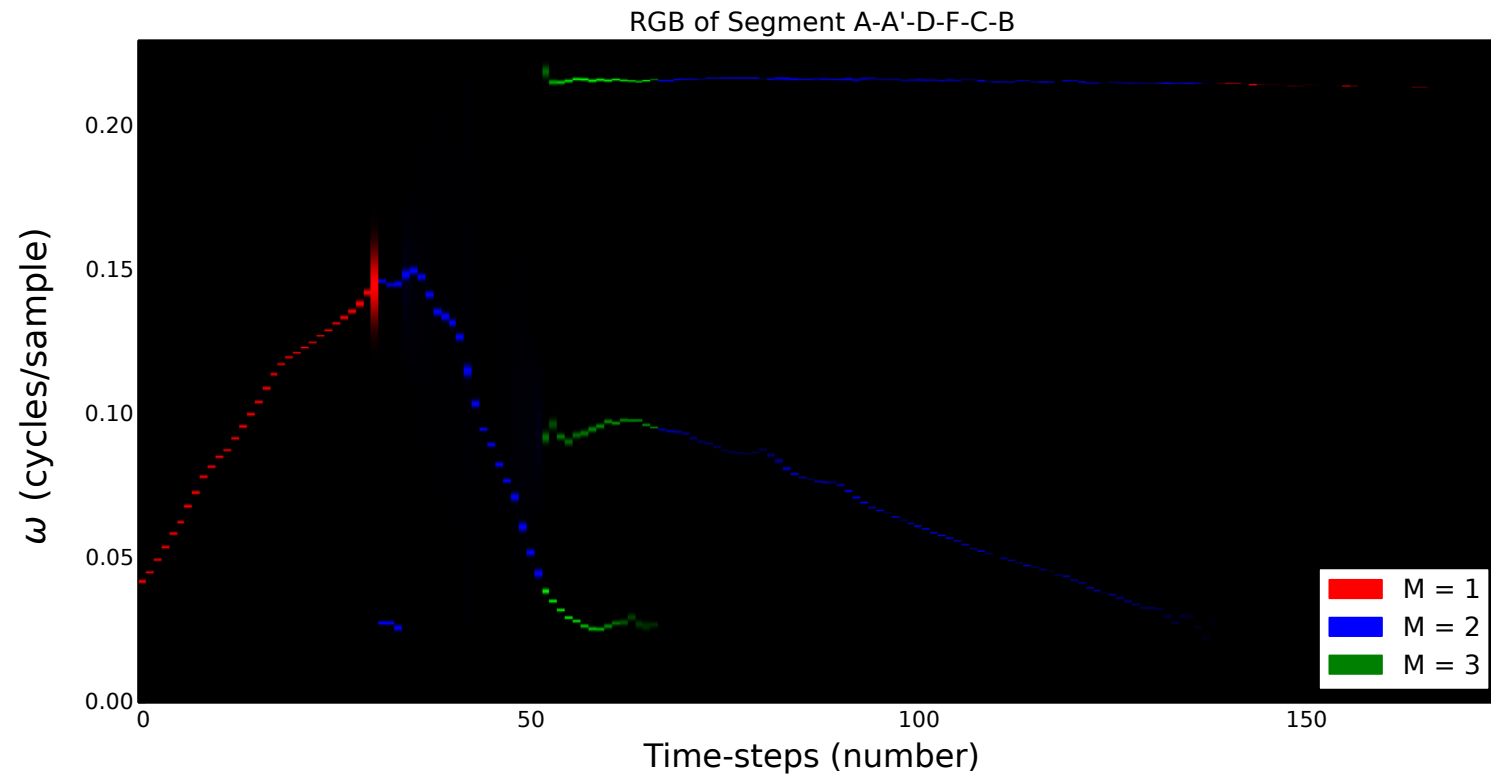


Figure 9.28: PGM analysis of Segments A, D, F, C, and B of Doppler radar data. The missed-target model is able to smooth the position of the club head where it was not visible to BSA from time-steps 34 to 51. The smoothed position is, however, so faint that one cannot see it in this figure. The segment transitions were not estimated and needed to be specified. The large variance near time-step 30 (the last step for $M = 1$) is likely due to a combination of the faint frequency component which can be seen in the spectrogram and the jump in frequency to the new BSA model.

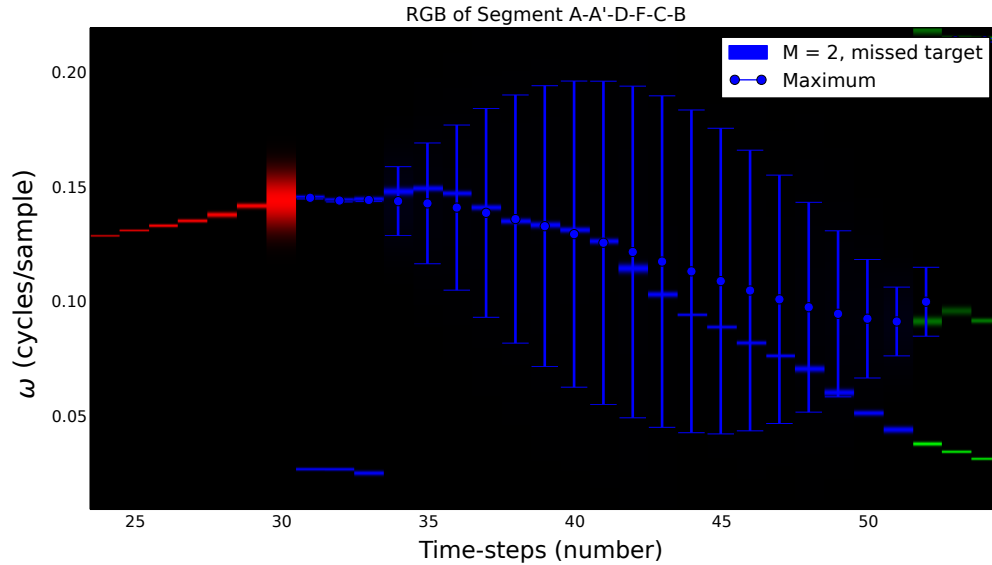


Figure 9.29: *This is an enlargement of the first region of Figure 9.28. The smoothed club head is very faint, and we therefore included the error bars and the mean. Note that the PGM is more certain of the club-head position at the edges of the missing-target model's segment.*

Figure 9.29 is an enlargement of the first region of Figure 9.28. We include error bars in order to demonstrate the uncertainty of the missed-target model. In this example, once again, the PGM did not change the initial BSA estimates very much. Decreasing the process noise will also not work in this example such as in Section 9.3.1.

9.4 Conclusion

We had moderate success in smoothing the Doppler radar frequency components where some parts of the signals were more easily smoothed and others. It struggles somewhat more on the simulated data without the factors adding extra information. The methods developed, namely the left-to-right Markov chain and the missed-target model, allowed us to improve upon the conventional multitarget tracking. These methods compensated for our uncertainty surrounding the behaviour of the Doppler radar signals. With more domain knowledge, which could be more easily incorporated in the PGM framework, the smoothing of the Doppler radar frequency components could be improved even further.

Chapter 10

Conclusion and Recommendations

10.1 Conclusion

In this thesis we demonstrated a Bayesian approach to signal processing. The signals of interest, containing the Doppler shifts of objects moving during a golf swing, were analysed using BSA and then smoothed using PGMs. In order to develop our approach we combined three different methods, namely BSA, state-space estimation, and PGMs. Our approach resulted in spectrogram-like Bayesian analyses of the frequency components in the Doppler radar signals. While we did not automate all the processes related to the multitarget tracking of the golf swing, we were able to demonstrate how a Bayesian approach can be advantageous. This is a systematic approach where the principles can be applied to spectrum analyses and multitarget tracking. Some of the problems we solved include how to model the progression of different parts of the signal (segment transitions), and what to do when the measurements are missing (missed-target model) or too confident (we increased the model noise).

10.1.1 Doppler Radar Signals

The signals we analysed were measured Doppler radar signals where the Doppler effect allowed us to relate relative frequency shifts to the physical objects that move in a golf swing. These signals, however, contain effects such as thermal noise, multi-path effects, clipping, and frequency components that are not well understood. Classical methods such as the DFT and spectrogram have no concept of noise, unwanted target, or missed target. Analysis of the Doppler radar signals can benefit from the techniques we explored in this thesis, as they do contain these concepts.

10.1.2 Bayesian Spectrum Analysis

We discussed Bayesian probability theory and demonstrated the use of Bretthorst's BSA [8]. Bayes' theorem allows us to estimate a model's parameters by treating them as random variables and to compare different competing models. We demonstrated these concepts in BSA on simulated and Doppler radar data signals where the models did not make the implicit assumption of single sinusoid in white noise such as is made by the periodogram [8].

We performed two types of analyses on the Doppler radar data: a long segment analysis and a shorter time-step spectrogram-like analysis. The long segments relate to the number of targets and dynamics present in each part of the signal. The mismatch between our models and the true signal was seen as noise by BSA, which is not present as a concept in the periodogram or spectrogram.

In the long segment analyses, the selected models suggested that the underlying complexity of the signal was greater than our BSA models could explain. The model selection favours the more complex models despite the penalties Bayesian probability theory imposes for having more complexity (such as having a larger number of parameters). This is not a failure of BSA, but a case of our assumptions about the signals being too strict and the frequency behaviour being more complex than is explained by our models. The shorter analyses reduced the effects of the non-linearities in the Doppler radar signal. However, the short signal analyses still suggested that the Doppler radar data contain more complexity than our models are able to explain, such as that there are possibly multiple very close signals (so close that their peaks overlap in the DFT). The appearance of the multiple very close frequency components is likely due to an object reflecting a range of different frequencies that are very close to one another and that relate to the different velocities of different parts of the object. Explicitly specifying the amount of noise in the signal in our BSA models allowed the simpler signals to be selected more often, which supports our hypothesis about the signal complexity. Less of the signal thus needs to be attributed to the signal model when increasing the noise power.

There tended to be a single strong frequency or group of very close frequencies created by the object that we are interested in. This proved to be invaluable for use in our PGMs, since with the time-step analyses the problem can be framed as a multitarget state-space estimation with the frequency components seen as targets.

10.1.3 State-space Estimation

We discussed how multitarget state-space estimation can be framed as Bayesian inference [11]. In conjunction with the BSA, we could use state-space estimation for smoothing the Doppler radar data parameters in a form in which PGMs can be used.

We discussed three different approaches used in the literature, TOMHT [11], switching state-space estimation [18], and JMPDT [19]. Our own model is a combination of multitarget tracking with a left-to-right transition of signal segments which includes additional temporal information.

10.1.4 Probabilistic Graphical Models

The PGMs we used are called cluster graphs. They are loopy, undirected graphs, which can represent the state-space estimation transitions as probability factors that factorise the joint probability density. We discussed the two types of factors used in this thesis and their operations – namely the discrete factor and the hybrid discrete Gaussian factor. We discussed the different algorithms that can be used to perform inference on them, specifically LBP [33], LBU [14], as well as the method of dynamically scheduling messages, RBP [35]. We used LBU with RBP scheduling.

The concepts from Bayesian state-space estimation allowed us to model our approach to smoothing the Doppler radar data's frequency components as PGMs. We developed a model in which the traditional state-space estimation transitions are combined with a discrete Markov chain for the signal segments discussed in Section 4.6. In order to exercise our intuition about the problem, we represent the model as a Bayes network and then implement it as cluster graphs. Using the parallel Markov chain, we could enforce a left-to-right progression of the Doppler radar data signal segments which relate to the different parts in the signal. The factors used for the discrete and hybrid models differed in that more conditional independences were included in the discrete model to allow for computational tractability. We also constructed a missed-target model for the hybrid model.

The challenges we faced included message-passing constraints and process noise for the hybrid PGM. As expected, the discrete model had more computational constraints in respect of the number of frequency components we could represent. We also needed to consider the different effects that LBU and LBP have on the approximate message passing of our hybrid PGM. We make use of LBU, since both [14] and [40] suggest that the LBU approximation potentially focusses on the more relevant parts of the probability space.

While our approach could not generalise to the whole Doppler radar signal, we were able to successfully demonstrate the concepts discussed in the above sections on parts of the signals. The models we constructed allowed us to include domain knowledge in a probabilistic and systematic manner. Our use of the segment transitions and missed-target model improved the standard multitarget tracking. The components developed in this thesis can be used in other multitarget tracking applications.

10.2 Recommendations and Future Work

In this section we discuss approaches that could further assist in achieving the goal of fully automated target tracking of the Doppler radar data. Of the recommendations mentioned here, we suggest that the most promising approaches would be systematically quantifying the measurement and process noise for the model dynamics, and using explicit duration models.

Future work could aim to:

- Employ better methods in estimating process noise. As mentioned in Section 8.3.3, one of the difficulties in creating our state-space estimation models lay in selecting our process noise for our linear Gaussian transitions. The process noise is also not necessarily the same throughout the entire signal.
- Use explicit duration models (See Appendix E), which can include more temporal information on the duration of the signal segments.
- Use more complex BSA models to better capture the Doppler radar data behaviour. The additional parameters could also be used within the PGMs to better smooth the frequency components.
- Alternatively explore the noise in the Doppler radar data further by estimating the signal noise given the BSA model. See Section A.4 for Bretthorst's [8] formulation of estimating the noise variance.
- Include the level of trust in the BSA results as a factor in the PGMs. This could effectively increase the "measurement" noise in the models and mitigate the sharp peaks in the BSA results.
- Use alternate discretisation approaches of the factors such as Particle Filters [27] or MCMC techniques [42]. These methods have other challenges to overcome, such as determining the burn-in period and the curse of dimensionality [43].

- Use Gaussian mixture models [44], which would be able to approximate multi-modal probability density functions and avoid weak marginalisation.
- Model the position of the objects being tracked and not merely smooth the relative frequency shifts. This would require using the multiple channels in the Doppler radar unit and the phase differences between them.

Appendices

Appendix A

BSA Detail

In this chapter we give a more detailed overview of Bretthorst's generalised approach to BSA in the case of a general signal model and will closely follow the calculations done in [8, 21, 22].

A.1 BSA Parameter Estimation

With N data points, $D = \{d_1, \dots, d_N\}$, sampled from $y(t)$, the continuous signal, at discrete times $\{t_1, \dots, t_N\}$, the signal model $M(t)$ and signal noise $e(t)$, we have the following equation to describe our data:

$$d_i = M(t_i) + e(t_i), \quad (1 \leq i \leq N). \quad (\text{A.1.1})$$

All of the functions in the above equation are functions of time that have been sampled at discrete times. The noise is assumed to be additive and non-correlated.

The general form of the model as from [8] is:

$$M(t_i) = \sum_{j=1}^m B_j G_j(t_i, \Theta) \quad (\text{A.1.2})$$

with B_j the amplitude multiplied with the j th model function $G_j(t_i, \Theta)$, out of m possible model functions. The set of r parameters that are going to be estimated are

$$\Theta = \{\Theta_0, \dots, \Theta_{r-1}\}.$$

The model parameters can be frequencies, chirp rates, signal decay or any arbitrary set of parameters used to describe the signal.

The case of estimating the frequency of a single sinusoid in noise gives us the following model [8]:

$$M(t_i) = B_1 \cos(2\pi\omega t_i) + B_2 \sin(2\pi\omega t_i),$$

with $m = 2$ model functions, cos and sin, with unknown amplitudes, B_1 and B_2 , and a single model parameter for the frequency of the signal, ω . Note that the assumption is made that each of the amplitudes in the set $\mathbf{B} = \{B_1, B_2, \dots\}$ is constant throughout the signal. This is not a limitation, as one can include and parameterise the change in amplitude in the model function, $G_j(t_i, \Theta_j)$.

The parameter estimation from Equation 3.2.4 is

$$p(\boldsymbol{\Omega}|D, M_j, I) = \frac{p(D|\boldsymbol{\Omega}, M_j, I)p(\boldsymbol{\Omega}|M_j, I)}{p(D|M_j, I)}.$$

Here the model parameters, $\boldsymbol{\Theta}$ and model amplitudes \mathbf{B} , are our hypothesis so that $\boldsymbol{\Omega} = \{\boldsymbol{\Theta}, \mathbf{B}\}$. We will not write M_j on the right of the conditioning bar, since for parameter estimation the implicit assumption that the correct model is selected can be encoded in I . Using the above equation, the probability of the model parameters, $\boldsymbol{\Theta}$, and nuisance amplitude parameters, \mathbf{B} , given the data and other prior information is

$$p(\boldsymbol{\Theta}, \mathbf{B}|D, I) = \frac{p(D|\boldsymbol{\Theta}, \mathbf{B}, I)p(\boldsymbol{\Theta}, \mathbf{B}|I)}{p(D|I)}.$$

The probability of parameters given only the information I represents the state of knowledge about the parameters before the data were measured, which is our prior, $p(\boldsymbol{\Theta}, \mathbf{B}|I)$. The other factors in the equation are $p(D|\boldsymbol{\Theta}, \mathbf{B}, I)$ the likelihood function, $p(\boldsymbol{\Theta}, \mathbf{B}|D, I)$ the posterior, and $p(D|I)$ the normalising constant [21]. Our goal is to calculate the parameters of interest $\boldsymbol{\Theta}$ and marginalise out the amplitudes \mathbf{B} to get

$$p(\boldsymbol{\Theta}|D, I) = \int d\mathbf{B} p(\boldsymbol{\Theta}, \mathbf{B}|D, I),$$

which is the probability of the parameters given the data.

A.1.1 The Likelihood Function

The model parameter and amplitude likelihood function is constructed by taking the difference between the model function, $M(t_i)$ or the “true signal”, and the data, D , which gives us the noise, $e(t_i)$. Jaynes [7] and Bretthorst [8] justify using a Gaussian prior for the noise by using the principle of maximum entropy and the central limit theorem as discussed in Section 3.2.6. Generally not much is known about the noise except that the noise carries finite total power.

The following calculations come from [8]. The probability density of a noise sample at time t_i given prior information I is

$$p(e(t_i)|\sigma, I) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{e(t_i)^2}{2\sigma^2}\right\},$$

where σ is the standard deviation of the noise values $e(t_i)$. *Everything* not explained by the parameter values of the model will be made part of the noise, and the accuracy of the parameter estimates depends on the estimated noise variance [21].

Using the product rule of probability theory for the whole set of $e(t_i)$ values, for $1 \leq i \leq N$, and supposing that they are independent gives us:

$$p(e(t_1), \dots, e(t_N) | \sigma, I) = \prod_{i=1}^N \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{e(t_i)^2}{2\sigma^2} \right\} \right]. \quad (\text{A.1.3})$$

From Equation A.1.1, we can see that the error $e(t_i)$ is the difference between our model and the data observed. It follows that the probability that one should obtain this particular noise sample, which is also the likelihood L of the parameters under consideration, is

$$p(D | \mathbf{B}, \Theta, \sigma, I) = L(\mathbf{B}, \Theta, \sigma) = (2\pi\sigma^2)^{-(N/2)} \prod_{i=1}^N \left[\exp \left\{ -\frac{[d_i - M(t_i)]^2}{2\sigma^2} \right\} \right].$$

The multiplication can be written as a sum inside the exponent function:

$$L(\mathbf{B}, \Theta, \sigma) = (2\pi\sigma^2)^{-(N/2)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^N [d_i - M(t_i)]^2 \right\} \quad (\text{A.1.4})$$

Substituting A.1.2 in our model equation for $M(t_i)$ and multiplying out the square of the difference between the data and the model function gives us

$$L(\mathbf{B}, \Theta, \sigma) = (2\pi\sigma^2)^{-(N/2)} \exp \left\{ -\frac{NQ(\Theta)}{2\sigma^2} \right\}, \quad (\text{A.1.5})$$

where we define

$$Q(\Theta) \equiv \overline{d^2} - \frac{2}{N} \sum_{j=1}^m \sum_{i=1}^N B_j d_i G_j(t_i, \Theta) + \frac{1}{N} \sum_{j=1}^m \sum_{k=1}^m g_{jk}(\Theta) B_j B_k \quad (\text{A.1.6})$$

and

$$g_{jk}(\Theta) \equiv \sum_{i=1}^N G_j(t_i, \Theta) G_k(t_i, \Theta), \quad (\text{A.1.7})$$

where j and k are the row and column of the matrix respectively. The mean-square of the data $\overline{d^2}$ is calculated as follows:

$$\overline{d^2} = \frac{1}{N} \sum_{i=1}^N d_i^2.$$

We have yet to marginalise the nuisance parameters – the model amplitudes \mathbf{B} . In the following section we will simplify Equation A.1.5 by making all terms which are not on the diagonal zero in the matrix $g_{jk}(\Theta)$.

A.1.2 The Orthonormal Model Equations

The following calculations are still from [8]. Supposing redundant model functions are removed so that $g_{jk}(\Theta)$ is positive-definite and of rank m (the number of model functions), let $e_{kj}(\Theta)$ be the j th component of the k th normalised eigenvector of $g_{jk}(\Theta)$ and $\lambda_j(\Theta)$ the corresponding eigenvalue:

$$\sum_{k=1}^m g_{jk}(\Theta) e_{lk}(\Theta) = \lambda_l(\Theta) e_{lj}(\Theta).$$

Note that g_{jk} is a function of the model parameters of interest and thus λ and e are too. We can define new *orthonormal* (orthogonal and normalised) model functions:

$$H_j(t_i, \Theta) = \frac{1}{\sqrt{\lambda_j(\Theta)}} \sum_{k=1}^m e_{jk}(\Theta) G_k(t_i, \Theta) \quad (\text{A.1.8})$$

and have a new guaranteed orthonormal $g_{jk}(\Theta)$ matrix redefined as:

$$\varphi_{jk}(\Theta) = \sum_{i=1}^N H_j(t_i, \Theta) H_k(t_i, \Theta).$$

The model Equation A.1.2 can now be rewritten in terms of the orthonormal functions $H_j(\Theta)$:

$$M_{orth}(t_i) = \sum_{j=1}^m A_j H_j(t_i, \Theta). \quad (\text{A.1.9})$$

Where the new amplitudes \mathbf{A} are linearly related to the old amplitudes \mathbf{B} for each set of values of the model parameters:

$$B_k = \sum_{j=1}^m \frac{A_j e_{jk}(\Theta)}{\sqrt{\lambda_j(\Theta)}} \quad \text{and} \quad A_k = \sqrt{\lambda_k(\Theta)} \sum_{j=1}^m B_j e_{jk}(\Theta). \quad (\text{A.1.10})$$

A.1.3 Elimination of Nuisance Parameters

We still follow the calculations from [8], except where indicated. We now have our orthonormal model functions and are able to eliminate the parameters that we do not wish to estimate. Typically this would be the now orthonormal amplitude parameters, \mathbf{A} , as we wish to estimate parameters such as the frequency, chirp, decay, etc. We will marginalise the amplitude parameters for the general case with an arbitrary number of model functions.

Substituting Equation A.1.9, our orthonormalised model functions, into Equation A.1.4, our likelihood function, we have

$$L(\mathbf{B}, \Theta, \sigma) = (2\pi\sigma^2)^{-(N/2)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^N [d_i - M_{orth}(t_i)]^2 \right\},$$

and since we can rewrite the equation in terms of the new amplitudes \mathbf{A} , we have

$$L(\mathbf{A}, \Theta, \sigma) = (2\pi\sigma^2)^{-(N/2)} \times \exp \left\{ -\frac{N}{2\sigma^2} \left[\bar{d}^2 - \frac{2}{N} \sum_{j=1}^m A_j h_j(\Theta) + \frac{1}{N} \sum_{j=1}^m A_j^2 \right] \right\}, \quad (\text{A.1.11})$$

with $h_j(\Theta)$ defined as:

$$h_j(\Theta) \equiv \sum_{i=1}^N d_i H_j(t_i, \Theta), \quad (1 \leq j \leq m). \quad (\text{A.1.12})$$

Here $h_j(\Theta)$ is the projection of the data onto the orthonormal model function $H_j(t_i, \Theta)$. This will help us calculate the sufficient statistic and summarise all of the information in the data for inferences about the nonlinear parameters [21]. This is the function that is especially important with regard to finding the most likely set of parameters given the data.

In order to remove the amplitudes, we need a prior probability density function for their values. We use Gaussian probability density functions, as the amplitudes are location parameters. It is assumed the data determine the parameters much better than the prior, so the variance on the prior for the amplitudes is assumed to be very large. Therefore, over the region where the direct probability of the data is at its peak, the prior will look like a uniform prior [21]. In the limit of the Gaussian probability density function's variance to infinity, the prior will be an improper uniform prior. This unbound uniform prior is strictly speaking not a probability distribution at all. This is not a problem for parameter estimation, as the infinities introduced always cancel out when the distribution is normalised. This is not true for model selection, and improper priors cannot be used. We will use properly bounded priors in Section A.3. The amplitudes can be marginalised analytically using multivariate Gaussian integrals (see Appendix B.1). For now we ignore the normalising constants. Approximating the prior for the amplitudes and integrating them out, we obtain [21]

$$L(\Theta, \sigma) \propto \sigma^{-N+m} \times \exp \left\{ -\frac{N\bar{d}^2 - m\bar{h}^2(\Theta)}{2\sigma^2} \right\}, \quad (\text{A.1.13})$$

where we define

$$\bar{h}^2(\Theta) \equiv \frac{1}{m} \sum_{j=1}^m [h_j(\Theta)]^2 \quad (\text{A.1.14})$$

as the mean-square of the observed projections.

If σ is known, then the joint probability of the Θ parameters conditional on the data and the noise is:

$$p(\Theta|D, I) \propto \exp \left\{ \frac{m\bar{h}^2(\Theta)}{2\sigma^2} \right\}. \quad (\text{A.1.15})$$

If σ is not known, it is removed as a nuisance parameter. It is a scale parameter and restricted to positive values. The uninformative prior probability for scale parameters is the Jeffrey's prior. Marginalising σ given the Jeffrey's prior ($1/\sigma$) gives us

$$p(\Theta|D, I) \propto \left[1 - \frac{m\bar{h}^2(\Theta)}{Nd^2} \right]^{\frac{m-N}{2}}. \quad (\text{A.1.16})$$

The above function only becomes singular if one has a perfect knowledge of the Θ parameters. Equations A.1.15 and A.1.16 will allow us to calculate the frequency and chirp parameters of an arbitrary number of sinusoids.

A.2 BSA Relation to Periodogram

In this section we will demonstrate the relation between the periodogram and BSA. We wish to calculate the sufficient statistic for a model to an arbitrary resolution using Equation A.1.14:

$$\bar{h}^2(\Theta) \equiv \frac{1}{m} \sum_{j=1}^m [h_j(\Theta)]^2.$$

As per our model function Equation A.1.2, we can calculate the sufficient statistic for a signal model. For a single sinusoid model we would construct the model:

$$M_{1\sin}(t_i) = B_1 \sin(2\pi\omega) + B_2 \cos(2\pi\omega)$$

where $m = 2$ is the number of model functions.

The first step would be calculating the g_{ij} matrix using Equation A.1.7 to create our orthonormal model functions:

$$g_{ij}(\omega) = \begin{bmatrix} \sum_{k=T}^{k=-T} \cos^2(2\pi\omega k) & \sum_{k=T}^{k=-T} \cos(2\pi\omega k) \sin(2\pi\omega k) \\ \sum_{k=T}^{k=-T} \cos(2\pi\omega k) \sin(2\pi\omega k) & \sum_{k=T}^{k=-T} \sin^2(2\pi\omega k) \end{bmatrix},$$

where i and j are the row and column of the matrix respectively. For uniform time sampling the off-diagonal terms are zero and the diagonal can be approximated as

$$g_{ij}(\omega) \approx \begin{bmatrix} \frac{N}{2} + \frac{\sin(2\pi N\omega)}{2\sin(2\pi\omega)} & 0 \\ 0 & \frac{N}{2} - \frac{\sin(2\pi N\omega)}{2\sin(2\pi\omega)} \end{bmatrix}.$$

Calculating our orthonormal model functions using Equation A.1.8, we have

$$H_1(t, \omega) \approx \frac{\cos(2\pi\omega t)}{\sqrt{\frac{N}{2} + \frac{\sin(2\pi N\omega)}{2\sin(2\pi\omega)}}},$$

$$H_2(t, \omega) \approx \frac{\sin(2\pi\omega t)}{\sqrt{\frac{N}{2} - \frac{\sin(2\pi N\omega)}{2\sin(2\pi\omega)}}}.$$

When $N \gg 1$ then our model functions are approximately

$$H_1(t, \omega) \approx \frac{\sqrt{2} \cos(2\pi\omega t)}{\sqrt{N}},$$

$$H_2(t, \omega) \approx \frac{\sqrt{2} \sin(2\pi\omega t)}{\sqrt{N}}.$$

The projection of the data on the orthonormal model functions are, from Equation A.1.12:

$$h_1(\omega) \approx \sum_{i=1}^N d_i \frac{\sqrt{2} \cos(2\pi\omega t_i)}{\sqrt{N}},$$

$$h_2(\omega) \approx \sum_{i=1}^N d_i \frac{\sqrt{2} \sin(2\pi\omega t_i)}{\sqrt{N}}.$$

Note the similarities to the projection of the data for the periodogram that can be calculated as:

$$C(\omega) = \frac{1}{N} [R(\omega)^2 + I(\omega)^2] = \frac{1}{N} \left| \sum_{j=1}^N d_j e^{2\pi i \omega t_j} \right|^2 = \frac{1}{N} |\text{DFT}|^2$$

and where the real and imaginary projections are:

$$R(\omega) = \sum_{j=1}^N d_j \cos(2\pi\omega t_j)$$

and

$$I(\omega) = \sum_{j=1}^N d_j \sin(2\pi\omega t_j).$$

From Equation A.1.14 we then calculate

$$\overline{h^2}(\omega) \approx \frac{1}{2} \left[\sum_{i=1}^N d_i \frac{\sqrt{2} \cos(2\pi\omega t_i)}{\sqrt{N}} \right]^2 + \frac{1}{2} \left[\sum_{i=1}^N d_i \frac{\sqrt{2} \sin(2\pi\omega t_i)}{\sqrt{N}} \right]^2 = C(\omega)$$

and arrive at the periodogram $C(\omega)$ as the sufficient statistic for a single sinusoid in white noise.

One need not approximate g_{ij} and the orthonormal functions H_j , and can calculate them exactly. For the case of the single sinusoid, the error caused by the approximations is negligible and becomes smaller for larger values of N .

A.3 BSA Model Selection

If one has a set of possible BSA models, how does one determine which one is most likely to be the correct model? Given the data and the available prior information, how strongly is that model supported relative to the alternatives? The questions asked for model selection are different to that of parameter estimation but are solved using the same methods. This section is based on [22], except where otherwise indicated.

The questions can be solved by calculating the joint probability density function of the data and the parameters, and then integrating out the parameters. For BSA this cannot be done analytically. Using Bayes' theorem we have the likelihood function for our model under consideration:

$$p(D|M_j, I) = \int d\Theta p(\Theta|M_j, I) \times \int d\mathbf{A} p(\mathbf{A}|\Theta, M_j, I)p(D|\mathbf{A}, \Theta, M_j, I), \quad (\text{A.3.1})$$

where M_j is the j th model function under consideration with the parameters of interest in the previous section Θ and the model function amplitudes \mathbf{A} being marginalised out.

Unlike our method of parameter estimation from the previous section, all the numerical factors must be kept as they do not necessarily cancel during model selection. One can consider all the parameters as nuisance parameters to be marginalised. Since we are ignorant as to which model is the correct one, we use a uniform prior for the models. The following equation for the model probability has the constants cancelling out:

$$p(M_j|\sigma, D, I) = \frac{p(D|\sigma, M_j, I)p(M_j|I)}{\sum_{k=1}^s p(D|\sigma, M_k, I)p(M_k|I)} = \frac{p(D|\sigma, M_j, I)}{\sum_{k=1}^s p(D|\sigma, M_k, I)},$$

where s is the number of hypotheses under review and the prior for our models is $p(M_j|I) = 1/s$.

Our parameter likelihood from Equation A.1.11 is multiplied with the prior probability density functions over the parameters we wish to marginalise, and then we integrate them out to determine the model likelihood function in Equation A.3.1 as follows:

$$p(D|M_j, \sigma, I) = (2\pi\sigma^2)^{-(N/2)} \int d\Theta d\mathbf{A} p(\Theta, \mathbf{A}|M_j, I) \times \exp \left\{ -\frac{N}{2\sigma^2} \left[\bar{d}^2 - \frac{2}{N} \sum_{j=1}^m A_j h_j(\Theta) + \frac{1}{N} \sum_{j=1}^m A_j^2 \right] \right\}. \quad (\text{A.3.2})$$

Because any information we have about our model parameters Θ will not influence our knowledge about the amplitudes, \mathbf{A} , we can write our prior knowledge about these independent parameters as:

$$p(\mathbf{A}, \Theta | M_j, I) = p(\mathbf{A}, | M_j, I) p(\Theta | M_j, I).$$

We have vague information about the amplitudes \mathbf{A} . They are location parameters and Gaussian probability density functions are therefore appropriate. As Gaussian priors do not restrict random variables to be either negative or positive and we have little information about the mean, we select a zero mean Gaussian for each amplitude. Assuming that the different amplitudes are independent of each other, the probability of all the amplitudes is a product of Gaussian probability density functions and thus

$$p(\mathbf{A} | \gamma, M_j, I) = (2\pi\gamma^2)^{\frac{m}{2}} \exp \left\{ - \sum_{k=1}^m \frac{A_k^2}{2\gamma^2} \right\}, \quad (\text{A.3.3})$$

is our prior. This introduces a new parameter γ , which is the standard deviation of our prior Gaussian densities and is a hyperparameter. We are assuming little prior information so that we have $\sigma \ll \gamma$ and the same precision for all the amplitudes.

Substituting our prior A.3.3 into A.3.2 we have

$$\begin{aligned} p(D | M_j, \sigma, \gamma, I) &= (2\pi\sigma^2)^{-(N/2)} (2\pi\gamma^2)^{\frac{m}{2}} \int d\Theta p(\Theta | M_j, I) \\ &\times \int_{-\infty}^{\infty} d\mathbf{A} \exp \left\{ - \sum_{k=1}^m \frac{A_k^2}{2\gamma^2} \right\} \\ &\times \exp \left\{ - \frac{N}{2\sigma^2} \left[\bar{d}^2 - \frac{2}{N} \sum_{j=1}^m A_j h_j(\Theta) + \frac{1}{N} \sum_{j=1}^m A_j^2 \right] \right\} \end{aligned} \quad (\text{A.3.4})$$

as the likelihood of our model parameters and model amplitudes given the model and the hyperparameter γ . Integrating out the nuisance amplitudes which are Gaussian integrals (see Appendix B.1), we obtain:

$$\begin{aligned} p(D | M_j, \sigma, \gamma, I) &= (2\pi\sigma^2)^{-(N/2)} \left[\frac{\sigma^2 + \gamma^2}{\sigma^2} \right]^{-(m/2)} \\ &\times \int d\Theta p(\Theta | M_j, I) \exp \left\{ - \frac{N\bar{d}^2}{2\sigma^2} + \frac{m\bar{h}^2\gamma^2}{2\sigma(\gamma^2 + \sigma^2)} \right\} \end{aligned} \quad (\text{A.3.5})$$

Since we assume $\sigma \ll \gamma$, we can simplify the equations by means of the following approximations:

$$\frac{\sigma^2 + \gamma^2}{\sigma^2} = 1 + \frac{\gamma^2}{\sigma^2} \approx \frac{\gamma^2}{\sigma^2}$$

and

$$\frac{\gamma^2}{\sigma^2(\gamma^2 + \sigma^2)} = \frac{(\sigma^2 + \gamma^2)}{\sigma^2(\sigma^2 + \gamma^2)} - \frac{\sigma^2}{\sigma^2(\sigma^2 + \gamma^2)} = \frac{1}{\sigma^2} - \frac{1}{\sigma^2 + \gamma^2} \approx \frac{1}{\sigma^2} - \frac{1}{\gamma^2}$$

thus giving us

$$p(D|M_j, \sigma, \gamma, I) \approx (2\pi)^{-(N/2)} \sigma^{m-N} \gamma^{-m} \times \int d\Theta p(\Theta|M_j, I) \exp \left\{ -\frac{[N\bar{d}^2 - m\bar{h}^2]}{2\sigma^2} - \frac{m\bar{h}^2}{2\gamma^2} \right\} \quad (\text{A.3.6})$$

We now only need to marginalise the parameters Θ which we cannot do analytically. To avoid numeric integration, we can perform a Taylor expansion on the peak of \bar{h}^2 . The values that maximise the posterior probability for the model M_j are designated as $\hat{\Theta} \equiv \{\hat{\Theta}_0, \dots, \hat{\Theta}_{r-1}\}$ where r is the total number of parameters. One then obtains:

$$\bar{h}^2 \approx \bar{h}^2 \Big|_{\hat{\Theta}} - \sum_{k=1}^r \sum_{l=1}^r \frac{b_{kl}}{m} (\hat{\Theta}_k - \Theta_k) (\hat{\Theta}_l - \Theta_l), \quad (\text{A.3.7})$$

where defining

$$b_{kl} \equiv - \frac{\partial^2 m\bar{h}^2}{2\partial\Theta_k \partial\Theta_l} \Big|_{\hat{\Theta}}. \quad (\text{A.3.8})$$

With this approximation, the global likelihood of the data Equation A.3.6 may be written as

$$p(D|M_j, \gamma, \sigma, I) \approx (2\pi)^{-(N/2)} \sigma^{m-N} \gamma^{-m} \exp \left\{ -\frac{N\bar{d}^2}{2\sigma^2} + \frac{mu\bar{h}^2}{2} \right\} \Big|_{\hat{\Theta}} \times \int d\Theta p(\Theta|M_j, I) \exp \left\{ -u \sum_{k,l=1}^r \frac{b_{jk}(\hat{\Theta}_k - \Theta_k)(\hat{\Theta}_l - \Theta_l)}{2} \right\}, \quad (\text{A.3.9})$$

where

$$u \equiv \frac{1}{\sigma^2} - \frac{1}{\gamma^2}.$$

To be able to perform the integrals, the prior probability $p(\Theta|M_j, I)$ must be assigned. These parameters are usually location parameters where the appropriate prior would once again be a Gaussian distribution. This is especially true if the parameters can be either positive or negative and they have a finite mean-square value.

In our our application we know the possible ranges for our parameters, for example the frequencies: $0 < \omega \leq 0.5$. It is only when the information in the prior is comparable to the data that the prior probability will make much

difference. The prior $p(\Theta|M_j, I)$ is a constant and can be placed outside the integral. The value of the uniform prior depends on which model is under consideration, and the bounds of its parameters:

$$p(D|M_j, \gamma, \sigma, I) \approx (2\pi)^{-(N/2)} \sigma^{m-N} \gamma^{-m} p(\Theta|M_j, I) \times \int d\Theta \exp \left\{ -\frac{m\bar{h}^2}{2\gamma^2} - \left[\frac{N\bar{d}^2 - m\bar{h}^2}{2\sigma^2} \right] \right\} \Big|_{\hat{\Theta}} \quad (\text{A.3.10})$$

Where v_j is the j th eigenvalue of the matrix b_{jk} . We can now do the integral over the non-linear model parameters to obtain:

$$p(D|M_j, \gamma, \sigma, I) \approx (2\pi)^{-(N/2)} \sigma^{m-N} \gamma^{-m} v_1^{-0.5} \dots v_r^{-0.5} \times \exp \left\{ -\frac{m\bar{h}^2}{2\gamma^2} - \left[\frac{N\bar{d}^2 - m\bar{h}^2}{2\sigma^2} \right] \right\} \Big|_{\hat{\Theta}} \quad (\text{A.3.11})$$

This is analogous to Laplace's Method [25] where the function is approximated by a Gaussian (see Appendix B.2). To get rid of the variances, σ and γ , the integrals are approximated as follows:

$$\int_{L_\gamma}^{H_\gamma} d\gamma \frac{\gamma^{-m} \exp(-m\bar{h}^2/s\gamma^2)}{\gamma \log(H_\gamma/L_\gamma)} \approx \frac{\Gamma(m/2)}{2 \log(H_\gamma/L_\gamma)} \left[m\bar{h}^2/2 \right]^{-m/2}$$

and

$$\int_{L_\sigma}^{H_\sigma} d\sigma \frac{\sigma^{m-N} \exp(-[N\bar{d}^2 - m\bar{h}^2]/2\sigma^2)}{\sigma \log(H_\sigma/L_\sigma)} \approx \frac{\Gamma\left(\frac{N-m}{2}\right)}{2 \log(H_\sigma/L_\sigma)} \left[\frac{N\bar{d}^2 - m\bar{h}^2}{2} \right]^{\frac{m-N}{2}}.$$

Cancelling the parts of model comparison that are not parameter dependent, we then have a model likelihood that is comparable between models:

$$p(D|M_j, I) \approx \Gamma\left(\frac{m}{2}\right) \left(\frac{N-m}{2}\right) v_1^{-0.5} \dots v_r^{-0.5} \times \left[\frac{m\bar{h}^2}{2} \right] \left[\frac{N\bar{d}^2 - m\bar{h}^2}{2} \right]^{\frac{m-N}{2}} \Big|_{\hat{\Theta}}. \quad (\text{A.3.12})$$

For models with multiple maxima due to symmetry, we can multiply the model likelihood by the number of peaks. For example, the two-peaked example in Figure 3.1 would be multiplied by 2, as otherwise we would only approximate the integral of one of the two identical peaks.

A.4 BSA Estimated Noise Variance $\langle \sigma^2 \rangle$

The following formulation of estimating the noise variance comes from [8]. Suppose we estimated our signal model parameters Θ , we are now interested

in the noise standard deviation σ . Bretthorst derives the variance directly. Bretthorst calculates the posterior moment of $\langle \sigma^s \rangle$ with any power of s . Using A.1.13 (The general likelihood function before integrating σ out) and a Jeffrey's prior $\frac{1}{\sigma}$ then integrating to get the expected value we have:

$$E(\sigma^s | \Theta, D, I) = \langle \sigma^s \rangle = \frac{\int_0^\infty d\sigma \sigma^{s-1} L(\sigma | \Theta, D, I)}{\int_0^\infty d\sigma \sigma^{-1} L(\sigma | \Theta, D, I)}$$

we obtain

$$\langle \sigma^s \rangle(\Theta) = \Gamma\left(\frac{N-m-s}{2}\right) \Gamma\left(\frac{N-m}{2}\right)^{-1} \left[\frac{N\bar{d}^2 - m\bar{h}^2}{2} \right]^{s/2}.$$

So for $s = 2$ one would have

$$\langle \sigma^2 \rangle(\Theta) = \frac{1}{N-m-2} \left[\frac{N\bar{d}^2 - m\bar{h}^2}{2} \right].$$

This estimation is dependant on the parameters, Θ . This matches our intuition as the degree to how well our model would explain the data is different depending on values for the parameters. The lowest amount of noise is where the model parameters' probability density function is at its peak and a better fit therefore needs to attribute less of the measured data to noise.

Appendix B

Miscellaneous Integrals and Algorithms

B.1 Gaussian Integrals

The integral of an arbitrary Gaussian function is:

$$\int_{-\infty}^{\infty} dx a e^{-(x-b)^2/2c^2} = \sqrt{2a} |c| \sqrt{\pi}.$$

B.2 Laplace's Method

From [25] we briefly discuss Laplace's Method. We wish to approximate a function as a Gaussian probability density function. For Laplace's Method we have an unnormalised probability density $P^*(x)$, whose normalising constant is

$$Z_P \equiv \int dx P^*(x)$$

and which has a peak at a point x_0 . We Taylor-expand the logarithm of $P^*(x)$ around the peak and only keep the first two terms

$$\log(P^*(x)) \simeq \log(P^*(x_0)) - \frac{1}{2}(x - x_0)^T \mathbf{A}(x - x_0),$$

where \mathbf{A} is the second derivative (the Hessian) of

$$-\log(P^*(x))|_{x=x_0}.$$

The first-order term of the Taylor-expansion is zero as we take the derivative at the maximum of the function. The normalizing constant can be approximated by

$$Z_P \simeq P^*(x_0) \sqrt{\frac{(2\pi)^{\dim}}{\det \mathbf{A}}}$$

and mean and covariance matrix of the Gaussian approximation is x_0 and \mathbf{A} respectively.

B.3 Nelder-Mead Optimisation

The Nelder-Mead algorithm [26] is a method used to find the minimum or maximum of an objective function. For n dimensions the algorithm maintains $n + 1$ test points arranged as a *simplex*, which is a special polytope with $n + 1$ vertices. It extrapolates the objective function's behavior at each test point in order to find a new test point which will replace one of the old test points. The Nelder-Mead algorithm typically requires one or two function evaluations at each step. In this thesis we used the *minimize* function from the SciPy [45] library with the Nelder-Mead method.

Appendix C

Simulated Signal BSA

C.1 Introduction

In this appendix we have examples of BSA applied to simulated sinusoids with additive white noise. We use the noiseless model from Equation 3.5.7 for parameter estimation and Equation 3.7.2 to approximate model selection. In the results we can see that typically the correct model is selected.

C.2 Single Stationary Sinusoid

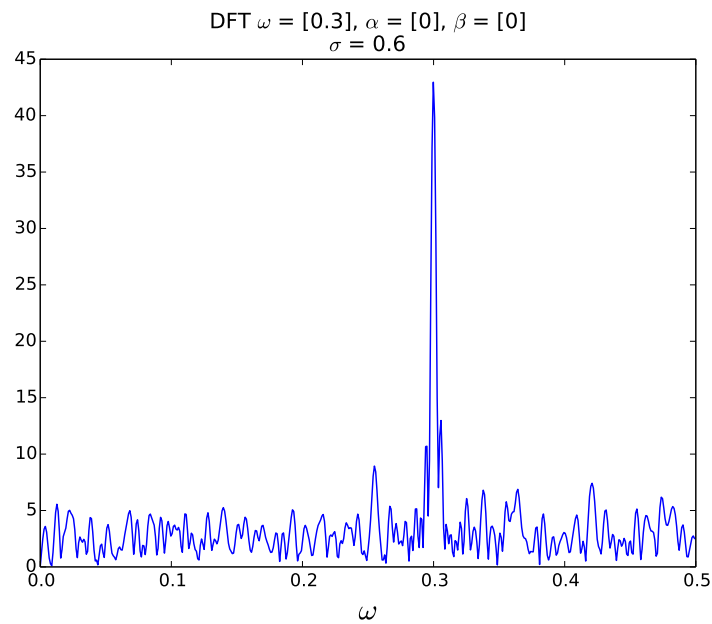


Figure C.1: *DFT of a simulated single stationary sinusoid. $\omega_0 = 0.3, \sigma = 0.6$*

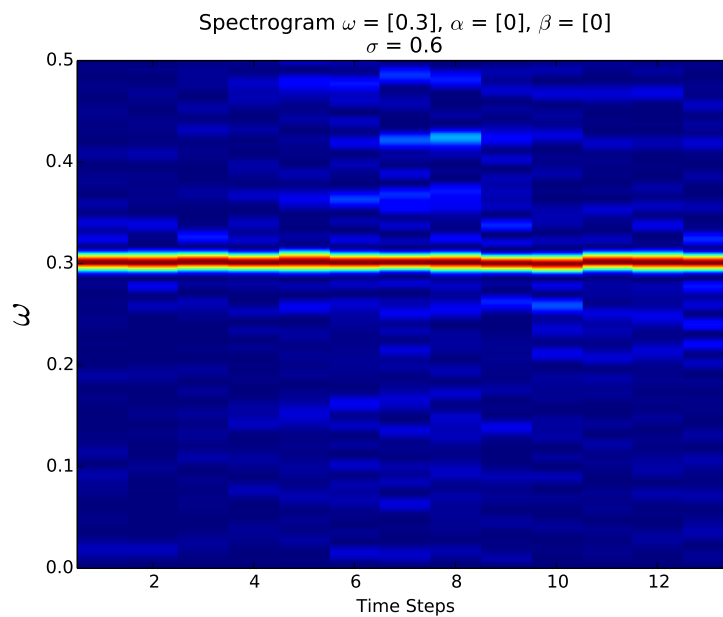


Figure C.2: *Spectrogram of a simulated single stationary sinusoid. $\omega_0 = 0.3, \sigma = 0.6$*

Table C.1: *Estimated parameters and model selection of a simulated single stationary sinusoid. The most likely model is indicated in blue.*

	True Model	1 Stationary	1 Chirp	1 Qua Chirp
Probability	-	9.93598e-01	2.02742e-05	3.18200e-12
ω_0	3.00000e-01	2.99955e-01	2.99947e-01	2.99956e-01
α_0	0.00000e+00	-	-1.21168e-05	-1.21431e-05
β_0	0.00000e+00	-	-	-5.12003e-09
ω_1	0.00000e+00	-	-	-
α_1	0.00000e+00	-	-	-
β_1	0.00000e+00	-	-	-
ω_2	0.00000e+00	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	2 Stationary	2 Chirp	2 Qua Chirp
Probability	-	6.34758e-03	1.60834e-11	5.84138e-24
ω_0	3.00000e-01	2.99956e-01	2.54592e-01	2.55627e-01
α_0	0.00000e+00	-	9.56709e-05	3.39535e-05
β_0	0.00000e+00	-	-	-7.64918e-03
ω_1	0.00000e+00	2.54931e-01	2.99952e-01	2.99930e-01
α_1	0.00000e+00	-	-1.32510e-05	-8.75546e-06
β_1	0.00000e+00	-	-	1.85851e-08
ω_2	0.00000e+00	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	3 Stationary	3 Chirp	3 Qua Chirp
Probability	-	3.36834e-05	6.24590e-17	4.96921e-33
ω_0	3.00000e-01	2.99962e-01	1.90820e-01	2.99941e-01
α_0	0.00000e+00	-	-4.52856e-02	-7.07495e-06
β_0	0.00000e+00	-	-	-3.68935e-09
ω_1	0.00000e+00	2.54924e-01	2.35950e-01	3.83395e-01
α_1	0.00000e+00	-	-1.63652e-03	2.52114e-03
β_1	0.00000e+00	-	-	-8.64191e-04
ω_2	0.00000e+00	4.20401e-01	2.99949e-01	2.21199e-01
α_2	0.00000e+00	-	-1.26409e-05	-2.07116e-03
β_2	0.00000e+00	-	-	-3.88624e-03

C.3 Two Stationary Sinusoids

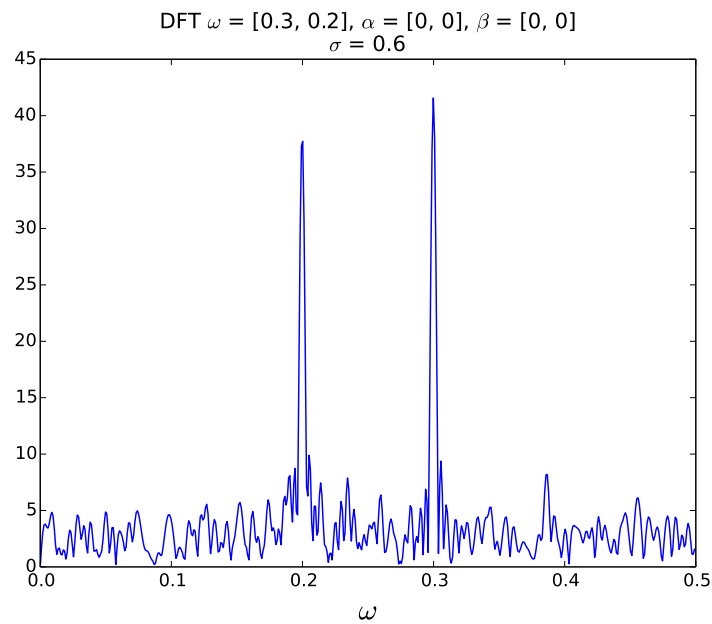


Figure C.3: *DFT of two simulated stationary sinusoids. $\omega_0 = 0.3, \omega_1 = 0.2, \sigma = 0.6$*

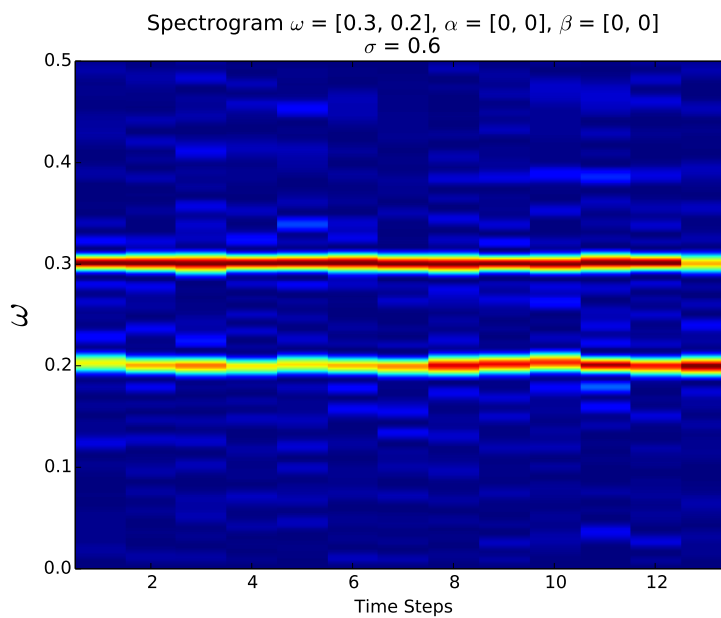


Figure C.4: *Spectrogram of two simulated stationary sinusoids. $\omega_0 = 0.3, \omega_1 = 0.2, \sigma = 0.6$*

Table C.2: *Estimated parameters and model selection of two simulated stationary sinusoids. The most likely model is indicated in blue.*

	True Model	1 Stationary	1 Chirp	1 Qua Chirp
Probability	-	2.62487e-40	4.44905e-45	1.20737e-51
ω_0	3.00000e-01	2.99867e-01	2.99866e-01	2.99714e-01
α_0	0.00000e+00	-	-4.37176e-06	-4.01071e-06
β_0	0.00000e+00	-	-	9.92394e-08
ω_1	2.00000e-01	-	-	-
α_1	0.00000e+00	-	-	-
β_1	0.00000e+00	-	-	-
ω_2	0.00000e+00	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	2 Stationary	2 Chirp	2 Qua Chirp
Probability	-	9.91538e-01	1.44053e-10	4.92731e-24
ω_0	3.00000e-01	2.99853e-01	2.99852e-01	1.99673e-01
α_0	0.00000e+00	-	-1.42964e-06	6.52387e-06
β_0	0.00000e+00	-	-	6.01076e-08
ω_1	2.00000e-01	1.99778e-01	1.99770e-01	2.99767e-01
α_1	0.00000e+00	-	6.60783e-06	-1.39242e-06
β_1	0.00000e+00	-	-	5.39858e-08
ω_2	0.00000e+00	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	3 Stationary	3 Chirp	3 Qua Chirp
Probability	-	8.46164e-03	1.08317e-14	1.53447e-34
ω_0	3.00000e-01	2.99853e-01	1.99714e-01	1.99728e-01
α_0	0.00000e+00	-	2.93097e-06	7.43383e-06
β_0	0.00000e+00	-	-	4.98801e-08
ω_1	2.00000e-01	1.99778e-01	1.58722e-01	2.62885e-01
α_1	0.00000e+00	-	1.24913e-03	1.68332e-05
β_1	0.00000e+00	-	-	1.59978e-04
ω_2	0.00000e+00	3.86466e-01	2.99861e-01	2.99799e-01
α_2	0.00000e+00	-	-1.95381e-06	-4.37508e-07
β_2	0.00000e+00	-	-	5.99437e-08

C.4 Three Stationary Sinusoids

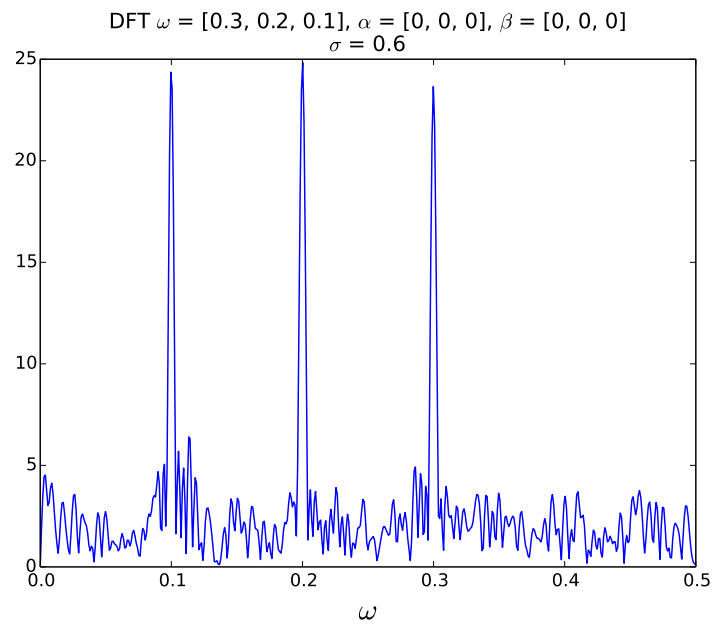


Figure C.5: *DFT of three simulated stationary sinusoids.* $\omega_0 = 0.3, \omega_1 = 0.2, \omega_2, \sigma = 0.6$.

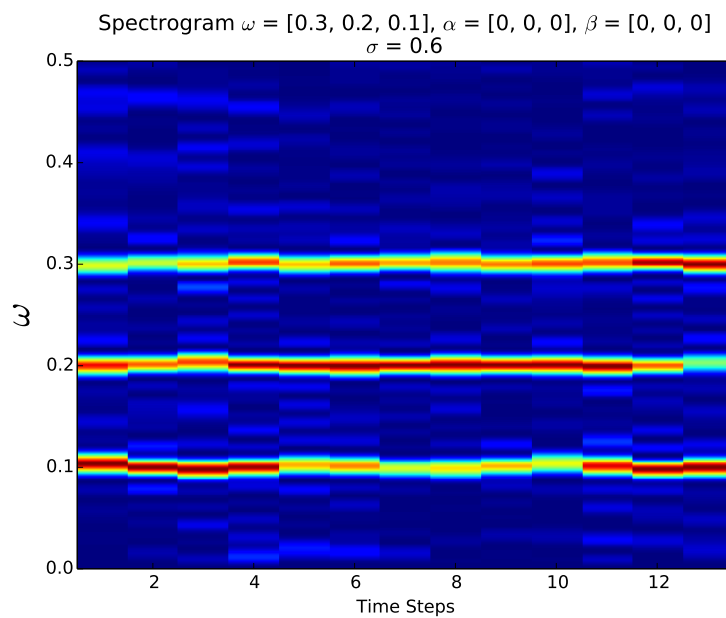


Figure C.6: *Spectrogram of three simulated stationary sinusoids.* $\omega_0 = 0.3, \omega_1 = 0.2, \omega_2, \sigma = 0.6$.

Table C.3: *Estimated parameters and model selection of three simulated stationary sinusoids. The most likely model is indicated in blue.*

	True Model	1 Stationary	1 Chirp	1 Qua Chirp
Probability	-	8.27621e-55	1.89592e-59	8.68450e-66
ω_0	3.00000e-01	2.00018e-01	2.00016e-01	1.99650e-01
α_0	0.00000e+00	-	-1.13041e-06	4.15793e-07
β_0	0.00000e+00	-	-	2.50871e-07
ω_1	2.00000e-01	-	-	-
α_1	0.00000e+00	-	-	-
β_1	0.00000e+00	-	-	-
ω_2	1.00000e-01	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	2 Stationary	2 Chirp	2 Qua Chirp
Probability	-	9.63051e-36	3.40526e-45	4.48875e-58
ω_0	3.00000e-01	9.99796e-02	9.99822e-02	1.99662e-01
α_0	0.00000e+00	-	-1.02823e-05	1.41510e-06
β_0	0.00000e+00	-	-	2.34452e-07
ω_1	2.00000e-01	2.00011e-01	2.00008e-01	9.96498e-02
α_1	0.00000e+00	-	-6.14636e-07	-1.02617e-05
β_1	0.00000e+00	-	-	2.11876e-07
ω_2	1.00000e-01	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	3 Stationary	3 Chirp	3 Qua Chirp
Probability	-	1.00000e+00	3.09933e-15	7.46976e-35
ω_0	3.00000e-01	2.00020e-01	9.99643e-02	1.99700e-01
α_0	0.00000e+00	-	-1.23550e-05	-5.65848e-07
β_0	0.00000e+00	-	-	2.06555e-07
ω_1	2.00000e-01	2.99873e-01	2.99869e-01	3.00091e-01
α_1	0.00000e+00	-	3.99170e-06	4.76510e-06
β_1	0.00000e+00	-	-	-1.30740e-07
ω_2	1.00000e-01	9.99629e-02	2.00015e-01	9.96988e-02
α_2	0.00000e+00	-	-1.60306e-06	-1.20395e-05
β_2	0.00000e+00	-	-	1.65745e-07

C.5 Single Chirped Sinusoid

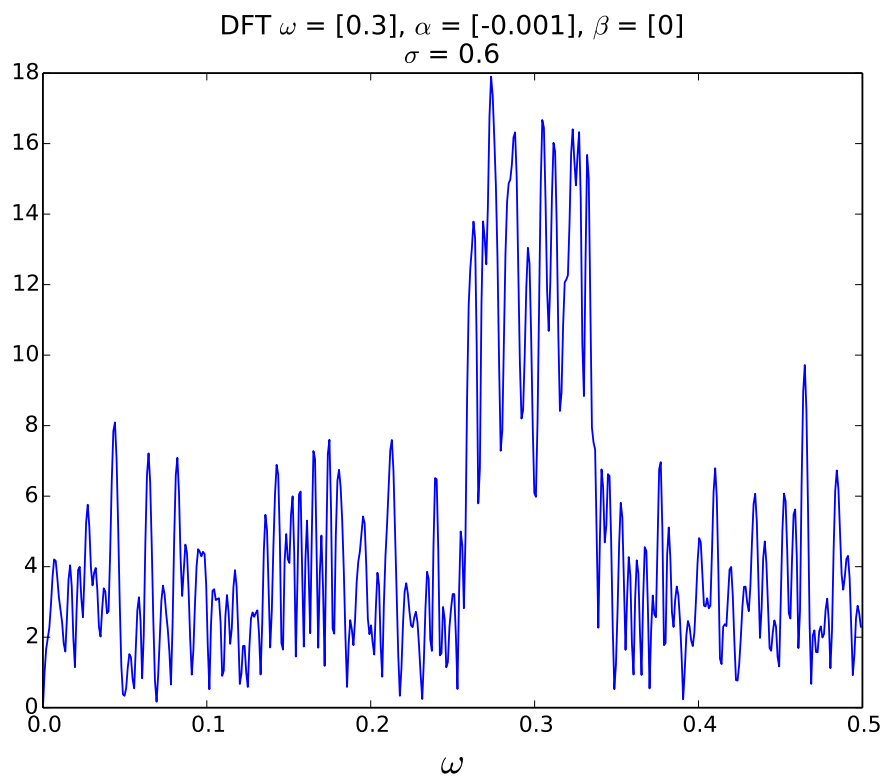


Figure C.7: *DFT of a single simulated chirped sinusoid. $\omega_0 = 0.3$, $\alpha = -0.001$.*

Table C.4: *Estimated parameters and model selection of a single simulated chirped sinusoid. The most likely model is indicated in blue.*

	True Model	1 Stationary	1 Chirp	1 Qua Chirp
Probability	-	3.68023e-47	9.99999e-01	1.50480e-07
ω_0	3.00000e-01	2.73556e-01	2.99889e-01	2.99875e-01
α_0	-1.00000e-03	-	-1.00927e-03	-1.00929e-03
β_0	0.00000e+00	-	-	8.88859e-09
ω_1	0.00000e+00	-	-	-
α_1	0.00000e+00	-	-	-
β_1	0.00000e+00	-	-	-
ω_2	0.00000e+00	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	2 Stationary	2 Chirp	2 Qua Chirp
Probability	-	6.13254e-48	7.66185e-07	9.19785e-19
ω_0	3.00000e-01	2.73949e-01	2.61976e-01	2.86466e-01
α_0	-1.00000e-03	-	-1.07922e-02	1.75017e-04
β_0	0.00000e+00	-	-	2.42885e-04
ω_1	0.00000e+00	3.04866e-01	2.99917e-01	2.99942e-01
α_1	0.00000e+00	-	-1.00908e-03	-1.00127e-03
β_1	0.00000e+00	-	-	-3.42771e-08
ω_2	0.00000e+00	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	3 Stationary	3 Chirp	3 Qua Chirp
Probability	-	1.94936e-44	4.32574e-13	7.36697e-29
ω_0	3.00000e-01	3.24848e-01	2.99901e-01	2.99630e-01
α_0	-1.00000e-03	-	-1.01222e-03	2.80488e-03
β_0	0.00000e+00	-	-	2.61168e-06
ω_1	0.00000e+00	3.24848e-01	2.42505e-01	4.15430e-01
α_1	0.00000e+00	-	-2.13823e-03	-5.88997e-03
β_1	0.00000e+00	-	-	-6.86799e-03
ω_2	0.00000e+00	3.12734e-01	4.14774e-01	2.99883e-01
α_2	0.00000e+00	-	-1.39062e-02	-1.00282e-03
β_2	0.00000e+00	-	-	-1.64266e-08

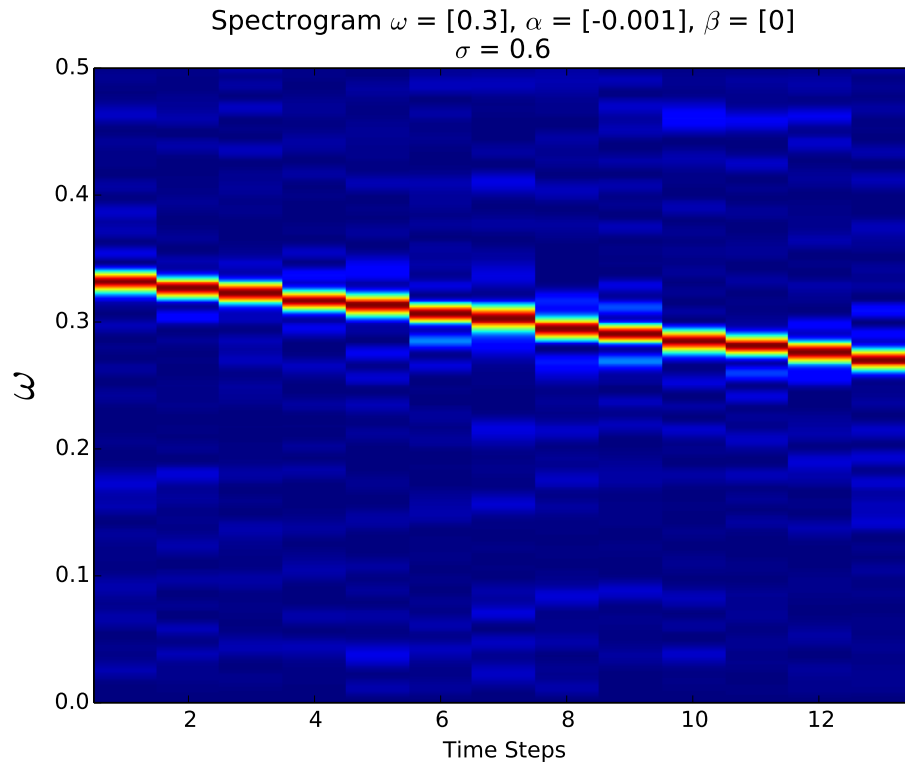


Figure C.8: Spectrogram of a single simulated chirped sinusoid. $\omega_0 = 0.3$, $\alpha = -0.001$.

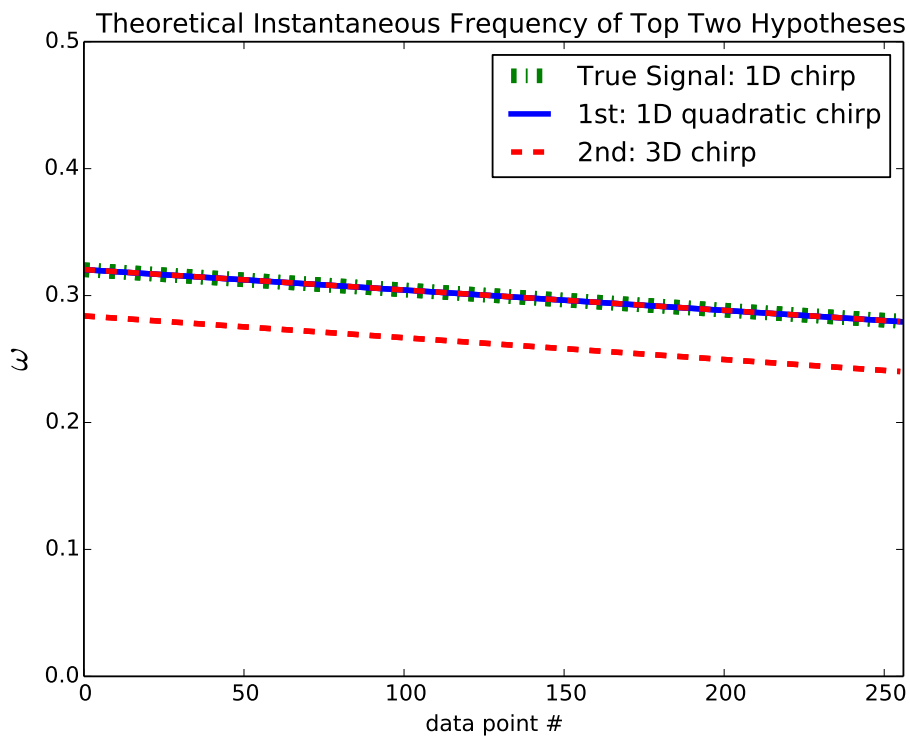


Figure C.9: Theoretical instantaneous frequency of the top two hypotheses.

C.6 Two Chirped Sinusoids

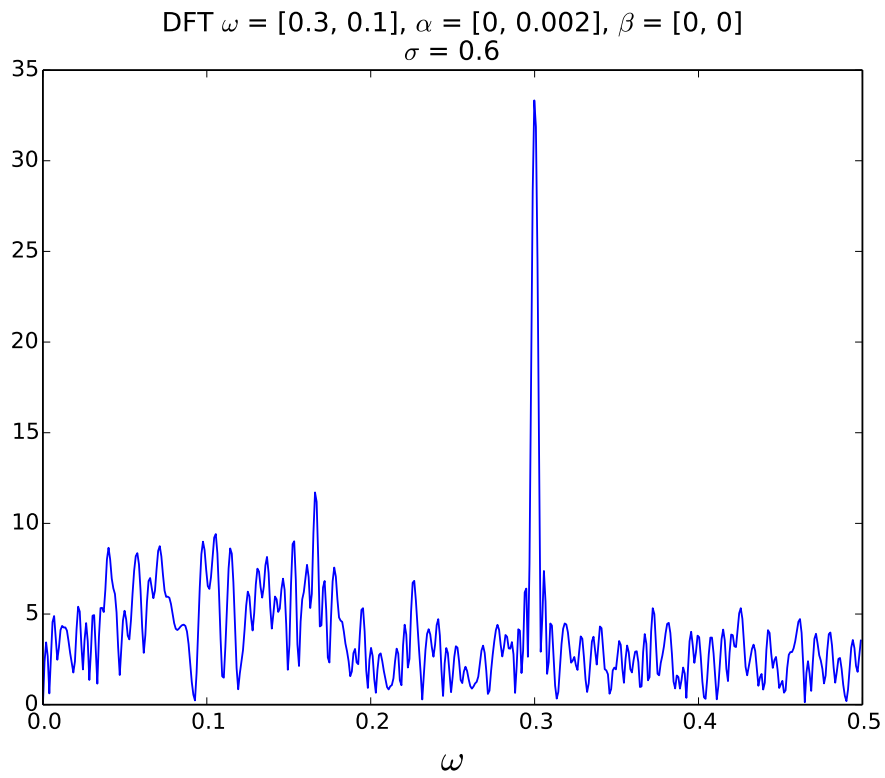


Figure C.10: *DFT of two simulated chirped sinusoids.*

Table C.5: *Estimated parameters and model selection of two simulated chirped sinusoids. The most likely model is indicated in blue.*

	True Model	1 Stationary	1 Chirp	1 Qua Chirp
Probability	-	1.03193e-27	7.81440e-32	5.10969e-39
ω_0	3.00000e-01	3.00082e-01	1.00199e-01	3.00069e-01
α_0	0.00000e+00	-	2.01079e-03	4.56158e-06
β_0	0.00000e+00	-	-	1.45866e-08
ω_1	1.00000e-01	-	-	-
α_1	2.00000e-03	-	-	-
β_1	0.00000e+00	-	-	-
ω_2	0.00000e+00	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	2 Stationary	2 Chirp	2 Qua Chirp
Probability	-	6.71500e-29	9.98912e-01	3.84042e-14
ω_0	3.00000e-01	1.66342e-01	1.00197e-01	1.00195e-01
α_0	0.00000e+00	-	2.00873e-03	2.00867e-03
β_0	0.00000e+00	-	-	9.74745e-09
ω_1	1.00000e-01	3.00083e-01	3.00078e-01	3.00066e-01
α_1	2.00000e-03	-	7.20653e-06	7.18612e-06
β_1	0.00000e+00	-	-	1.39862e-08
ω_2	0.00000e+00	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	3 Stationary	3 Chirp	3 Qua Chirp
Probability	-	3.07792e-27	1.08821e-03	2.34199e-23
ω_0	3.00000e-01	1.64503e-01	1.00253e-01	2.99921e-01
α_0	0.00000e+00	-	2.00387e-03	1.94807e-06
β_0	0.00000e+00	-	-	6.65765e-08
ω_1	1.00000e-01	1.64503e-01	3.00026e-01	1.00345e-01
α_1	2.00000e-03	-	1.79772e-06	2.00419e-03
β_1	0.00000e+00	-	-	-6.36205e-08
ω_2	0.00000e+00	3.00084e-01	5.00678e-02	4.98346e-02
α_2	0.00000e+00	-	8.01603e-03	8.01732e-03
β_2	0.00000e+00	-	-	1.26541e-07

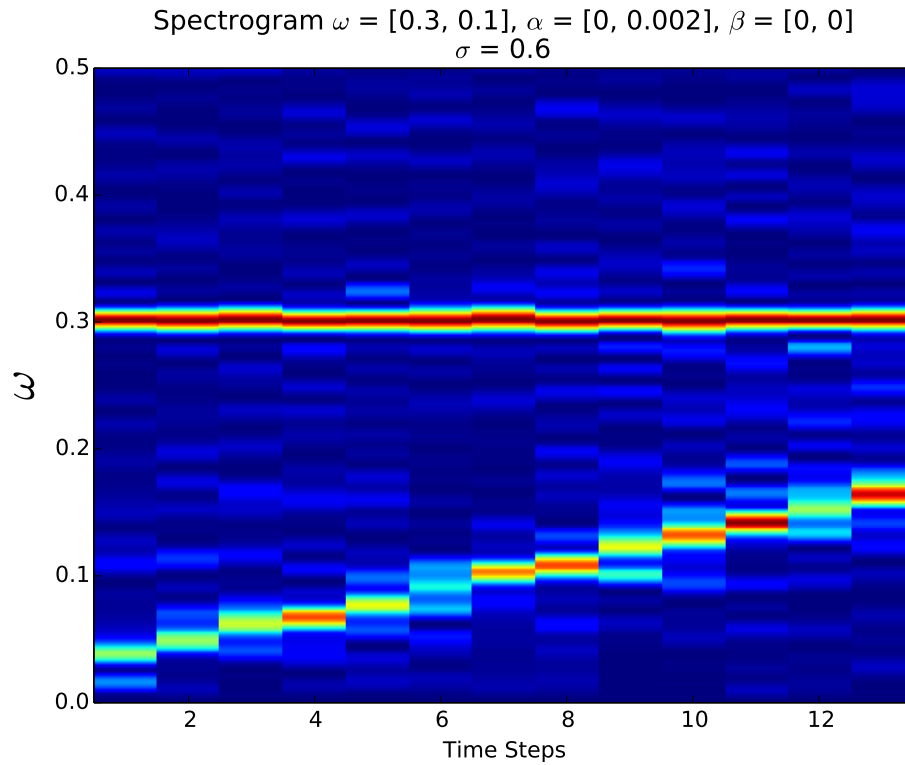


Figure C.11: Spectrogram of two simulated chirped sinusoids.

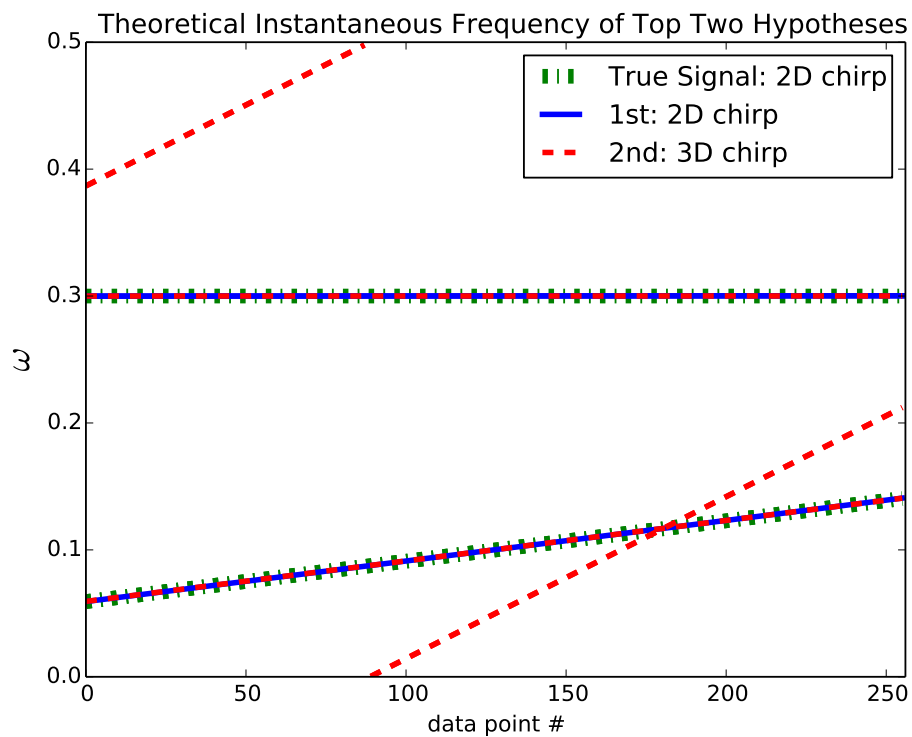


Figure C.12: Theoretical instantaneous frequency of the top two hypotheses.

C.7 Three Chirped Sinusoids

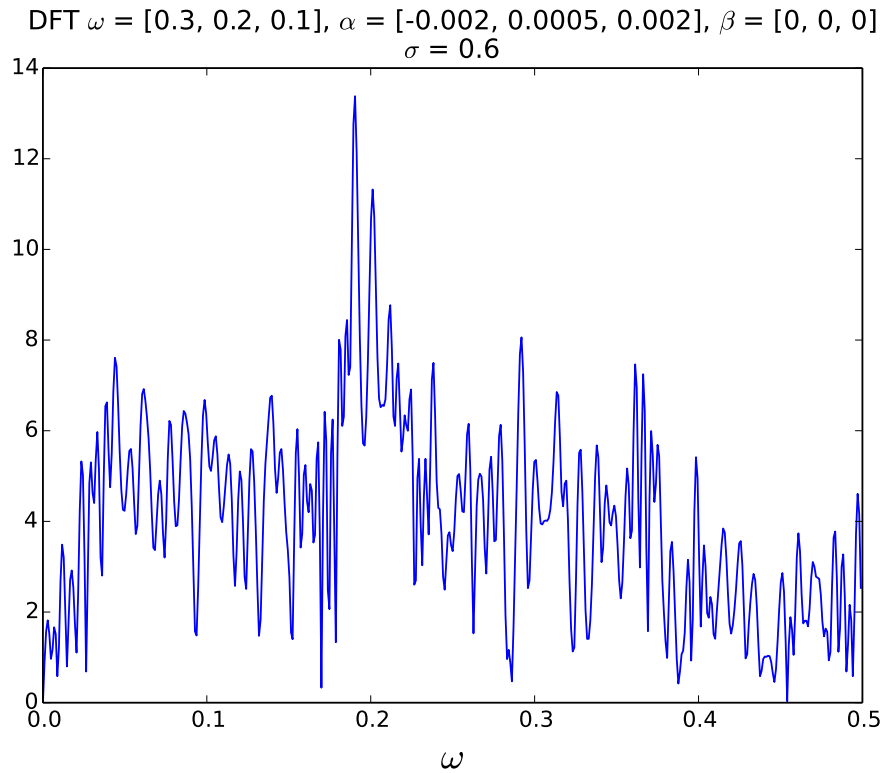


Figure C.13: *DFT of three simulated chirped sinusoids.*

Table C.6: *Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	True Model	1 Stationary	1 Chirp	1 Qua Chirp
Probability	-	3.83334e-54	2.65924e-45	6.02460e-60
ω_0	3.00000e-01	1.90214e-01	2.00297e-01	2.04168e-01
α_0	-2.00000e-03	-	5.32161e-04	1.17856e-03
β_0	0.00000e+00	-	-	3.78020e-06
ω_1	2.00000e-01	-	-	-
α_1	5.00000e-04	-	-	-
β_1	0.00000e+00	-	-	-
ω_2	1.00000e-01	-	-	-
α_2	2.00000e-03	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	2 Stationary	2 Chirp	2 Qua Chirp
Probability	-	1.85865e-55	3.16411e-27	3.92384e-40
ω_0	3.00000e-01	2.00958e-01	9.98582e-02	9.96457e-02
α_0	-2.00000e-03	-	1.98417e-03	1.98084e-03
β_0	0.00000e+00	-	-	1.42917e-07
ω_1	2.00000e-01	1.90259e-01	2.00233e-01	1.99824e-01
α_1	5.00000e-04	-	5.27852e-04	5.29073e-04
β_1	0.00000e+00	-	-	2.43423e-07
ω_2	1.00000e-01	-	-	-
α_2	2.00000e-03	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	3 Stationary	3 Chirp	3 Qua Chirp
Probability	-	2.37456e-57	1.00000e+00	1.59574e-20
ω_0	3.00000e-01	2.11362e-01	2.00007e-01	9.96245e-02
α_0	-2.00000e-03	-	5.07929e-04	1.98046e-03
β_0	0.00000e+00	-	-	1.73373e-07
ω_1	2.00000e-01	1.90242e-01	9.98667e-02	2.00216e-01
α_1	5.00000e-04	-	1.98180e-03	5.06532e-04
β_1	0.00000e+00	-	-	-1.24178e-07
ω_2	1.00000e-01	2.01185e-01	3.00075e-01	3.00301e-01
α_2	2.00000e-03	-	-2.00480e-03	-2.00679e-03
β_2	0.00000e+00	-	-	-1.51997e-07

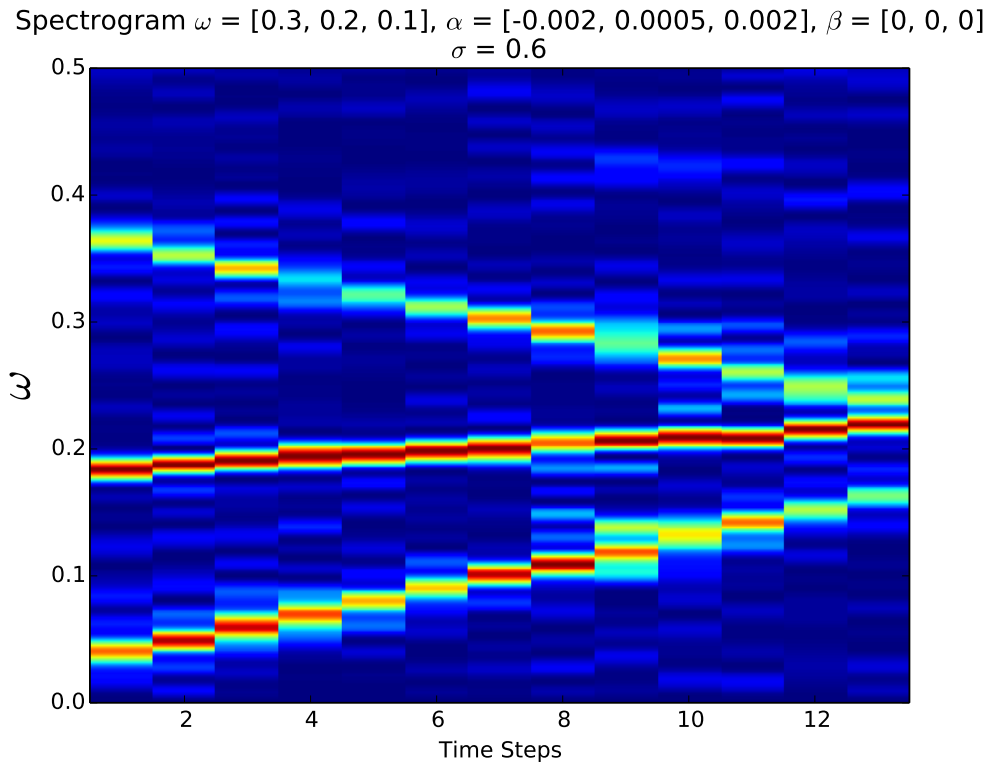


Figure C.14: Spectrogram of three simulated chirped sinusoids.

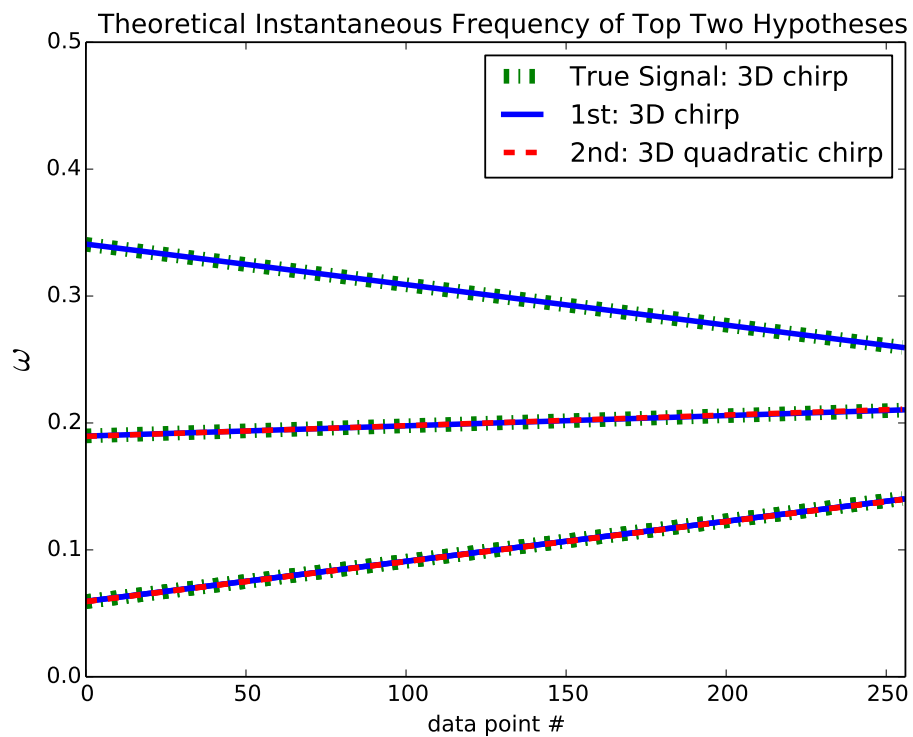


Figure C.15: Theoretical instantaneous frequency of the top two hypotheses.

C.8 One Quadratic Chirped Sinusoid

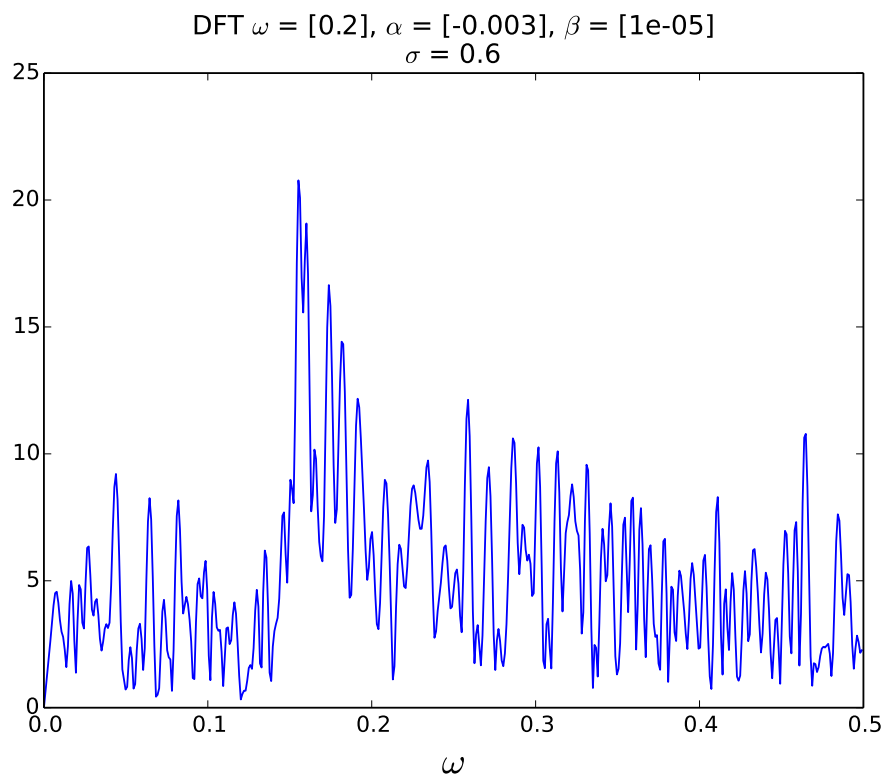


Figure C.16: *DFT of a single simulated quadratic chirped sinusoid.*

Table C.7: *Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	True Model	1 Stationary	1 Chirp	1 Qua Chirp
Probability	-	1.16112e-27	1.69484e-31	1.00000e+00
ω_0	2.00000e-01	1.55591e-01	1.55538e-01	1.99697e-01
α_0	-3.00000e-03	-	6.86067e-05	-2.96326e-03
β_0	1.00000e-05	-	-	1.02660e-05
ω_1	0.00000e+00	-	-	-
α_1	0.00000e+00	-	-	-
β_1	0.00000e+00	-	-	-
ω_2	0.00000e+00	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	2 Stationary	2 Chirp	2 Qua Chirp
Probability	-	8.93356e-26	3.02874e-30	2.30012e-12
ω_0	2.00000e-01	1.57739e-01	1.99693e-01	4.00038e-01
α_0	-3.00000e-03	-	-2.04223e-03	4.01834e-03
β_0	1.00000e-05	-	-	-6.29045e-07
ω_1	0.00000e+00	1.57746e-01	1.85008e-01	1.99707e-01
α_1	0.00000e+00	-	-4.93335e-03	-2.96307e-03
β_1	0.00000e+00	-	-	1.02623e-05
ω_2	0.00000e+00	-	-	-
α_2	0.00000e+00	-	-	-
β_2	0.00000e+00	-	-	-
	True Model	3 Stationary	3 Chirp	3 Qua Chirp
Probability	-	6.09888e-22	1.74159e-28	1.58673e-23
ω_0	2.00000e-01	1.57801e-01	1.74325e-01	4.14915e-02
α_0	-3.00000e-03	-	-6.10444e-04	-2.14417e-05
β_0	1.00000e-05	-	-	-1.02920e-02
ω_1	0.00000e+00	1.57801e-01	1.81061e-01	1.75765e-01
α_1	0.00000e+00	-	-2.89347e-04	1.33060e-04
β_1	0.00000e+00	-	-	4.97540e-04
ω_2	0.00000e+00	1.81610e-01	1.84939e-01	1.99874e-01
α_2	0.00000e+00	-	-4.92379e-03	-2.95907e-03
β_2	0.00000e+00	-	-	1.01643e-05

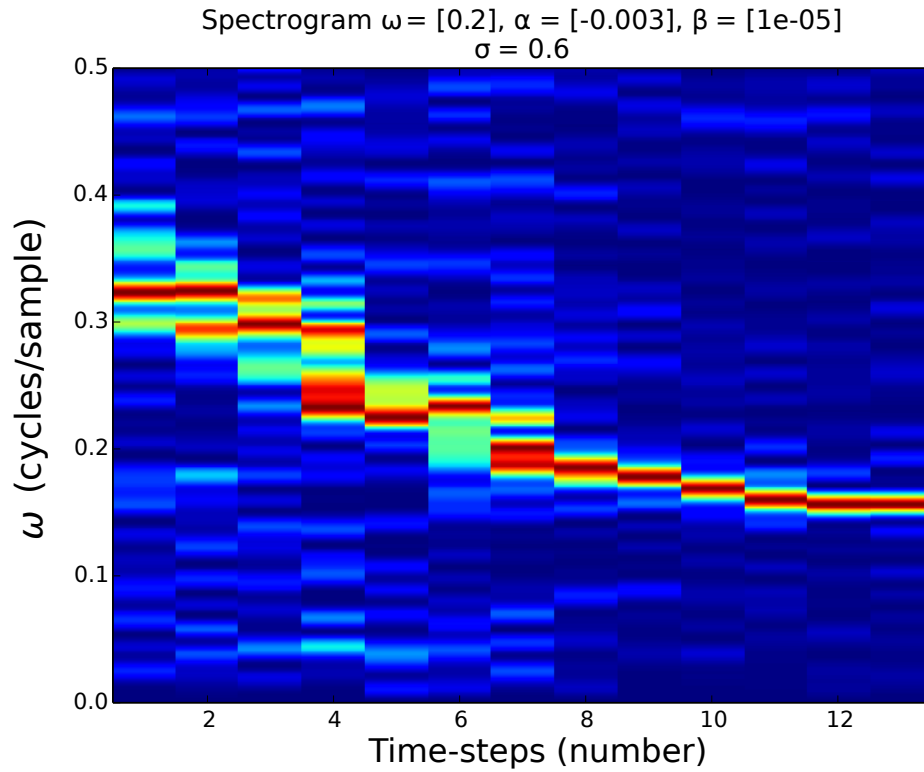


Figure C.17: Spectrogram of a single simulated quadratic chirped sinusoid

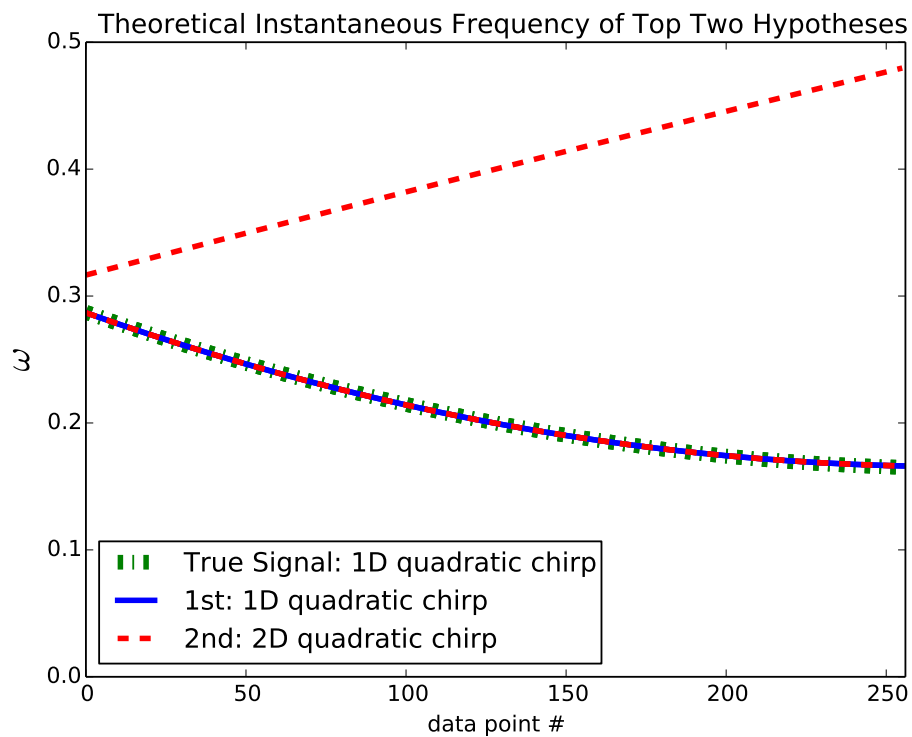


Figure C.18: Theoretical instantaneous frequency of the top two hypotheses.

Appendix D

Doppler Radar Signal BSA

D.1 Introduction

In this appendix we use BSA on Doppler radar data. We use the noiseless model from Equation 3.5.7 for parameter estimation and Equation 3.7.2 to approximate model selection. The results are summarised in Section 4.6.

D.2 Signal Segments

We divided the Doppler radar signals into seven possible segments see Figure D.1 for a golf swing hitting a ball and Figure D.2 where no ball was hit.

Segment A: the downward swing of the golf club towards the ball and possibly low velocity movement of the player.

Segment B: the golf ball only.

Segment C: the golf ball and post-impact golf club.

Segment D: the downward swing of the golf club and stick-flash pre-impact and possibly low velocity movement of the player.

Segment E: the golf club and stick-flash with no impact .

Segment F: the golf club, stick-flash and ball post-impact .

Segment G: the golf club with no impact and after stick-flash .

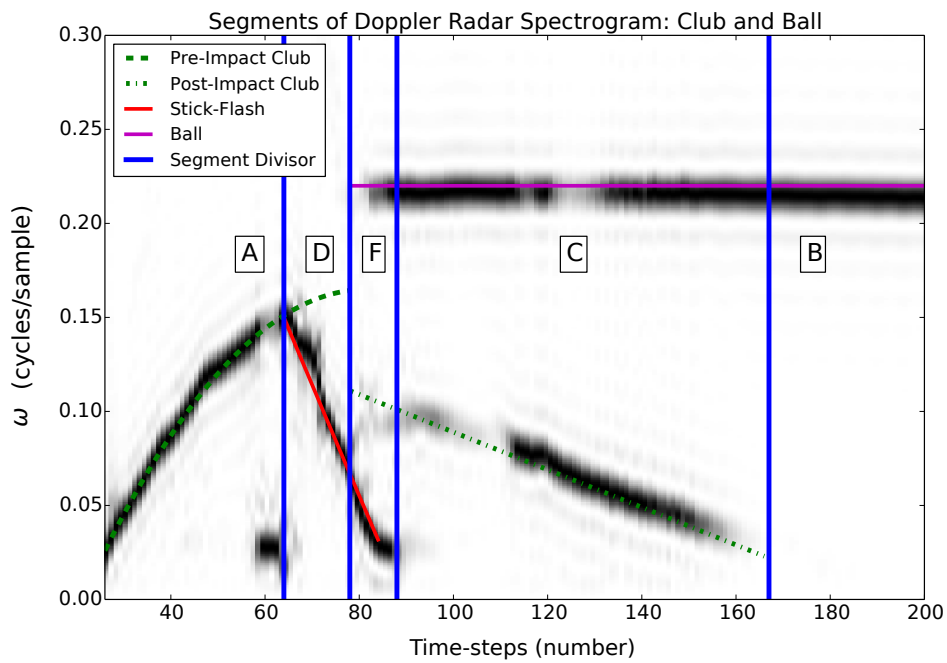


Figure D.1: Spectrogram of a golf swing hitting a golf ball segmented into different parts: Window Size = 64, Stepsize = 16.

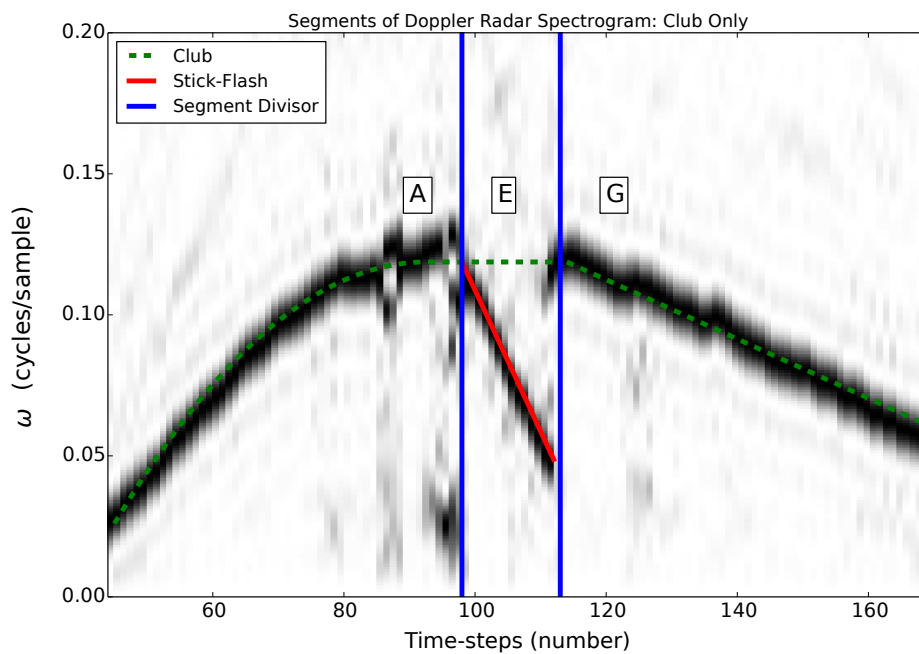


Figure D.2: Spectrogram of a golf swing without hitting a ball segmented into different parts: Window Size = 64, Stepsize = 16.

D.3 Segment A: Golf Club Pre-Impact

D.3.1 Example 1

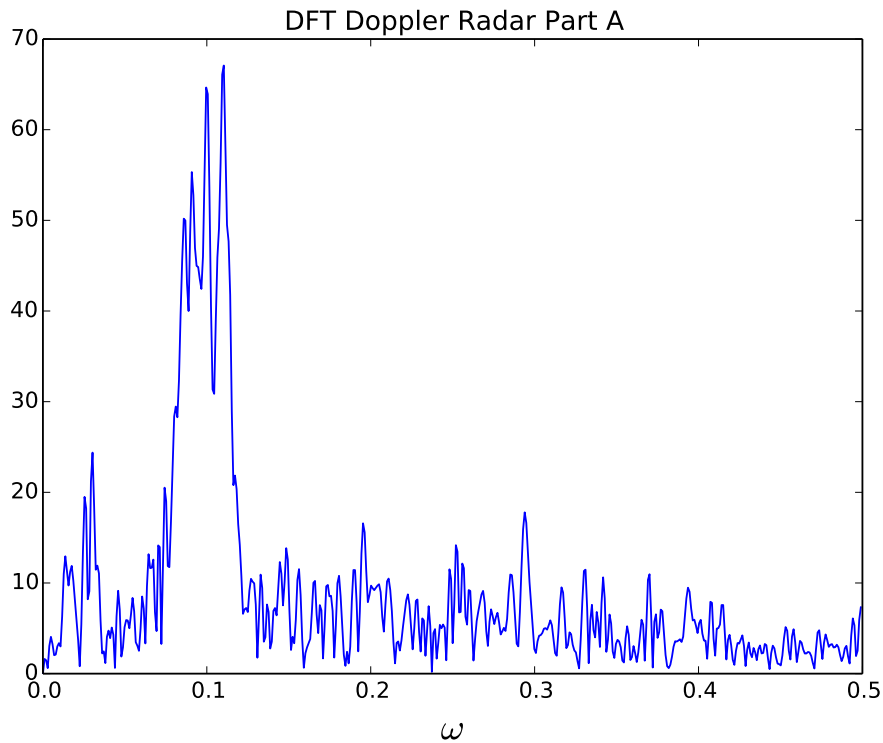


Figure D.3: *Segment A Example 1: DFT of Segment A containing a non-linear chirped signal around $\omega = 0.1$ and low frequency noise from player movement around $\omega = 0.01$*

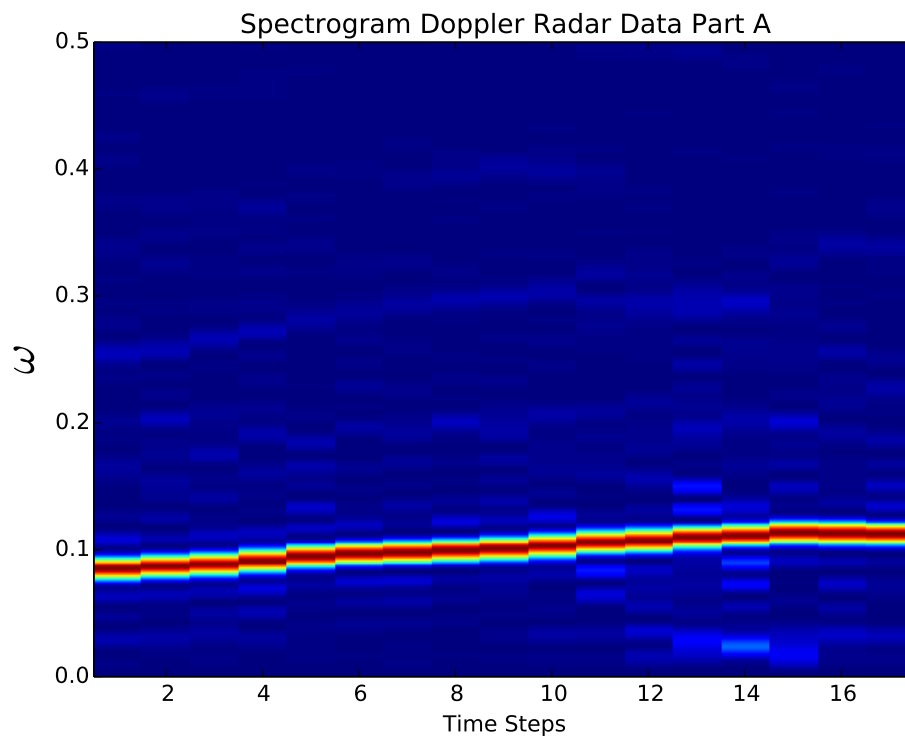


Figure D.4: *Segment A Example 1: Spectrogram of Segment A containing a non-linear chirped signal around $\omega = 0.1$ and low frequency noise from player movement around $\omega = 0.01$ time-step*

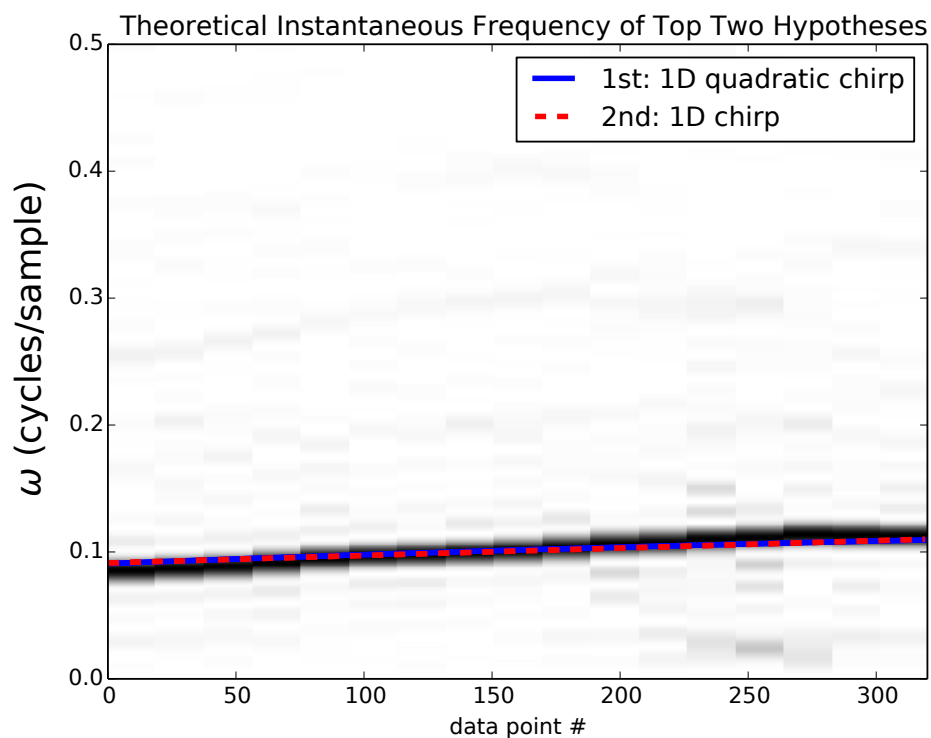


Figure D.5: *Segment A Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram*

Table D.1: *Segment A Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	1 Stationary	1 Chirp	1 Qua Chirp
Probability	5.84543e-88	1.31767e-08	1.00000e+00
ω_0	1.09936e-01	1.00709e-01	1.01549e-01
α_0	-	3.67783e-04	3.65845e-04
β_0	-	-	-3.40525e-07
ω_1	-	-	-
α_1	-	-	-
β_1	-	-	-
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	2 Stationary	2 Chirp	2 Qua Chirp
Probability	1.89992e-82	9.74774e-11	2.61342e-10
ω_0	1.09674e-01	1.00673e-01	1.01446e-01
α_0	-	3.56401e-04	3.65217e-04
β_0	-	-	-3.05506e-07
ω_1	1.00316e-01	1.02332e-01	2.85188e-02
α_1	-	5.24731e-04	3.55360e-03
β_1	-	-	-6.79413e-06
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	3 Stationary	3 Chirp	3 Qua Chirp
Probability	1.08009e-67	4.24825e-13	1.00635e-12
ω_0	1.11077e-01	9.92517e-02	2.95174e-01
α_0	-	1.42578e-04	4.45709e-04
β_0	-	-	-2.81844e-06
ω_1	1.11077e-01	1.00641e-01	1.01376e-01
α_1	-	3.35834e-04	3.65903e-04
β_1	-	-	-2.86003e-07
ω_2	9.95685e-02	1.02070e-01	2.03593e-02
α_2	-	5.31948e-04	3.55566e-03
β_2	-	-	-5.20240e-06

D.3.2 Example 2

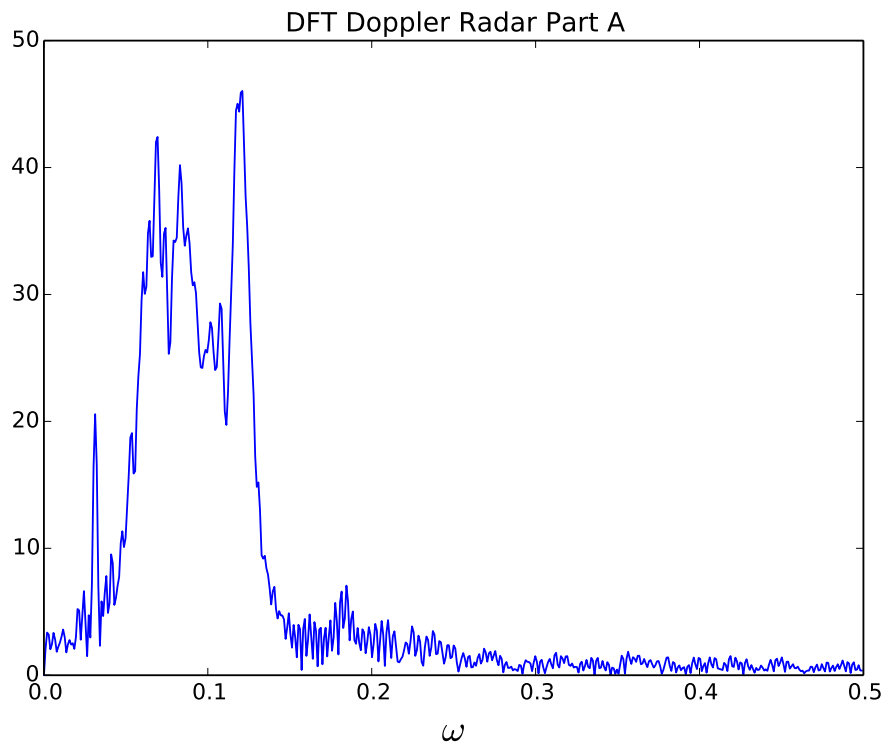


Figure D.6: *Segment A Example 2: DFT of Segment A containing a non-linear chirped signal around $\omega = 0.1$ and low frequency noise from player movement around $\omega = 0.03$*

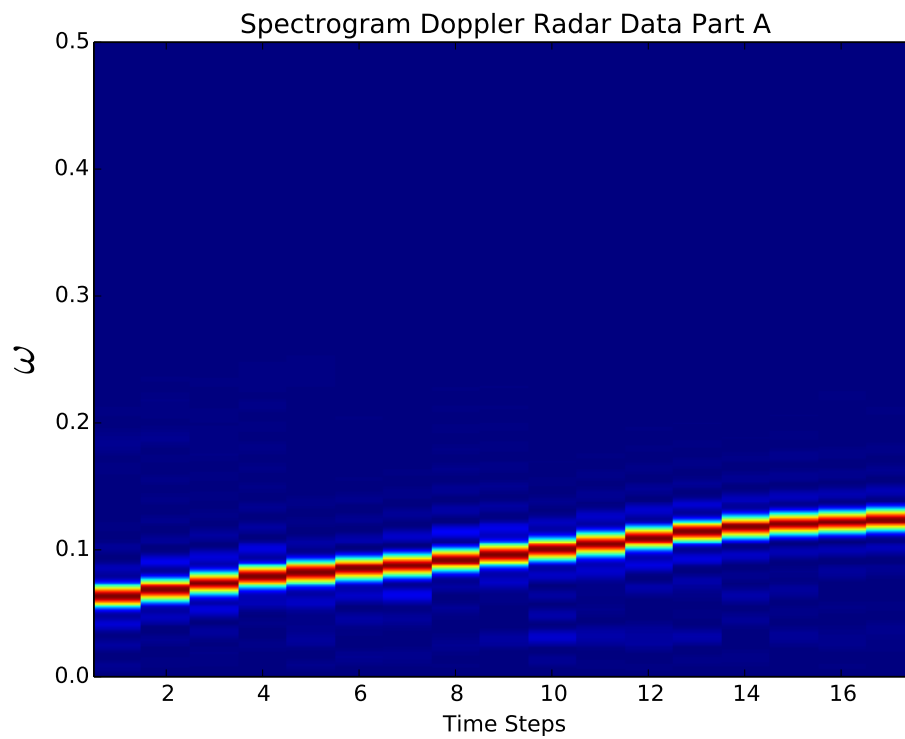


Figure D.7: *Segment A Example 2: Spectrogram of Segment A containing a non-linear chirped signal around $\omega = 0.1$ and low frequency noise from player movement around $\omega = 0.01$ time-step*

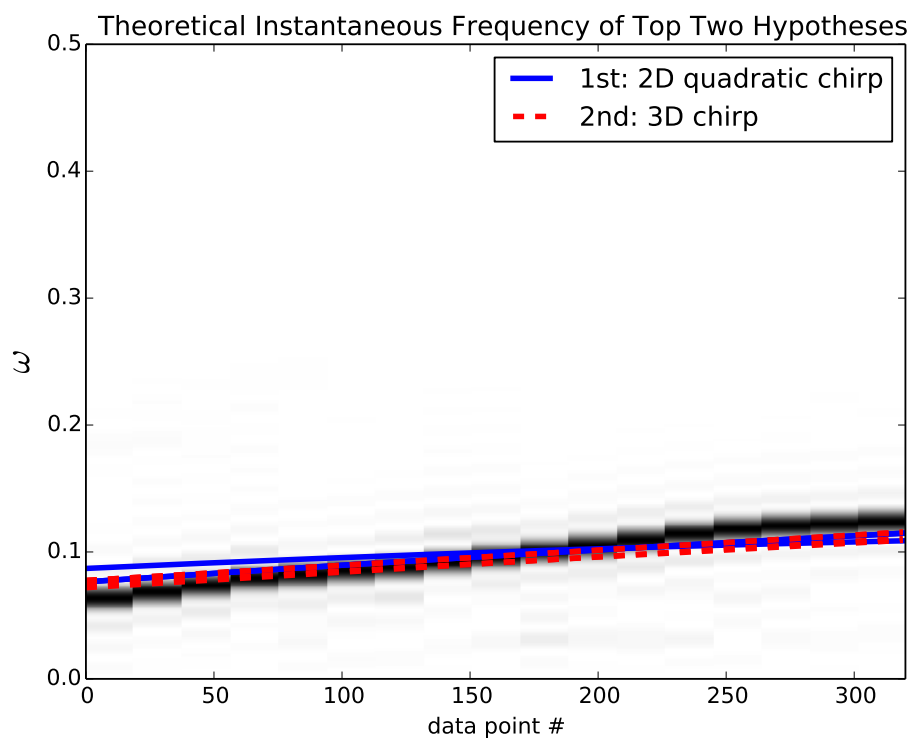


Figure D.8: *Segment A Example 2: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram*

Table D.2: *Segment A Example 2: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	1 Stationary	1 Chirp	1 Qua Chirp
Probability	3.56968e-176	1.72207e-41	7.08154e-08
ω_0	1.20646e-01	9.61201e-02	9.68794e-02
α_0	-	7.59337e-04	7.56767e-04
β_0	-	-	-3.06713e-07
ω_1	-	-	-
α_1	-	-	-
β_1	-	-	-
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	2 Stationary	2 Chirp	2 Qua Chirp
Probability	2.88416e-171	2.48736e-33	9.99622e-01
ω_0	1.19522e-01	1.00002e-01	9.99378e-02
α_0	-	3.48366e-04	4.28169e-04
β_0	-	-	-4.97058e-07
ω_1	1.19523e-01	9.61390e-02	9.69402e-02
α_1	-	7.58981e-04	7.54113e-04
β_1	-	-	-3.19767e-07
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	3 Stationary	3 Chirp	3 Qua Chirp
Probability	5.49060e-163	3.77804e-04	8.52046e-12
ω_0	1.19541e-01	9.13467e-02	1.03033e-01
α_0	-	7.65585e-04	2.68244e-04
β_0	-	-	1.20802e-06
ω_1	6.86320e-02	9.15815e-02	9.81887e-02
α_1	-	6.93738e-04	1.75204e-04
β_1	-	-	1.44669e-06
ω_2	1.19541e-01	9.66538e-02	9.72193e-02
α_2	-	7.48793e-04	7.48368e-04
β_2	-	-	-3.86999e-07

D.4 Segment B: Ball Only

D.4.1 Example 1

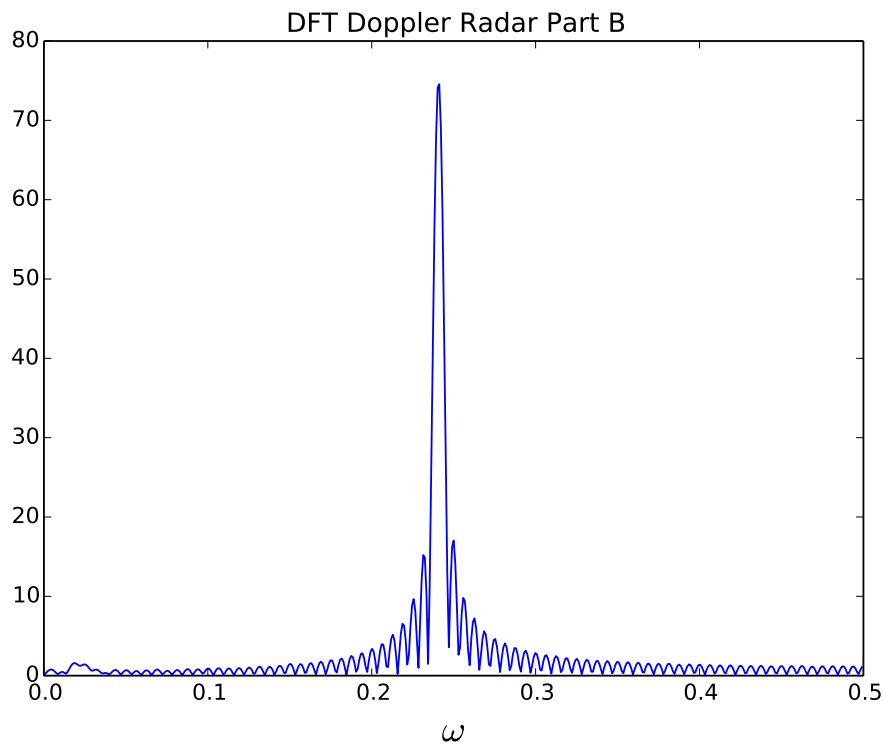


Figure D.9: *Segment B Example 1: DFT of Segment B a chirped signal around $\omega = 0.24$ and a low frequency component around $\omega = 0.02$*

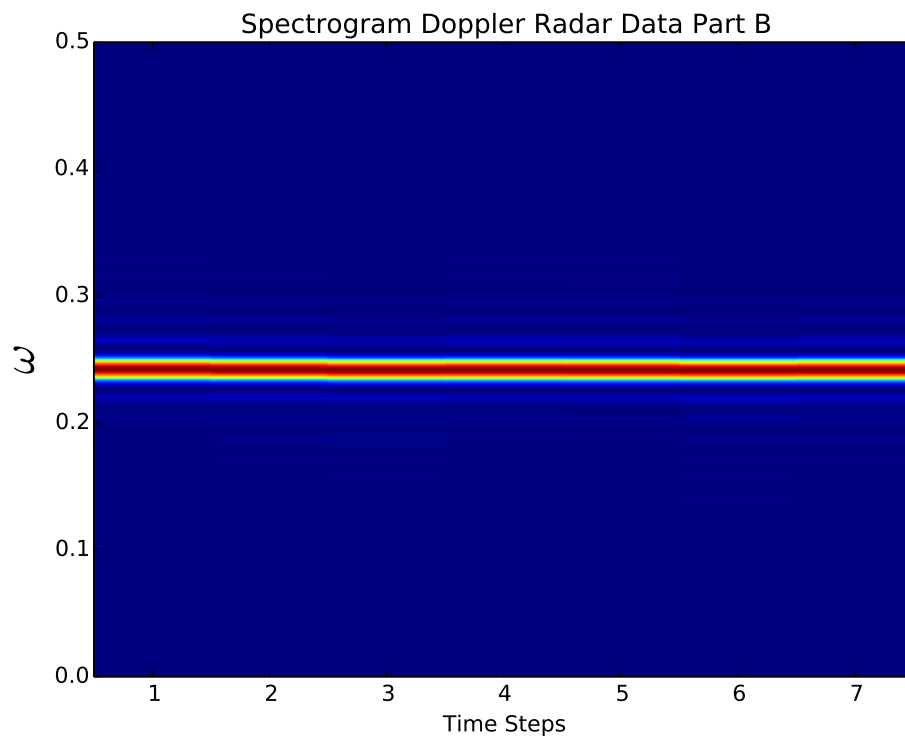


Figure D.10: *Segment B Example 1: Spectrogram of Segment B containing a chirped signal around $\omega = 0.24$*

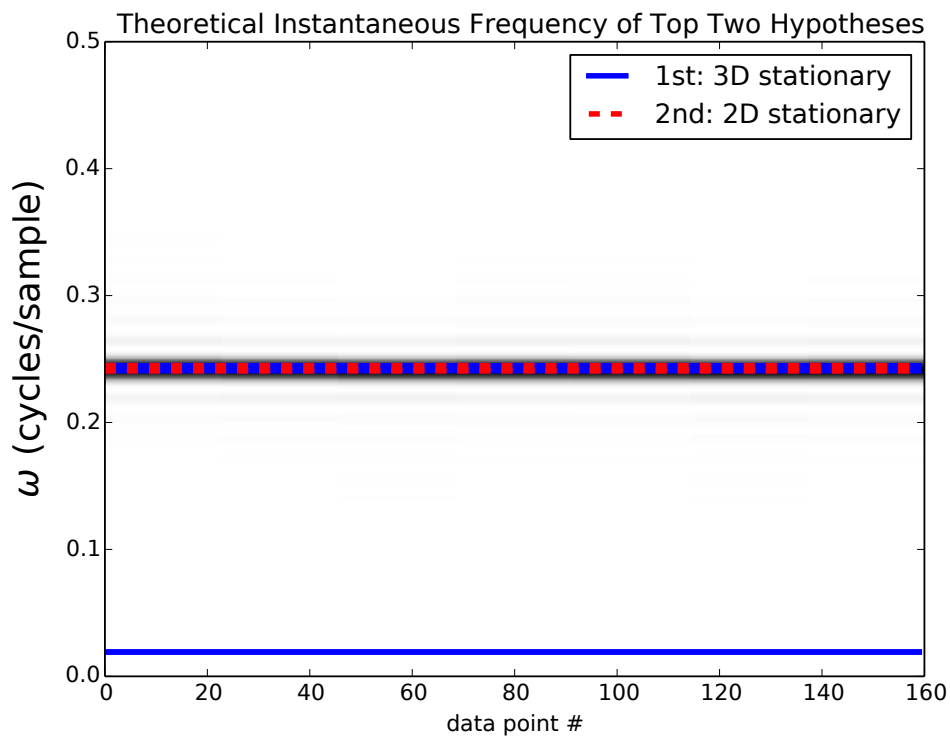


Figure D.11: *Segment B Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram*

Table D.3: *Segment B Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	1 Stationary	1 Chirp	1 Qua Chirp
Probability	1.09799e-52	1.09994e-48	9.60692e-56
ω_0	2.40773e-01	2.40765e-01	2.40754e-01
α_0	-	-1.53998e-05	-1.52208e-05
β_0	-	-	1.72591e-08
ω_1	-	-	-
α_1	-	-	-
β_1	-	-	-
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	2 Stationary	2 Chirp	2 Qua Chirp
Probability	1.54779e-14	1.86034e-18	2.84151e-77
ω_0	2.44819e-01	2.40775e-01	2.40806e-01
α_0	-	-1.22918e-05	-1.45535e-05
β_0	-	-	3.63290e-08
ω_1	2.40793e-01	2.42048e-01	2.41738e-01
α_1	-	-2.74134e-05	-2.04413e-05
β_1	-	-	-3.36357e-07
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	3 Stationary	3 Chirp	3 Qua Chirp
Probability	1.00000e+00	1.00635e-22	1.89992e-82
ω_0	2.44803e-01	2.40747e-01	2.40893e-01
α_0	-	-1.15483e-05	-2.04602e-05
β_0	-	-	9.33571e-08
ω_1	2.40793e-01	5.49429e-03	2.40037e-01
α_1	-	1.45386e-03	4.01027e-07
β_1	-	-	1.69379e-07
ω_2	1.91549e-02	2.42267e-01	3.22724e-02
α_2	-	-2.95890e-05	-7.43140e-03
β_2	-	-	5.05236e-05

D.5 Segment C: Club and Ball Post-Impact

D.5.1 Example 1

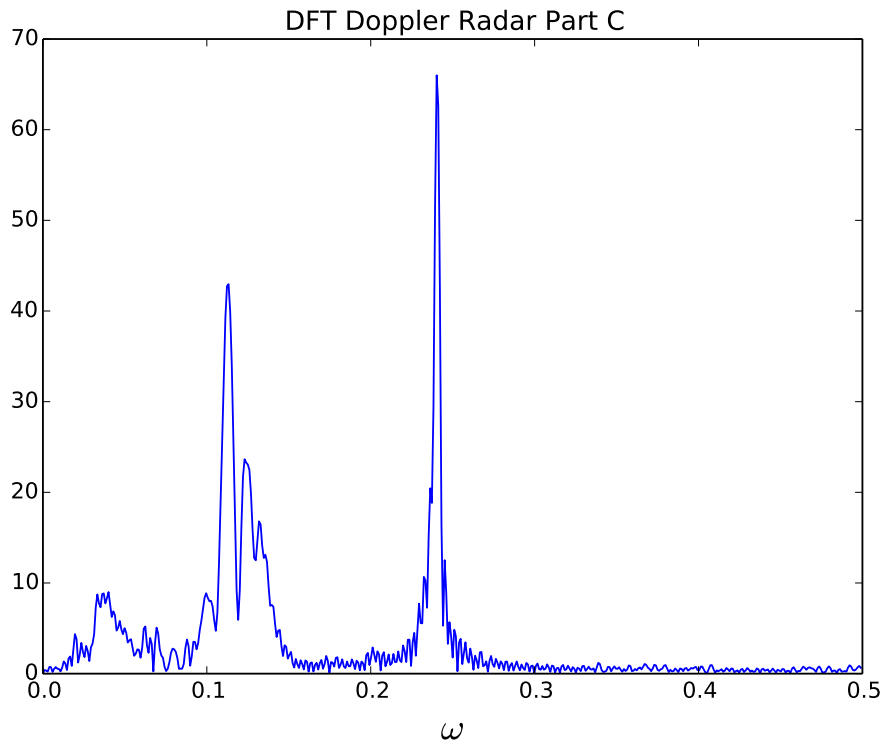


Figure D.12: *Segment C Example 1: DFT of Segment C a signal around $\omega = 0.24$, a chirped signal around $\omega = 0.12$ and low frequency noise from player movement around $\omega = 0.04$*

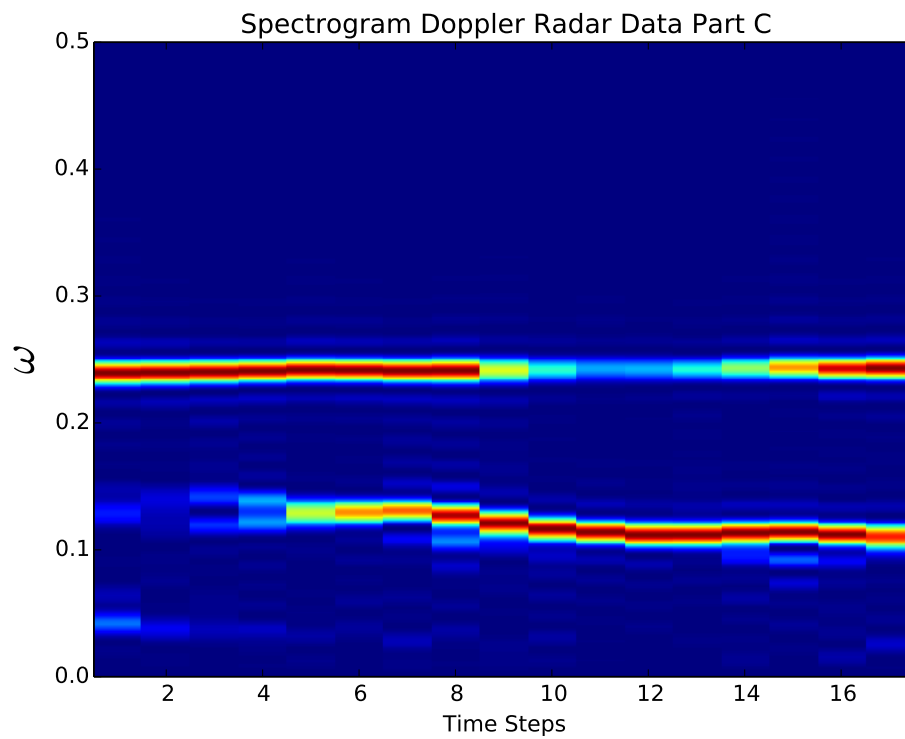


Figure D.13: *Segment C Example 1: Spectrogram of Segment C containing a signal around $\omega = 0.24$, a chirped single around $\omega = 0.12$ and low frequency noise from player movement around $\omega = 0.04$ time-steps 1 to 6*

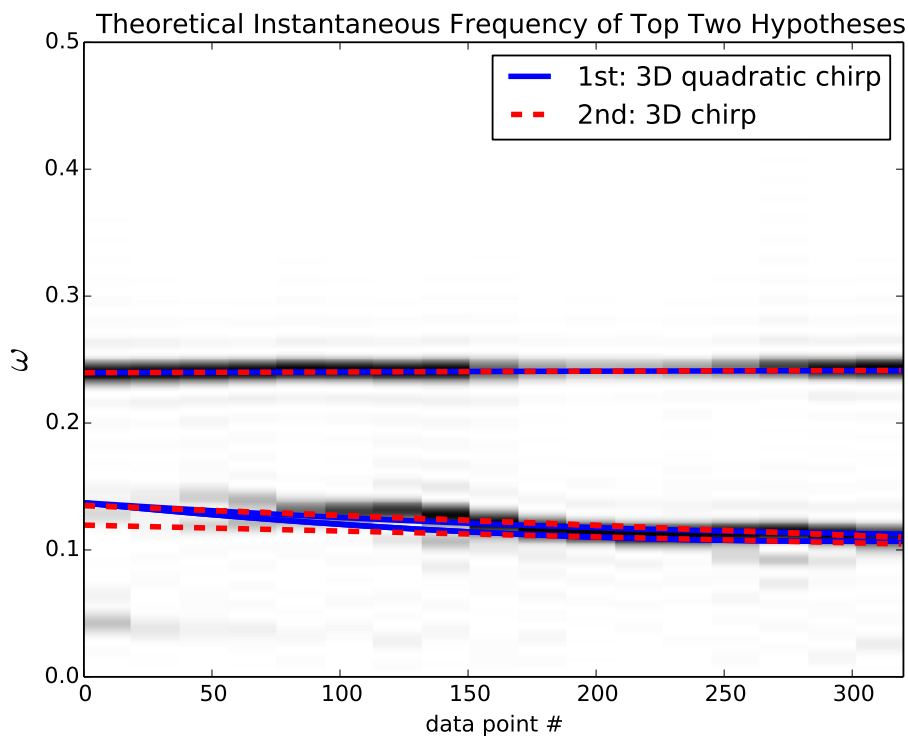


Figure D.14: *Segment C Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram*

Table D.4: *Segment C Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	1 Stationary	1 Chirp	1 Qua Chirp
Probability	1.95712e-63	2.06958e-65	4.95433e-70
ω_0	2.40547e-01	2.40573e-01	2.40726e-01
α_0	-	3.41154e-05	3.38097e-05
β_0	-	-	-6.49250e-08
ω_1	-	-	-
α_1	-	-	-
β_1	-	-	-
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	2 Stationary	2 Chirp	2 Qua Chirp
Probability	7.80940e-46	4.37377e-40	9.33170e-34
ω_0	1.12811e-01	1.22427e-01	2.40679e-01
α_0	-	-4.55851e-04	3.30849e-05
β_0	-	-	-3.75858e-08
ω_1	2.40538e-01	2.40569e-01	1.20425e-01
α_1	-	3.53177e-05	-4.83167e-04
β_1	-	-	1.14647e-06
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	3 Stationary	3 Chirp	3 Qua Chirp
Probability	4.65503e-32	2.94096e-10	1.00000e+00
ω_0	2.40537e-01	1.12123e-01	1.20306e-01
α_0	-	-2.92599e-04	-4.84716e-04
β_0	-	-	1.11141e-06
ω_1	1.11395e-01	2.40569e-01	2.40688e-01
α_1	-	3.51525e-05	3.33435e-05
β_1	-	-	-4.58812e-08
ω_2	1.14584e-01	1.22487e-01	1.13314e-01
α_2	-	-4.93172e-04	-5.90591e-04
β_2	-	-	2.12971e-06

D.5.2 Example 2

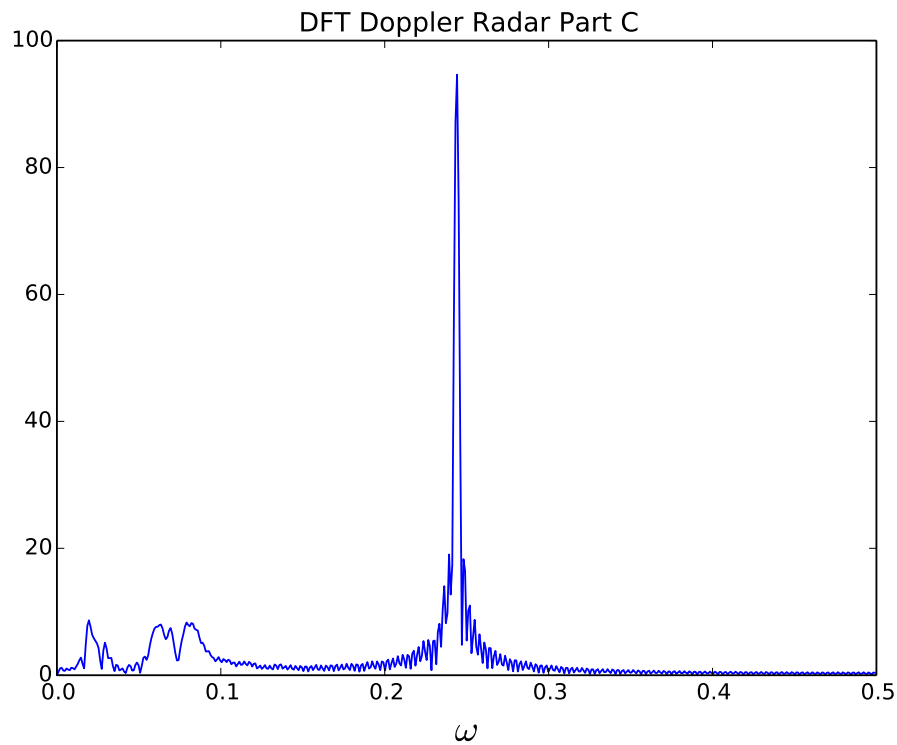


Figure D.15: *Segment C Example 2: DFT of Segment C a signal around $\omega = 0.24$, a chirped signal around $\omega = 0.07$ and low frequency noise around $\omega = 0.02$*

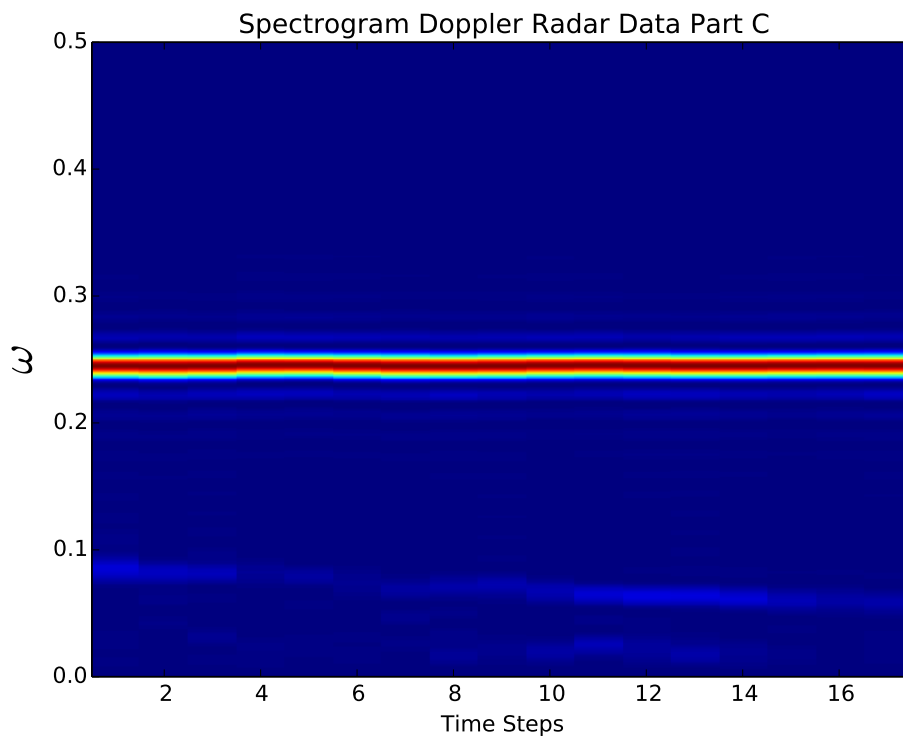


Figure D.16: *Segment C Example 2: Spectrogram of Segment C containing a signal around $\omega = 0.24$, a chirped single around $\omega = 0.07$ and low frequency noise around $\omega = 0.02$*

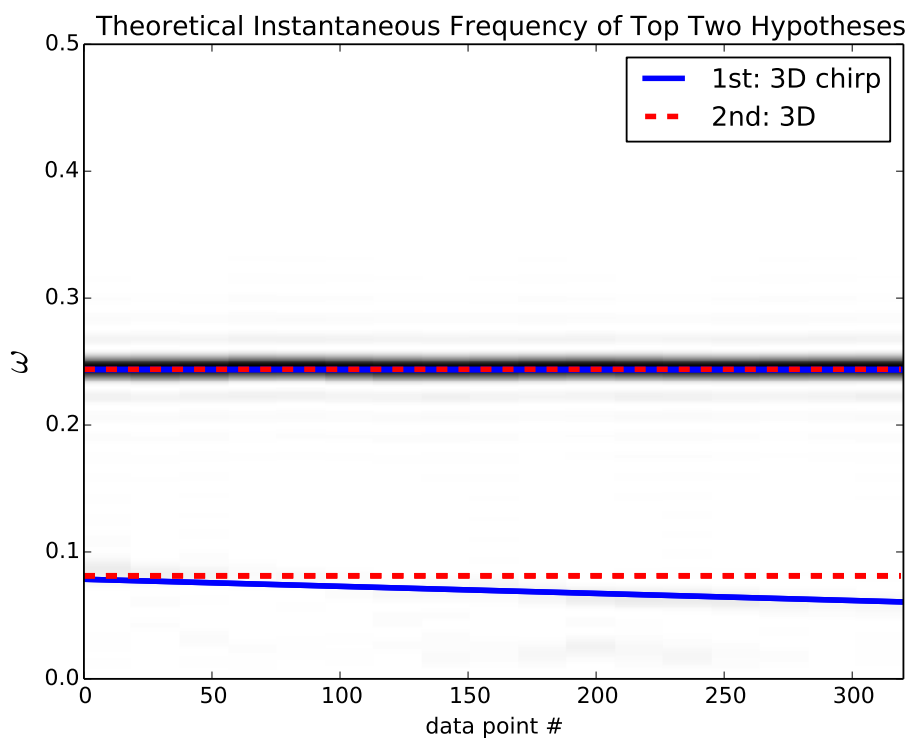


Figure D.17: *Segment C Example 2: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram*

Table D.5: *Segment B Example 2: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	1 Stationary	1 Chirp	1 Qua Chirp
Probability	1.03801e-51	2.70334e-57	1.42196e-59
ω_0	2.43910e-01	2.43912e-01	2.43914e-01
α_0	-	1.19885e-06	1.19824e-06
β_0	-	-	-5.44952e-10
ω_1	-	-	-
α_1	-	-	-
β_1	-	-	-
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	2 Stationary	2 Chirp	2 Qua Chirp
Probability	4.62880e-44	1.33358e-54	7.43686e-65
ω_0	2.43575e-01	2.43785e-01	2.43869e-01
α_0	-	1.22160e-06	1.89033e-06
β_0	-	-	-1.05086e-08
ω_1	2.44055e-01	2.44150e-01	2.44027e-01
α_1	-	2.53800e-06	3.02131e-06
β_1	-	-	3.76429e-08
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	3 Stationary	3 Chirp	3 Qua Chirp
Probability	2.34205e-42	1.00000e+00	5.95873e-57
ω_0	2.43910e-01	2.43943e-01	2.43913e-01
α_0	-	-4.49463e-07	1.37391e-06
β_0	-	-	1.65970e-09
ω_1	8.12714e-02	2.43429e-01	6.88473e-02
α_1	-	-3.97403e-06	-3.49320e-04
β_1	-	-	2.56082e-07
ω_2	8.12715e-02	6.95552e-02	4.40073e-02
α_2	-	-3.50840e-04	-1.41550e-03
β_2	-	-	-9.96614e-06

D.6 Segment D: Club and Stick-Flash Pre-Impact

D.6.1 Example 1

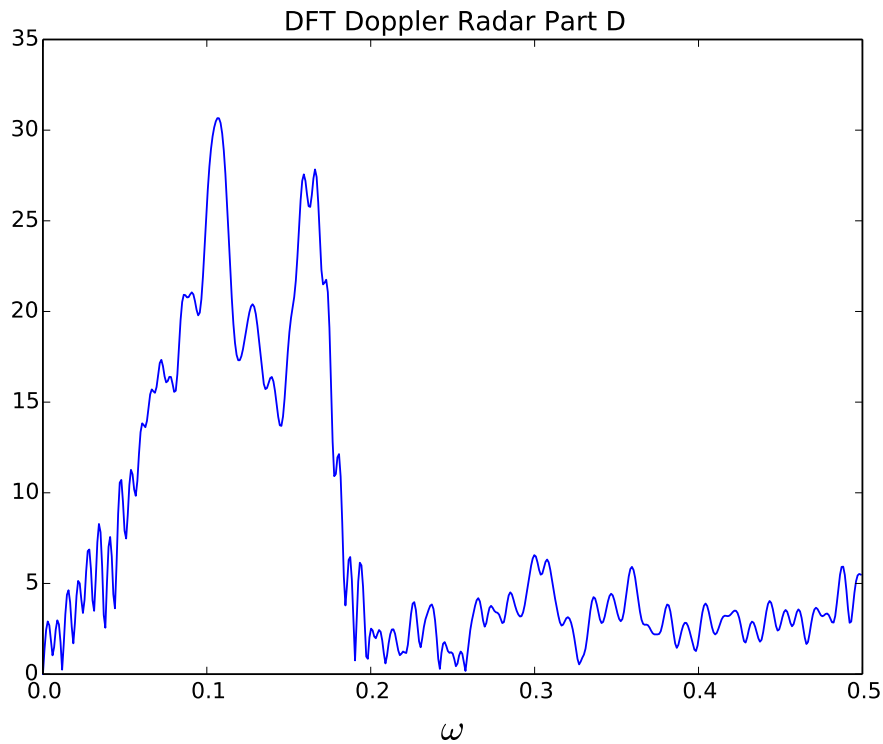


Figure D.18: *Segment D Example 1: DFT of Segment D a chirped signal around $\omega = 0.1$ and $\omega = 0.17$ and possibly a harmonic of that signal around $\omega = 0.3$*

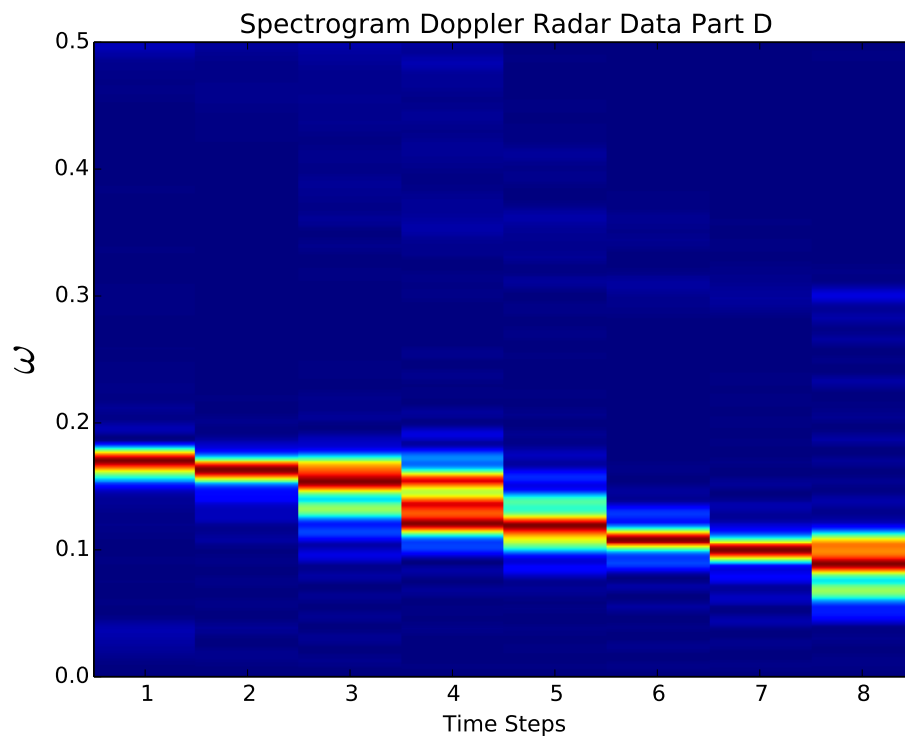


Figure D.19: *Segment D Example 1: Spectrogram of Segment D containing a chirped signal around $\omega = 0.2$ and $\omega = 0.1$ and a harmonic of that signal around $\omega = 0.4$ and $\omega = 0.3$*

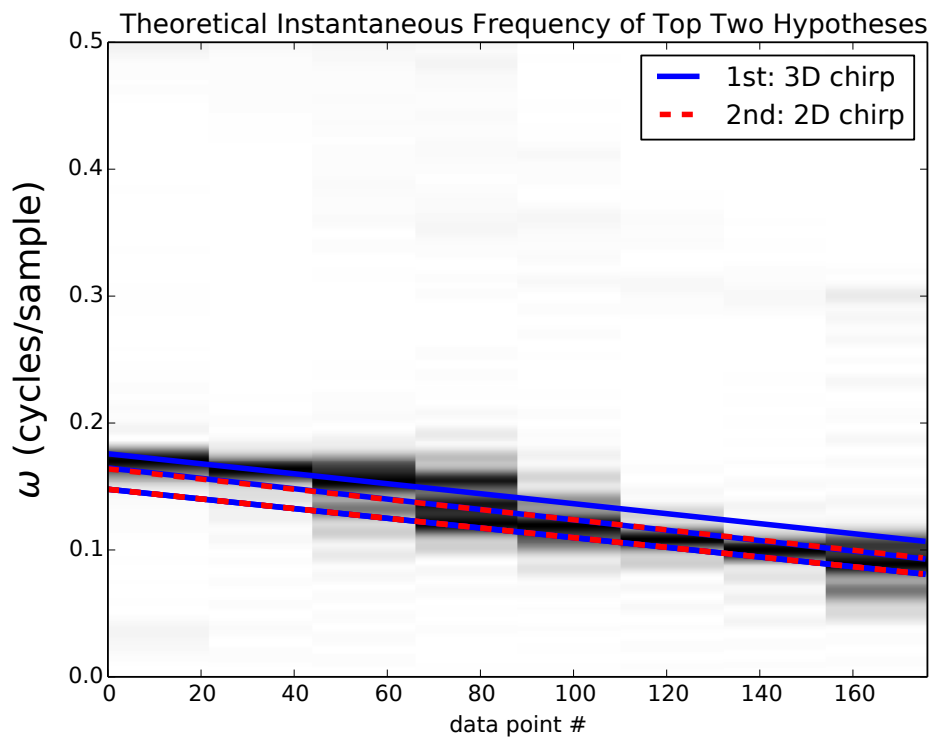


Figure D.20: *Segment D Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram*

Table D.6: *Segment D Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	1 Stationary	1 Chirp	1 Qua Chirp
Probability	1.91765e-54	3.04718e-12	1.22286e-52
ω_0	1.07006e-01	1.28813e-01	1.24580e-01
α_0	-	-2.51798e-03	-1.56109e-03
β_0	-	-	-7.54240e-06
ω_1	-	-	-
α_1	-	-	-
β_1	-	-	-
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	2 Stationary	2 Chirp	2 Qua Chirp
Probability	1.97541e-55	7.84298e-06	2.18709e-19
ω_0	1.02620e-01	1.41018e-01	1.28453e-01
α_0	-	-2.39893e-03	-4.34757e-03
β_0	-	-	-1.37072e-05
ω_1	1.09277e-01	1.28586e-01	1.28230e-01
α_1	-	-2.51685e-03	-2.46518e-03
β_1	-	-	6.82905e-07
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	3 Stationary	3 Chirp	3 Qua Chirp
Probability	2.46091e-62	9.99992e-01	1.85486e-20
ω_0	1.04758e-01	1.41185e-01	1.16052e-01
α_0	-	-2.46994e-03	-2.52113e-03
β_0	-	-	-2.17134e-06
ω_1	8.79806e-02	1.28631e-01	1.26516e-01
α_1	-	-2.56964e-03	-2.43075e-03
β_1	-	-	3.34600e-06
ω_2	8.79806e-02	1.14272e-01	1.30513e-01
α_2	-	-2.39537e-03	-3.01291e-03
β_2	-	-	1.06812e-06

D.6.2 Example 2

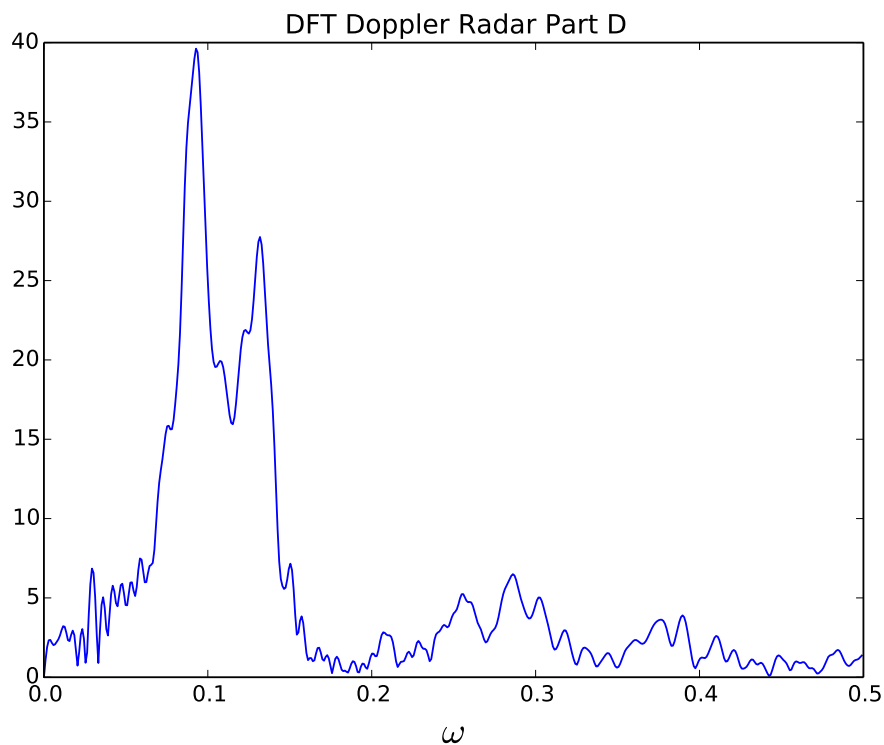


Figure D.21: *Segment D Example 2: DFT of Segment D a chirped signal around $\omega = 0.13$ and $\omega = 0.1$ and possibly a harmonic of that signal around $\omega = 0.27$*

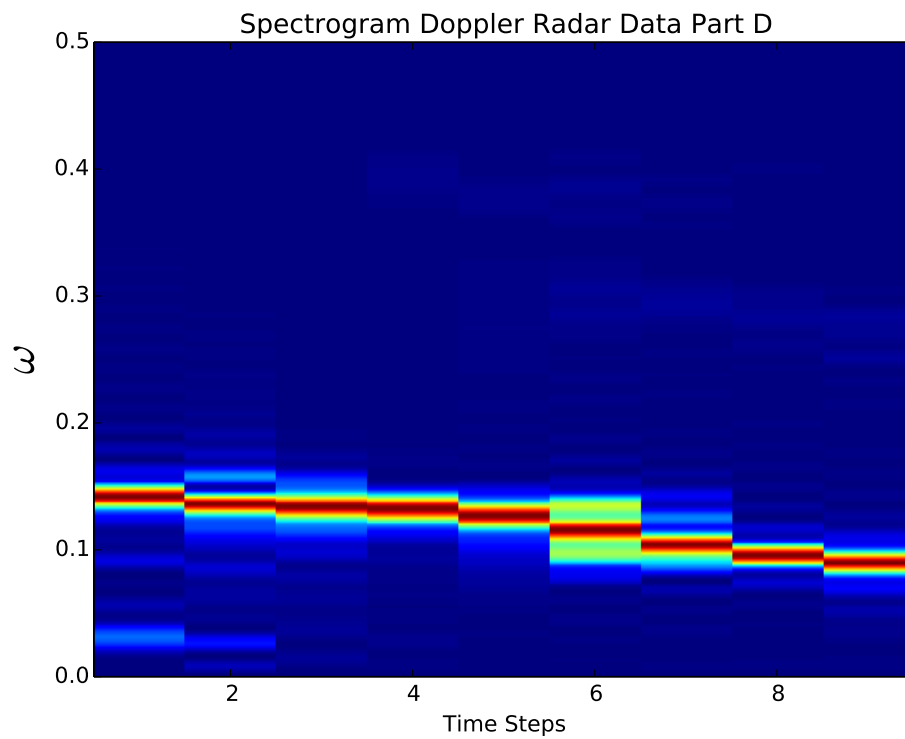


Figure D.22: *Segment D Example 2: Spectrogram of Segment D containing a chirped signal around $\omega = 0.13$ and $\omega = 0.1$ and a harmonic of that signal around $\omega = 0.3$ and $\omega = 0.27$*

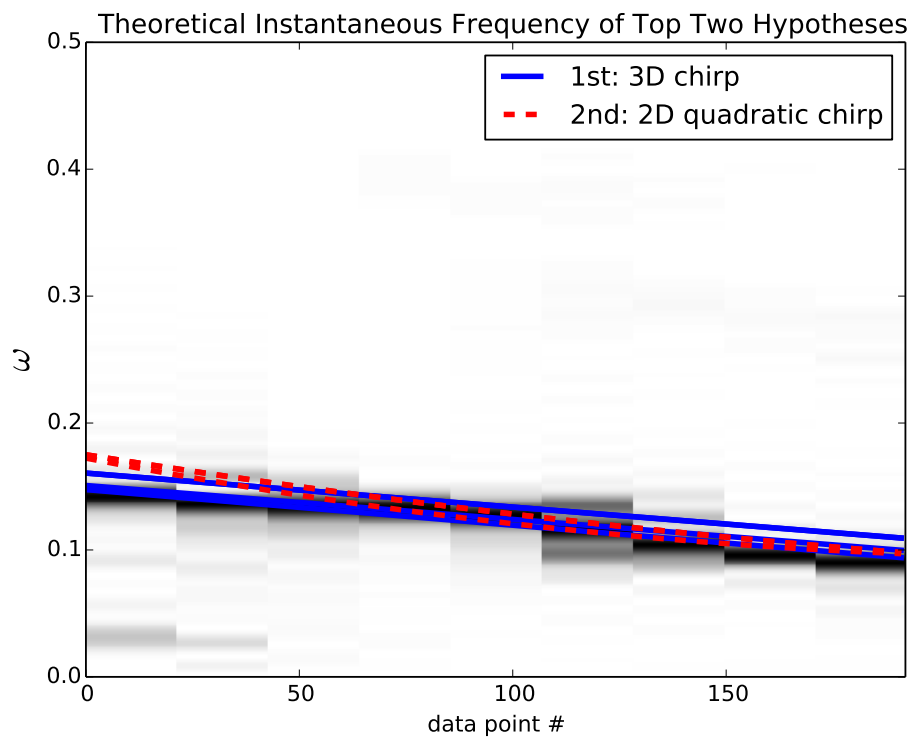


Figure D.23: *Segment D Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram*

Table D.7: *Segment D Example 2: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	1 Stationary	1 Chirp	1 Qua Chirp
Probability	3.41382e-57	2.42914e-38	4.22246e-42
ω_0	9.29244e-02	1.24324e-01	1.22775e-01
α_0	-	-1.81308e-03	-1.37696e-03
β_0	-	-	-3.55484e-06
ω_1	-	-	-
α_1	-	-	-
β_1	-	-	-
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	2 Stationary	2 Chirp	2 Qua Chirp
Probability	1.28718e-46	4.22925e-12	7.89238e-06
ω_0	9.06712e-02	1.25172e-01	1.22217e-01
α_0	-	-1.69984e-03	-2.46699e-03
β_0	-	-	8.47934e-06
ω_1	9.06712e-02	1.21081e-01	1.29845e-01
α_1	-	-1.78412e-03	-2.53152e-03
β_1	-	-	4.41835e-06
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	3 Stationary	3 Chirp	3 Qua Chirp
Probability	8.14462e-56	9.99992e-01	4.90119e-21
ω_0	1.05713e-01	1.25155e-01	1.29437e-01
α_0	-	-1.67801e-03	-3.23810e-04
β_0	-	-	3.14062e-06
ω_1	9.05732e-02	1.20603e-01	1.26362e-01
α_1	-	-1.76192e-03	-2.60763e-03
β_1	-	-	7.63363e-06
ω_2	9.05732e-02	1.34857e-01	1.26301e-01
α_2	-	-1.68182e-03	-2.60681e-03
β_2	-	-	7.66123e-06

D.7 Segment E: Club and Stick-Flash No Impact

D.7.1 Example 1

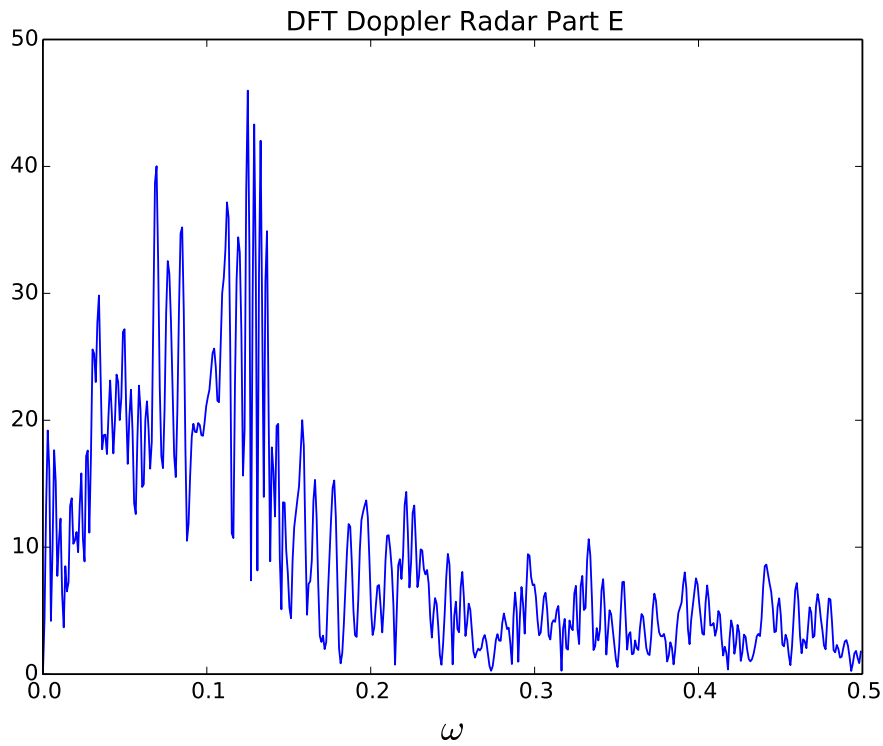


Figure D.24: *Segment E Example 1: DFT of Segment E containing multiple chirped signals*

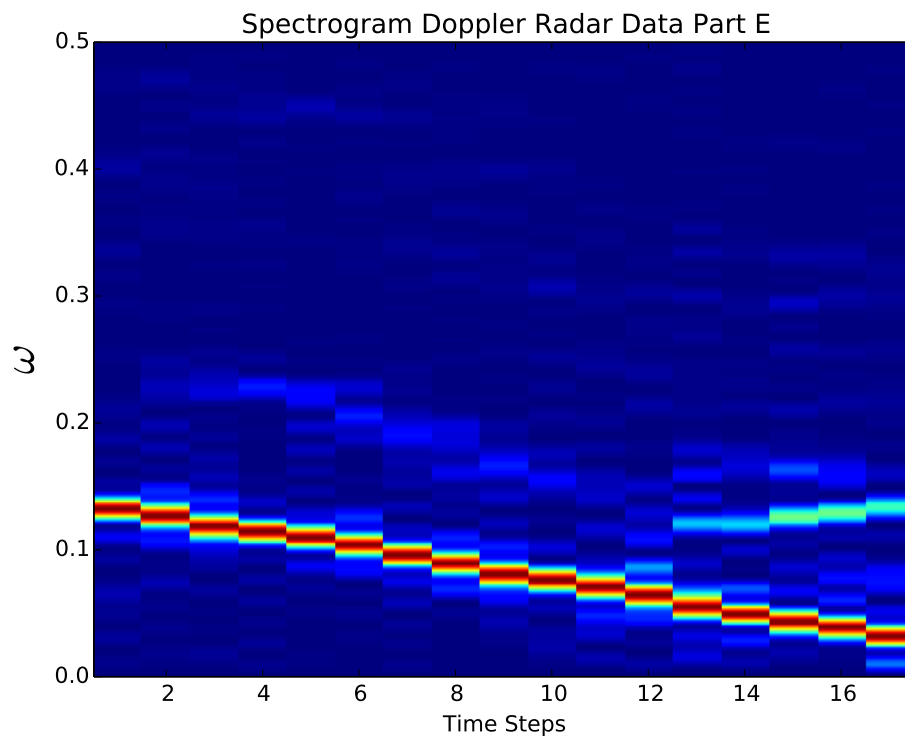


Figure D.25: *Segment E Example 1: Spectrogram of Segment E containing a chirped signal from around $\omega = 0.12$ to $\omega = 0.01$ and its harmonic from around $\omega = 0.22$, time-step 2, to $\omega = 0.12$, time-step 12 and a signal $\omega = 0.12$ from time-step 12 to 17.*

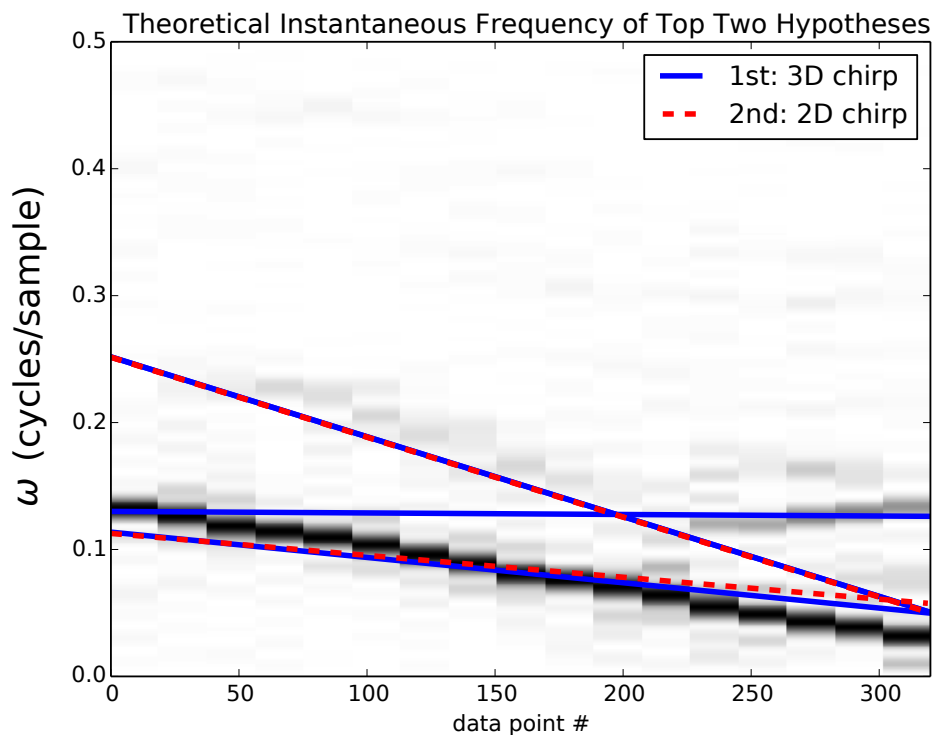


Figure D.26: *Segment E Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram*

Table D.8: *Segment E Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	1 Stationary	1 Chirp	1 Qua Chirp
Probability	1.78231e-54	9.43333e-59	9.87136e-53
ω_0	1.24916e-01	1.24976e-01	7.92215e-02
α_0	-	3.51032e-05	-7.41974e-04
β_0	-	-	-2.88852e-06
ω_1	-	-	-
α_1	-	-	-
β_1	-	-	-
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	2 Stationary	2 Chirp	2 Qua Chirp
Probability	9.76050e-49	2.06664e-45	9.10756e-48
ω_0	1.27009e-01	8.50343e-02	7.14325e-02
α_0	-	-1.08403e-03	-2.35601e-04
β_0	-	-	-4.09540e-06
ω_1	1.27009e-01	1.50726e-01	8.33530e-02
α_1	-	-3.95640e-03	-1.09775e-03
β_1	-	-	4.86832e-07
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	3 Stationary	3 Chirp	3 Qua Chirp
Probability	7.70643e-68	1.00000e+00	2.92082e+56
ω_0	1.27093e-01	1.28067e-01	1.26849e-01
α_0	-	-7.60878e-05	-2.53433e-04
β_0	-	-	-3.73998e-07
ω_1	1.27093e-01	8.17663e-02	1.27513e-01
α_1	-	-1.25331e-03	-2.77401e-05
β_1	-	-	3.99262e-07
ω_2	1.18433e-01	1.51091e-01	8.20565e-02
α_2	-	-3.95223e-03	-1.26471e-03
β_2	-	-	-1.03192e-07

D.7.2 Example 2

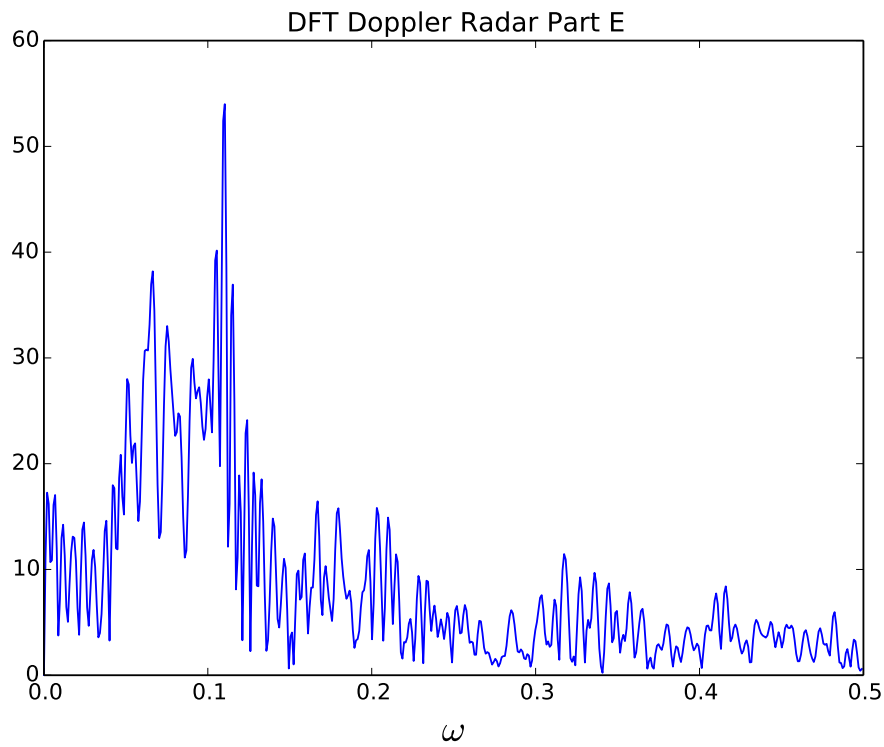


Figure D.27: *Segment E Example 2: DFT of Segment E containing multiple chirped signals*

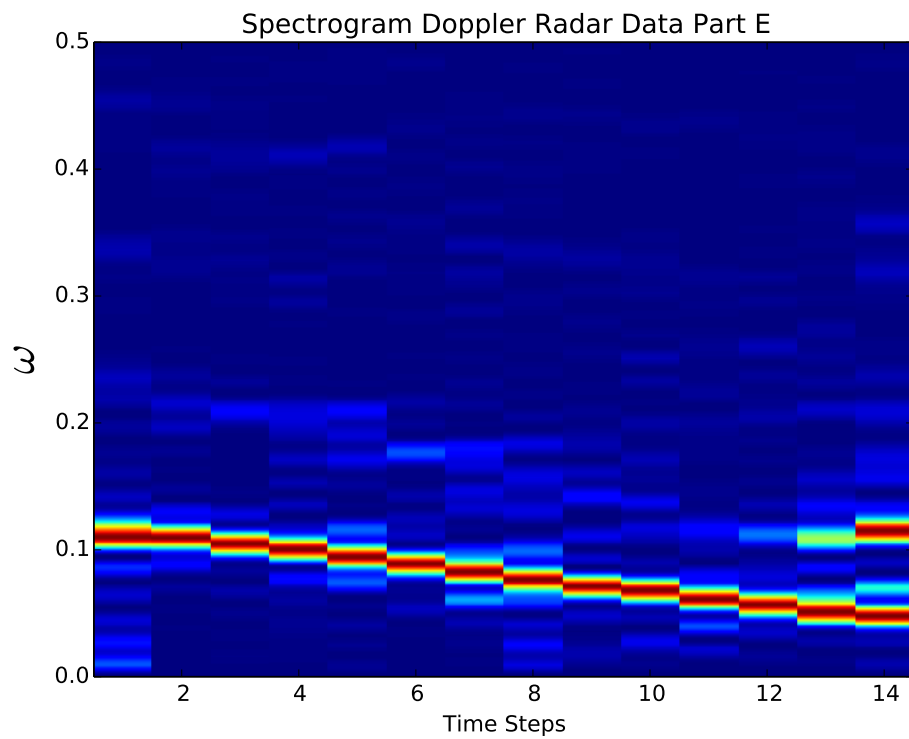


Figure D.28: *Segment E Example 2: Spectrogram of Segment E containing a chirped signal from around $\omega = 0.11$ to $\omega = 0.04$ and its harmonic from around $\omega = 0.2$, time-step 2, to $\omega = 0.11$, time-step 12 and a signal $\omega = 0.12$ from time-step 12 to 14.*

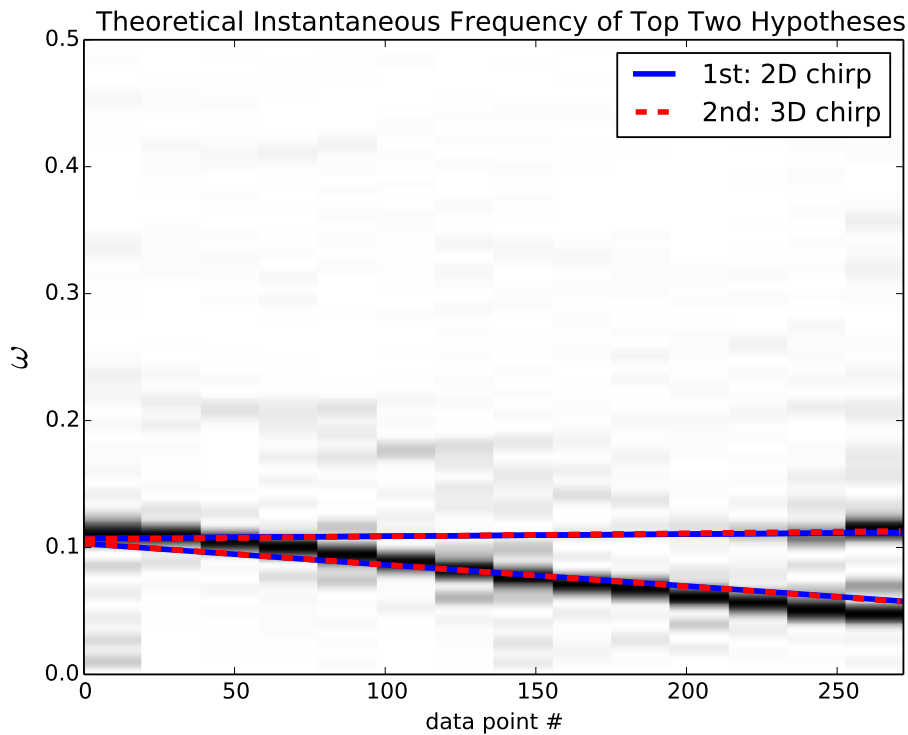


Figure D.29: *Segment E Example 2: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram*

Table D.9: *Segment E Example 2: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	1 Stationary	1 Chirp	1 Qua Chirp
Probability	2.35149e-31	1.61748e-33	5.31476e-40
ω_0	1.09951e-01	1.09833e-01	1.09767e-01
α_0	-	9.75306e-05	9.73808e-05
β_0	-	-	3.55186e-08
ω_1	-	-	-
α_1	-	-	-
β_1	-	-	-
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	2 Stationary	2 Chirp	2 Qua Chirp
Probability	8.82630e-31	9.39473e-01	3.49453e-12
ω_0	6.62759e-02	8.03286e-02	7.95695e-02
α_0	-	-1.04872e-03	-1.05529e-03
β_0	-	-	5.31333e-07
ω_1	1.09904e-01	1.09452e-01	1.09204e-01
α_1	-	1.04554e-04	9.60170e-05
β_1	-	-	1.81311e-07
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	3 Stationary	3 Chirp	3 Qua Chirp
Probability	2.69992e-30	6.05267e-02	1.26394e-17
ω_0	1.06329e-01	1.09609e-01	1.10754e-01
α_0	-	1.61543e-04	5.05496e-05
β_0	-	-	9.89374e-07
ω_1	1.09175e-01	1.09530e-01	7.96669e-02
α_1	-	1.34414e-04	-1.07213e-03
β_1	-	-	5.28679e-07
ω_2	6.62769e-02	8.02698e-02	1.10712e-01
α_2	-	-1.06450e-03	5.67457e-05
β_2	-	-	-2.00554e-07

D.8 Segment F: Club, Stick-Flash and Ball Post-Impact

D.8.1 Example 1

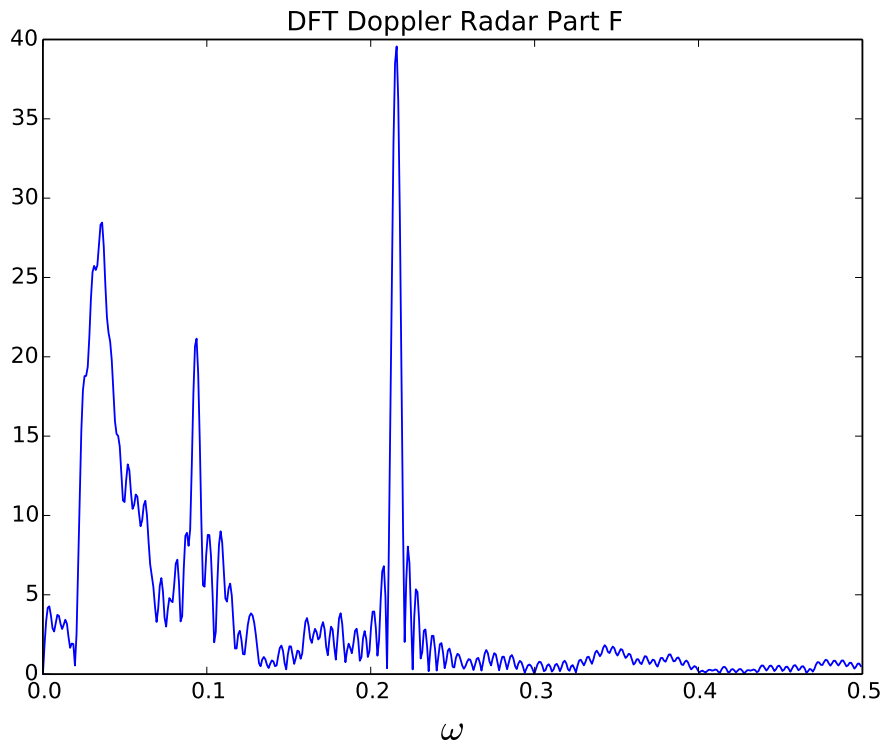


Figure D.30: *Segment F Example 1: DFT of Segment F containing three signals.*

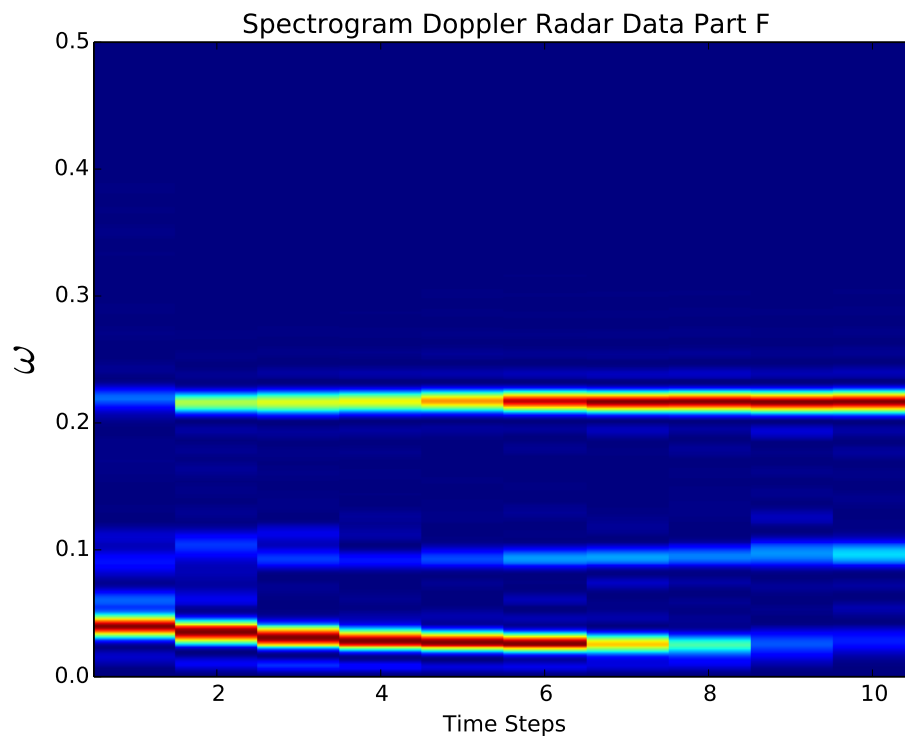


Figure D.31: *Segment F Example 1: Spectrogram of Segment F containing a signal around $\omega = 0.21$, a chirped signal around $\omega = 0.1$ and a chirped signal around $\omega = 0.04$.*

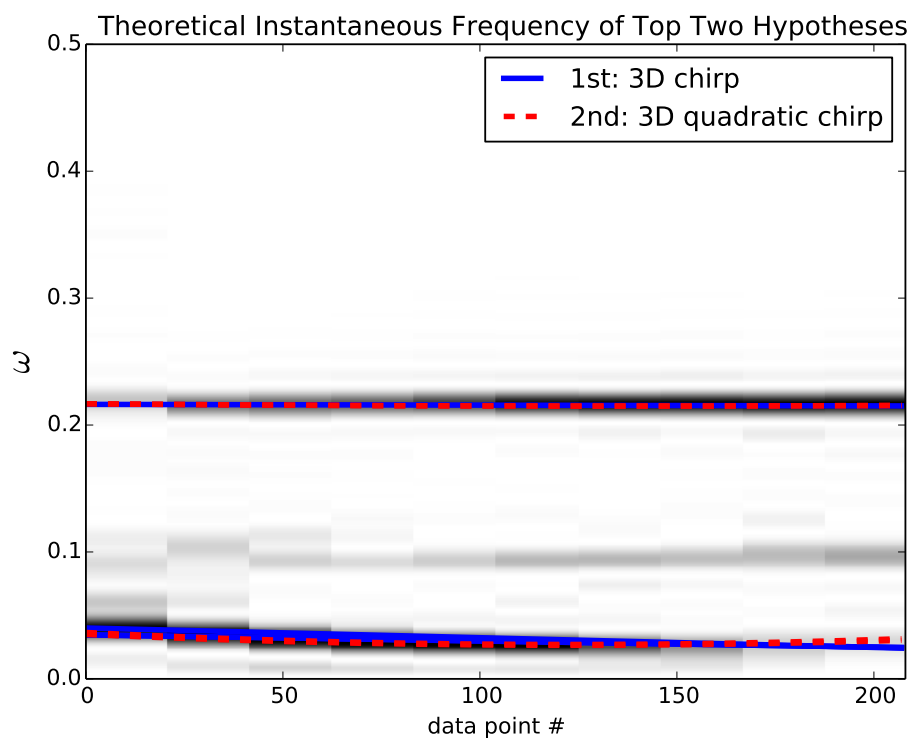


Figure D.32: *Segment F Example 1: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram.*

Table D.10: *Segment F Example 1: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	1 Stationary	1 Chirp	1 Qua Chirp
Probability	3.64181e-26	6.40605e-26	4.44928e-29
ω_0	2.15553e-01	2.86082e-02	2.65994e-02
α_0	-	-3.74075e-04	-2.43965e-04
β_0	-	-	2.70936e-06
ω_1	-	-	-
α_1	-	-	-
β_1	-	-	-
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	2 Stationary	2 Chirp	2 Qua Chirp
Probability	7.41855e-21	9.50721e-10	3.35897e-16
ω_0	3.56186e-02	2.86007e-02	2.66064e-02
α_0	-	-3.75445e-04	-2.44837e-04
β_0	-	-	2.71066e-06
ω_1	2.15545e-01	2.15658e-01	2.14995e-01
α_1	-	-2.94732e-05	-3.88583e-05
β_1	-	-	6.25976e-07
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	3 Stationary	3 Chirp	3 Qua Chirp
Probability	665028e-12	1.00000e+00	1.33521e-08
ω_0	3.75126e-02	2.15673e-01	2.66187e-02
α_0	-	-3.65779e-05	-1.37258e-04
β_0	-	-	3.54688e-06
ω_1	2.15601e-01	3.21418e-02	2.73138e-02
α_1	-	-4.81974e-04	-1.45972e-04
β_1	-	-	3.97048e-06
ω_2	3.75126e-02	2.96735e-02	2.15005e-01
α_2	-	-3.03075e-04	-3.94227e-05
β_2	-	-	6.36121e-07

D.8.2 Example 2

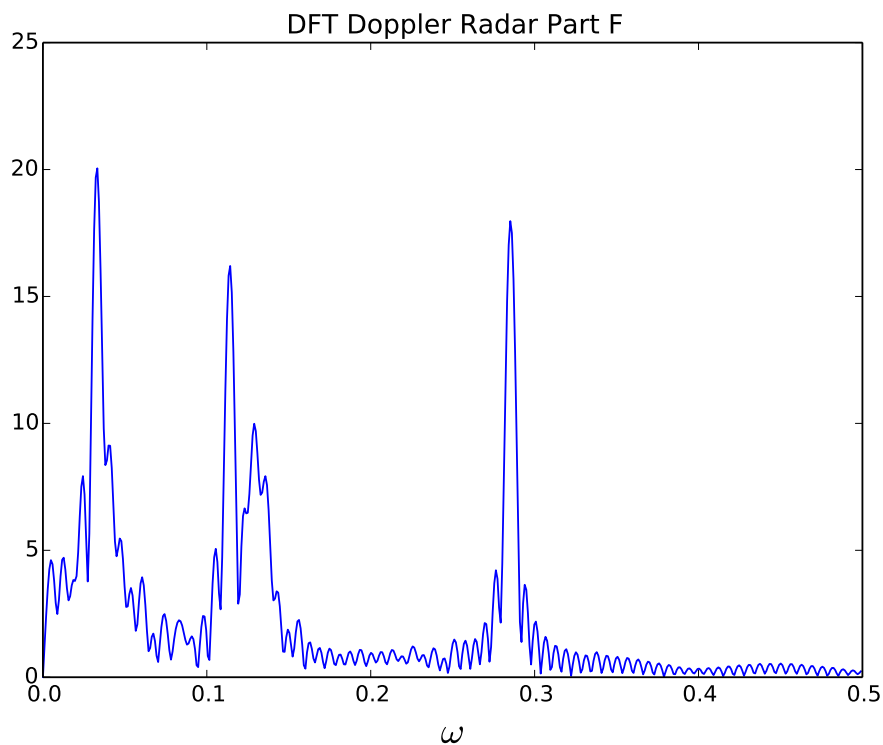


Figure D.33: *Segment F Example 2: DFT of Segment F containing three signals*

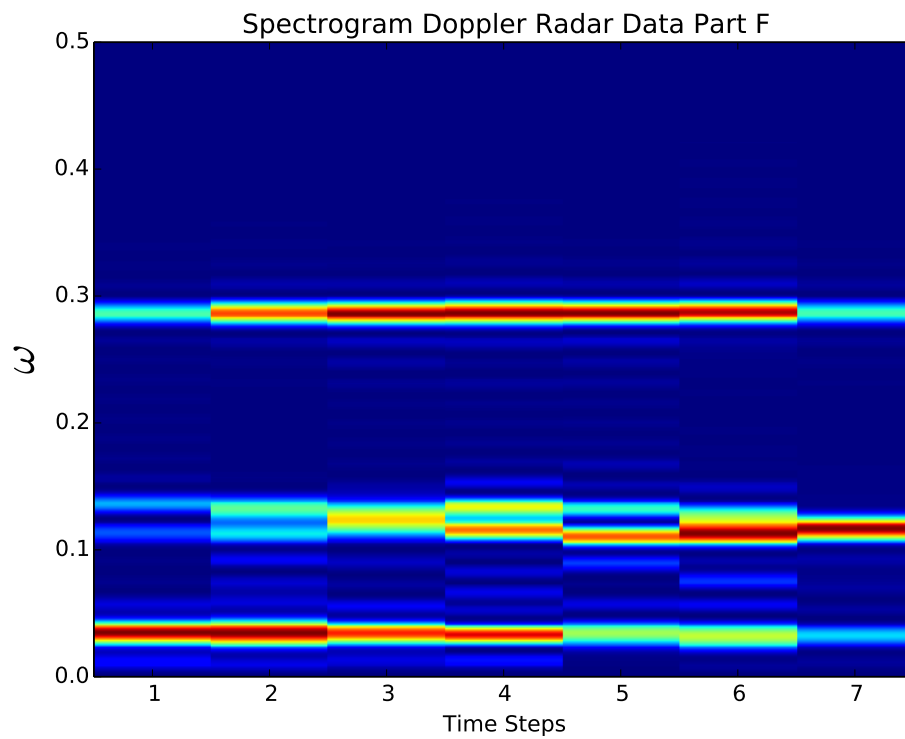


Figure D.34: *Segment F Example 2: Spectrogram of Segment F containing a signal around $\omega = 0.29$, a chirped signal around $\omega = 0.12$ and a chirped signal around $\omega = 0.03$.*

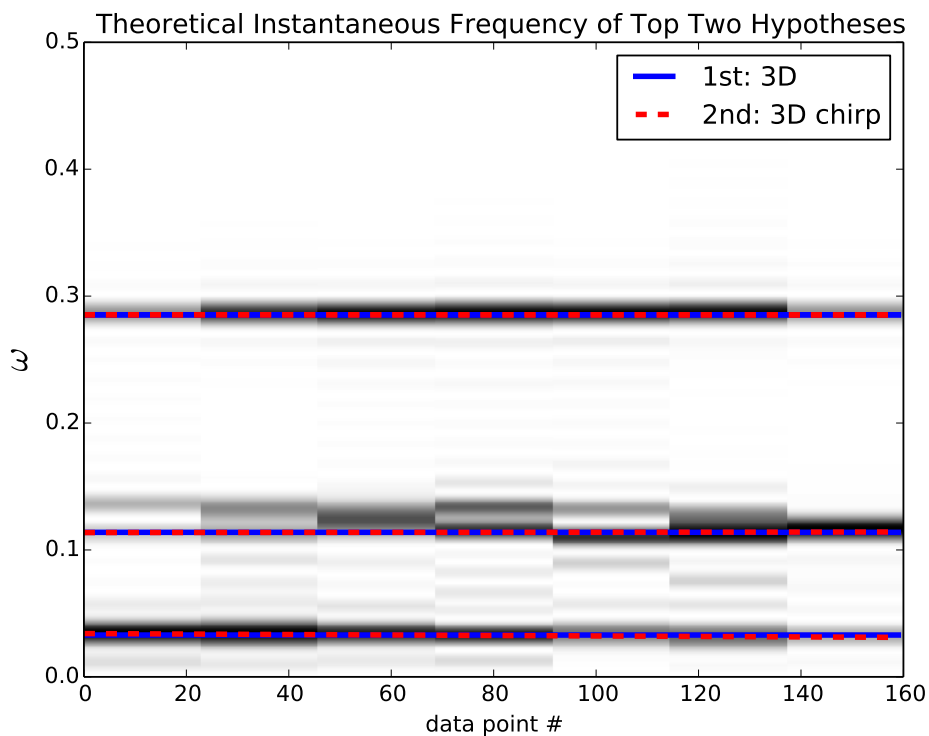


Figure D.35: *Segment F Example 2: Top two BSA hypotheses' instantaneous frequency super-imposed on the spectrogram.*

Table D.11: *Segment F Example 2: Estimated parameters and model selection of three simulated chirped sinusoids. The most likely model is indicated in blue.*

	1 Stationary	1 Chirp	1 Qua Chirp
Probability	4.51410e-22	1.23416e-25	1.84175e-31
ω_0	3.28959e-02	3.26290e-02	3.27703e-02
α_0	-	-1.17807e-04	-1.19950e-04
β_0	-	-	-2.15838e-07
ω_1	-	-	-
α_1	-	-	-
β_1	-	-	-
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	2 Stationary	2 Chirp	2 Qua Chirp
Probability	3.99187e-13	9.62903e-21	1.84500e-32
ω_0	3.28742e-02	3.26197e-02	3.28542e-02
α_0	-	-1.12511e-04	-1.16839e-04
β_0	-	-	-3.50862e-07
ω_1	2.85265e-01	2.85254e-01	2.85604e-01
α_1	-	4.23973e-06	-7.55904e-08
β_1	-	-	-5.72819e-07
ω_2	-	-	-
α_2	-	-	-
β_2	-	-	-
	3 Stationary	3 Chirp	3 Qua Chirp
Probability	1.00000e+00	3.55393e-12	9.41303e-30
ω_0	1.13881e-01	3.26251e-02	3.27881e-02
α_0	-	-1.18036e-04	-1.25503e-04
β_0	-	-	-2.53818e-07
ω_1	3.29699e-02	1.13927e-01	2.85606e-01
α_1	-	1.64782e-05	-6.56369e-06
β_1	-	-	-5.29721e-07
ω_2	2.85268e-01	2.85250e-01	1.13087e-01
α_2	-	-5.64955e-06	2.19024e-05
β_2	-	-	1.35435e-06

Appendix E

Explicit Duration Model

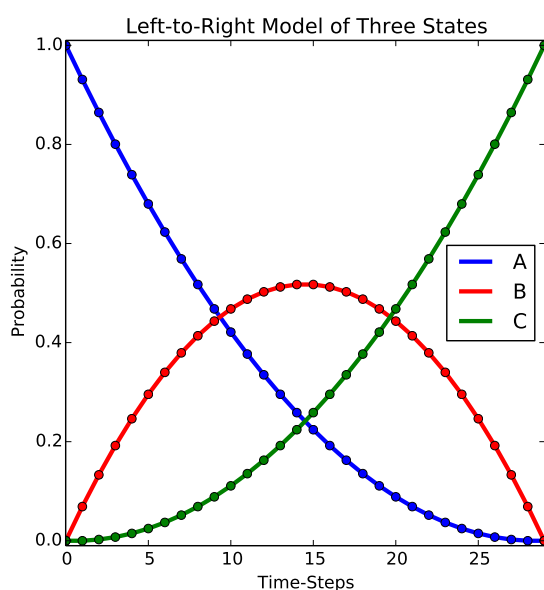


Figure E.1: *Left-to-right model prior probabilities.*

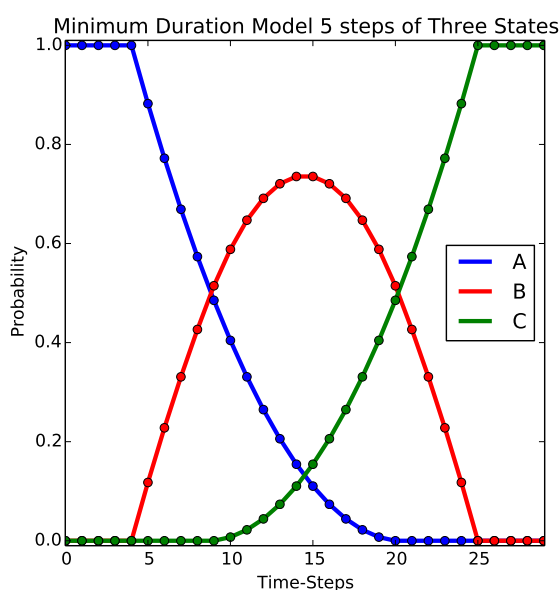


Figure E.2: *Minimum Duration model prior probabilities.*

E.1 Explicit Duration

Explicit duration models are also employed in HMMs such as the Ferguson model from [29]. Here one can have multiple states that produce the same output and are conceptually the sub-states of the same state. The transition probabilities can be fine-tuned to give one an arbitrary duration model for each state depending on the number of sub-states and their probabilities.

Figure E.3 is an example of an explicit duration model from [29]. The transition probabilities can be tuned to create arbitrary prior state probabilities.

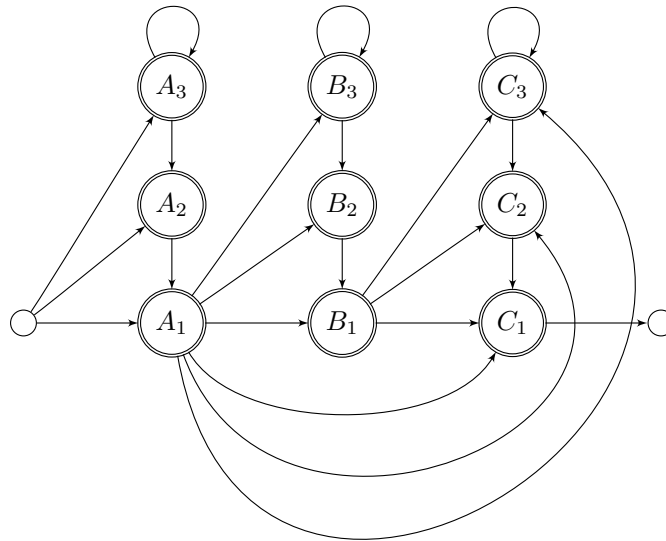


Figure E.3: *Ferguson Explicit Duration Markov chain*

E.2 Minimum Duration

A minimum duration model is a subset of the explicit duration models that forces the model to stay within a certain state for a fixed number of time-steps before being able to transition to the next time-step [29]. The model must step through each sub-state before reaching the next state. This is particularly useful as it disallows the model from staying in a certain state for only one time-step and moving to the next state too soon. Figure E.4 shows how one can set up the transitions for the minimum duration model. The model behaves just like the left-to-right model but needs to pass through all the sub-states before reaching the next state. Figure E.2 shows how the model is forced to stay in each state for at least five time-steps before moving to the next state. Contrast this with Figure E.1 where the model can immediately start to transition to the next state.

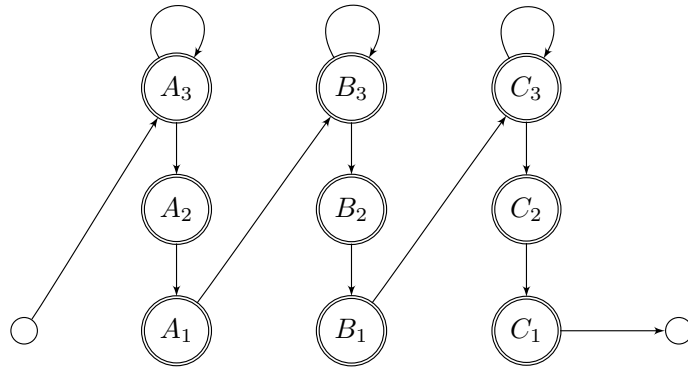


Figure E.4: *Ferguson Minimum Duration Markov chain*

Appendix F

Factor Details

F.1 Linear Transitions

In this section we list the possible linear relationships between our frequency components and their dynamics for different BSA model transitions. We only demonstrate for one- and two-frequency models, however, the same principles extend to the three-frequency model.

Single-frequency Model to single-frequency Model

$$\begin{bmatrix} \omega_{t+1} \\ \Delta\omega_{t+1} \\ a_{t+1} \\ \Delta a_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_t \\ \Delta\omega_t \\ a_t \\ \Delta a_t \end{bmatrix} + \mathbf{w}.$$

Two-frequency Model to Two-frequency Model

$$\begin{bmatrix} \omega_{t+1,0} \\ \Delta\omega_{t+1,0} \\ a_{t+1,0} \\ \Delta a_{t+1,0} \\ \omega_{t+1,1} \\ \Delta\omega_{t+1,1} \\ a_{t+1,1} \\ \Delta a_{t+1,1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{t,0} \\ \Delta\omega_{t,0} \\ a_{t,0} \\ \Delta a_{t,0} \\ \omega_{t,1} \\ \Delta\omega_{t,1} \\ a_{t,1} \\ \Delta a_{t,1} \end{bmatrix} + \mathbf{w}.$$

Two-frequency Model to One-frequency Model Top Frequency Survives

$$\begin{bmatrix} \omega_{t+1,0} \\ \Delta\omega_{t+1,0} \\ a_{t+1,0} \\ \Delta a_{t+1,0} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \omega_{t,0} \\ \Delta\omega_{t,0} \\ a_{t,0} \\ \Delta a_{t,0} \\ \omega_{t,1} \\ \Delta\omega_{t,1} \\ a_{t,1} \\ \Delta a_{t,1} \end{bmatrix} + \mathbf{w}.$$

Two-frequency Model to One-frequency Model Bottom Frequency Survives

$$\begin{bmatrix} \omega_{t+1,0} \\ \Delta\omega_{t+1,0} \\ a_{t+1,0} \\ \Delta a_{t+1,0} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \omega_{t,0} \\ \Delta\omega_{t,0} \\ a_{t,0} \\ \Delta a_{t,0} \\ \omega_{t,1} \\ \Delta\omega_{t,1} \\ a_{t,1} \\ \Delta a_{t,1} \end{bmatrix} + \mathbf{w}.$$

One-frequency Model to Two-frequency Model Multiply a one-frequency model to one-frequency model transition with a prior for the new frequency component.

F.2 Discrete Factor Examples

In Table 7.4 and F.2 we demonstrate how we calculate the discrete factors. The random variables on the right of the conditioning bar (denoted with the double line), determine the possible values of the random variables on the left of the conditioning bar for the given model transition. We only demonstrate $p(\omega_{t+1}|\omega_t, \Delta\omega_{t+1})$ and $p(\Delta\omega_{t+1}|\Delta\omega_t, \mathbf{a}_{t+1})$. However, the other transition factors have the same premise; we exhaustively list all possible values while making use of place-holder values and then normalise the factor as a CPD. The use of the place-holder values also ensure that the transitions are unique for each model transition.

Table F.1: This table describes how we calculate the possible values for the discrete factors of the form $p(\omega_{t+1}|\omega_t, \Delta\omega_{t+1})$. Note that there are two scenarios each for switching between the signal models; one involving the top frequency component and one the bottom.

Model Transition	$\omega_{t+1,0}$	$\omega_{t+1,1}$	$\omega_{t,0}$	$\omega_{t,1}$	$\Delta\omega_{t+1,0}$	$\Delta\omega_{t+1,1}$
1 freq to 1	ϕ	$\omega_{t,1} + \Delta\omega_{t,1}$	ϕ	$\omega_{t,1}$	ϕ	$\Delta\omega_{t,1}$
2 freq to 2	$\omega_{t,0} + \Delta\omega_{t,0}$	$\omega_{t,1} + \Delta\omega_{t,1}$	$\omega_{t,0}$	$\omega_{t,1}$	$\Delta\omega_{t,0}$	$\Delta\omega_{t,1}$
1 freq to 2 \mathfrak{A}	$\omega_{t+1,0}$	$\omega_{t,1} + \Delta\omega_{t,1}$	ϕ	$\omega_{t,1}$	\mathfrak{A}	$\Delta\omega_{t,1}$
1 freq to 2 \mathfrak{B}	$\omega_{t,1} + \Delta\omega_{t,1}$	$\omega_{t+1,1}$	ϕ	$\omega_{t,1}$	$\Delta\omega_{t,0}$	\mathfrak{B}
2 freq to 1 \mathfrak{C}	ϕ	$\omega_{t,1} + \Delta\omega_{t,1}$	$\omega_{t,0}$	$\omega_{t,1}$	\mathfrak{C}	$\Delta\omega_{t,1}$
2 freq to 1 \mathfrak{D}	ϕ	$\omega_{t,0} + \Delta\omega_{t,0}$	$\omega_{t,0}$	$\omega_{t,1}$	$\Delta\omega_{t,0}$	\mathfrak{D}

Table F.2: This table describes how we calculate the possible values for the discrete factors of the form $p(\Delta\omega_{t+1}|\Delta\omega_t, \mathbf{a}_{t+1})$. Note that it takes two time-steps for the dynamics to switch between different signal models.

Model Transition	$\Delta\omega_{t+1,0}$	$\Delta\omega_{t+1,1}$	$\Delta\omega_{t,0}$	$\Delta\omega_{t,1}$	$a_{t+1,0}$	$a_{t+1,1}$
1 freq to 1	ϕ	$\Delta\omega_{t,1} + a_{t,1}$	ϕ	$\Delta\omega_{t,1}$	ϕ	$a_{t,1}$
2 freq to 2	$\Delta\omega_{t,0} + a_{t,0}$	$\Delta\omega_{t,1} + a_{t,1}$	$\Delta\omega_{t,0}$	$\Delta\omega_{t,1}$	$a_{t,0}$	$a_{t,1}$
1 freq to 2 \mathfrak{A}	\mathfrak{A}	$\Delta\omega_{t,1} + a_{t,1}$	ϕ	$\Delta\omega_{t,1}$	\mathfrak{A}	$a_{t,1}$
1 freq to 2 \mathfrak{a}	$\Delta\omega_{t+1,0}$	$\Delta\omega_{t,1} + a_{t,1}$	\mathfrak{A}	$\Delta\omega_{t,1}$	\mathfrak{a}	$a_{t,1}$
1 freq to 2 \mathfrak{B}	$\Delta\omega_{t,1} + a_{t,0}$	\mathfrak{B}	ϕ	$\Delta\omega_{t,1}$	$a_{t,0}$	\mathfrak{B}
1 freq to 2 \mathfrak{b}	$\Delta\omega_{t,0} + a_{t,0}$	$\Delta\omega_{t+1,1}$	$\Delta\omega_{t,0}$	\mathfrak{B}	$a_{t,0}$	\mathfrak{b}
2 freq to 1 \mathfrak{C}	\mathfrak{C}	$\Delta\omega_{t,1} + a_{t,1}$	$\Delta\omega_{t,0}$	$\Delta\omega_{t,1}$	\mathfrak{C}	$a_{t,1}$
2 freq to 1 \mathfrak{c}	ϕ	$\Delta\omega_{t,1} + a_{t,1}$	\mathfrak{C}	$\Delta\omega_{t,1}$	\mathfrak{c}	$a_{t,1}$
2 freq to 1 \mathfrak{D}	$\Delta\omega_{t,0} + a_{t,0}$	\mathfrak{D}	$\Delta\omega_{t,0}$	$\Delta\omega_{t,1}$	$a_{t,0}$	\mathfrak{D}
2 freq to 1 \mathfrak{d}	$\Delta\omega_{t,0} + a_{t,0}$	ϕ	$\Delta\omega_{t,0}$	\mathfrak{D}	$a_{t,0}$	\mathfrak{d}

List of References

- [1] M. Mallick, B.-N. Vo, T. Kirubarajan, and S. Arulampalam, "Introduction to the issue on multitarget tracking," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 7, no. 3, pp. 373–375, 2013.
- [2] A.-r. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 1, pp. 14–22, 2012.
- [3] P. C. Gregory and T. J. Lored, "Bayesian periodic signal detection : Analysis of rosat observations of PSR 0540-693," *The Astrophysical Journal*, vol. 473, no. 2, pp. 1059–1066, 1996.
- [4] P. Gregory, "A. bayesian revolution in spectral analysis," in *AIP conference proceedings*. IOP INSTITUTE OF PHYSICS PUBLISHING LTD, 2001, pp. 557–568.
- [5] G. L. Bretthorst, "An introduction to model selection using probability theory as logic," in *Maximum entropy and Bayesian methods*. Springer, 1996, pp. 1–42.
- [6] E. Granqvist, G. E. Oldroyd, and R. J. Morris, "Automated bayesian model development for frequency detection in biological time series," *BMC systems biology*, vol. 5, no. 1, p. 1, 2011.
- [7] E. T. Jaynes, "Bayesian spectrum and chirp analysis," in *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems*. Springer, 1987, pp. 1–37.
- [8] G. L. Bretthorst, "Bayesian spectrum analysis and parameter estimation," *Lecture Notes in Statistics*, vol. 48, 1988.
- [9] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [10] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Fluids Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

- [11] L. D. Stone, C. A. Barlow, and T. L. Corwin, *Bayesian Multiple Target Tracking*. Artech House, 1999.
- [12] A. Frank, P. Smyth, and A. Ihler, “A graphical model representation of the track-oriented multiple hypothesis tracker,” in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*. IEEE, 2012, pp. 768–771.
- [13] J. L. Williams and R. A. Lau, “Data association by loopy belief propagation,” in *Information Fusion (FUSION), 2010 13th Conference on*, 2010, pp. 1–8.
- [14] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [15] F. V. Jensen, *An introduction to Bayesian networks*. UCL press London, 1996, vol. 210.
- [16] R. Kindermann, J. L. Snell *et al.*, *Markov random fields and their applications*. American Mathematical Society Providence, RI, 1980, vol. 1.
- [17] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 498–519, 2001.
- [18] R. Shumway and D. Stoffer, “Dynamic linear models with switching,” *Journal of the American Statistical Association*, vol. 86, no. 415, pp. 763–769, 1991.
- [19] C. Kreucher, K. Kastella, and A. O. Hero, “Multitarget tracking using the joint multitarget probability density,” *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 41, no. 4, pp. 1396–1414, 2005.
- [20] M. A. Richard, J. A. Scheer, and A. William, “Principles of modern radar: Basic principles, the new face of radar 101—modern and comprehensive,” 2010.
- [21] G. L. Bretthorst, “Bayesian analysis. I. parameter estimation using quadrature NMR models,” *Journal of Magnetic Resonance (1969)*, vol. 88, no. 3, pp. 533–551, 1990.
- [22] ———, “Bayesian analysis. II. signal detection and model selection,” *Journal of Magnetic Resonance (1969)*, vol. 88, no. 3, pp. 552–570, 1990.
- [23] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge University Press, 2005.

- [24] R. A. Fisher, “On the mathematical foundations of theoretical statistics,” *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 222, pp. 309–368, 1922.
- [25] D. J. MacKay, *Information theory, inference and learning algorithms*. Cambridge university press, 2003.
- [26] J. A. Nelder and R. Mead, “A simplex method for function minimization,” *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.
- [27] Z. Chen, “Bayesian filtering: From kalman filters to particle filters, and beyond,” *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.
- [28] Z. Ghahramani, “An introduction to hidden markov models and bayesian networks,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 15, no. 01, pp. 9–42, 2001.
- [29] J. D. Ferguson, “Variable duration models for speech,” in *Proceedings of the Symposium on the Application of Hidden Markov Models to Text and Speech*, pp. 143 – 179, 1980.
- [30] P. Del Moral, “Non-linear filtering: interacting particle resolution,” *Markov processes and related fields*, vol. 2, no. 4, pp. 555–581, 1996.
- [31] D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.
- [32] G. Heitz, *Graphical models for high-level computer vision*. ProQuest, 2009.
- [33] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988.
- [34] S. L. Lauritzen and D. J. Spiegelhalter, “Local computations with probabilities on graphical structures and their application to expert systems,” *Journal of the Royal Statistical Society*, vol. 50, no. 2, pp. 157–224, 1988.
- [35] G. Elidan, I. McGraw, and D. Koller, “Residual belief propagation: Informed scheduling for asynchronous message passing,” in *Proceedings of the Twenty-second Conference on Uncertainty in AI (UAI)*, 2006.
- [36] J. A. Du Preez, private communication, 2015.
- [37] T. B. Schön and F. Lindsten, “Manipulating the multivariate gaussian density,” Technical report, Linköping University, 2011., Tech. Rep., 2011.

- [38] W. Sun, “Implementation of clique tree algorithm for conditional linear gaussian model,” 2009.
- [39] I. Csiszàr, “I-divergence geometry of probability distributions and minimization problems.” *The Annals of Probability*, vol. 3, no. 1, pp. 146–158, 1975.
- [40] T. P. Minka, “Expectation propagation for approximate bayesian inference,” in *17th Conference on Uncertainty in Artificial Intelligence*, 2001, pp. 362–369.
- [41] D. Simon, “Kalman filtering,” *Embedded systems programming*, vol. 14, no. 6, pp. 72–79, 2001.
- [42] L. E. Baum and T. Petrie, “Statistical inference for probabilistic functions of finite state markov chains,” *Ann. Math. Statist.*, vol. 37, no. 6, pp. 1554–1563, 12 1966. [Online]. Available: <http://dx.doi.org/10.1214/aoms/1177699147>
- [43] D. Ceperley, Y. Chen, R. V. Craiu, X.-L. Meng, A. Mira, and J. Rosenthal, “Challenges and advances in high dimensional and high complexity monte carlo computation and theory,” in *Proceedings of the Workshop at the Banff International Research Station for Mathematical Innovation Discovery*, 2012.
- [44] R. van der Merwe and E. Wan, “Gaussian mixture sigma-point particle filters for sequential probabilistic inference in dynamic state-space models,” in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, vol. 6. IEEE, 2003, pp. VI–701.
- [45] E. Jones, T. Oliphant, P. Peterson *et al.*, “SciPy: Open source scientific tools for Python,” 2001–, [Online; accessed 2016-04-23]. [Online]. Available: <http://www.scipy.org/>