

The London School of Economics and Political Science



Limitations to Seasonal Weather Prediction
and Crop Forecasting due to Nonlinearity
and Model Inadequacy

S M W Higgins

A thesis submitted to the Department of Statistics of the London
School of Economics for the degree of Doctor of Philosophy.

London, September 2015

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

Sarah M. W. Higgins

London, September 2015

Word Count: 25,650

Acknowledgments

I would like to thank my supervisor Professor Leonard A. Smith for his constant and unwavering support and mentoring during my long thesis. I am very grateful to wise advice provided by Professor Henry Wynn and the valuable insights provided by Dr Joshua W. Elliott about the CERES-Maize model and its inputs.

In addition I would like to thank the cast of CATS, past and present, for their encouragement and thought provoking conversations especially Roman Binter, Daniel Bruynooghe, Liam Clarke, Lyn Grove, Alex Jarman, Falk Niehörster, Emma Suckling and Erica Thomson with a special nod to the insanity that is Hailiang Du.

Finally I would like to extend my deepest gratitude to my wonderful husband Mark and our children Lara and Jamie without whose love, support and understanding this thesis would never have been completed.

Nonlinear aspects of modelling crop yield on seasonal timescales

Sarah M W Higgins

Abstract

This Thesis examines the main issues surrounding crop modelling by detailed studies of (i) multi-model ensemble forecasting using a simple dynamical system as a proxy for seasonal weather forecasting, (ii) probabilistic forecasts for crop models and (iii) an analysis of changes in US yield. The ability to forecast crop yield accurately on a seasonal time frame would be hugely beneficial to society in particular farmers, governments and the insurance industry. In addition, advance warning of severe weather patterns that could devastate large areas of crops would allow contingency plans to be put in place before the onset of a widespread famine, potentially averting a humanitarian disaster.

There is little experience in the experimental design of ensembles for seasonal weather forecasting. Exploring the stability of the results varying, for example, the sample size aids understanding. For this a series of numerical experiments are conducted in an idealised world based around the Moran Ricker Map. The idealised world is designed to replicate the multi-model ensemble forecasting methods used in seasonal weather forecasting. Given

the complexity of the physical weather systems experiments are instead conducted on the Moran Ricker Map [56,70]. Additionally, experiments examine whether including climatology as a separate model or blending with climatology can increase the skill.

A method to create probabilistic forecasts from a crop model, the Crop Environment Resource Synthesis Maize model (CERES-Maize) [19,37] is proposed. New empirical models are created using historical US maize yield. The skill from equally weighting the crop model with a simple empirical model is investigated.

Background reviews of weather and yield data is presented in new ways for the largest maize growing state Iowa. A new method separating the impacts of favourable weather from technology increases in a crop yield time series is explored.

List of Figures	13
List of Tables	29
1 Introduction to dynamical systems and crop modelling	32
1.1 Introduction	32
1.1.1 Different crop model types	35
1.2 Maize yield time series	35
1.3 Meteorological observations	38
1.4 Gridded meteorological data sets	40
1.5 The crop model	41
1.6 Primary weaknesses in CERES-Maize model	45
1.7 The impact of technology advancements on the yield time series	46
1.8 A chaotic system: Moran Ricker Map	48
1.9 Ensemble forecasting	52
1.10 Converting an ensemble into a probabilistic forecast	54
1.11 Measuring skill from a probabilistic forecast	56
1.12 Seasonal weather models	58
1.13 Overview of introduction	60
1.14 Contributions	60
2 Illustrating challenges to multi-model forecasting when data are precious	63
2.1 Introduction	63
2.2 Models for the non-linear dynamical system	67
2.2.1 Initial conditions	71
2.2.2 Creating the ensemble	72

2.2.3	Creating an idealised forecast-outcome archive	77
2.2.4	How to form a naïve probability distribution using priors	77
2.3	Illustrating the restrictions of a small forecast - outcome archive	80
2.4	How to combine multiple models	94
2.5	Skill from equally weighted models	94
2.6	Skill from blending models with climatology	98
2.6.1	Blending parameters at different steps	101
2.7	Conclusions	104
3	Crop Modelling	107
3.1	Creating a probabilistic forecast from a crop model	109
3.2	Crop model skill	112
3.3	Creating a bench-mark model for crop modelling	116
3.3.1	Persistence Model	117
3.3.2	Skill of Crop Models	120
3.4	Empirical Crop Models	120
3.4.1	The Dynamic Climatology Model	120
3.4.2	Ratio Model	127
3.4.3	Asymmetric Model	129
3.4.4	Gamma Model	137
3.5	Estimating the kernel width	144
3.6	Kernel dressing methods and skill	146
3.7	Multi-model crop forecasts	148
3.8	Multi-model forecasting at US state level	152
3.9	Conclusions	160
4	Analysis of weather and yield	163

4.1	Meteorological observations from Iowa	165
4.1.1	Location of weather stations in Iowa	165
4.1.2	Maximum temperature	168
4.1.3	Quality Control for Meteorological Observations . . .	171
4.1.4	Precipitation	172
4.2	Causes of observational uncertainties	175
4.3	Interaction of meteorological observations and crop yield . . .	179
4.4	Maize yield by county in Iowa	183
4.4.1	How widespread is crop failure?	185
4.4.2	Maize yield in adverse weather years	189
4.4.3	Yield from irrigated fields	193
4.5	Introducing the prior return to model technology advance- ment in maize yield	196
4.5.1	State level	199
4.5.2	County Level	201
4.6	Conclusions	206
5	Creating initial conditions for gridded weather	209
5.1	Uncertainty in gridded minimum temperature	210
5.2	Uncertainty in gridded maximum temperature	215
5.3	Uncertainty in gridded precipitation	217
5.3.1	Estimating the probability for a dry or wet day	218
5.3.2	Estimating nine equally likely bins	218
5.3.3	Estimating the exponential distribution	221
5.4	Conclusions	226
6	Summary	227

A	Gridded Meteorological data sets	232
A.1	Gridded Precipitation	232
A.2	Gridded Temperature	233
A.3	Gridded solar radiation	233
B	Imperfect Models for the Moran Ricker Map	234
B.1	Model MR12: Taylor series expansion of the Moran Ricker Map	235
B.2	MRLM: Taylor series expansion of the log term	238
B.3	Model MRFT: A Fourier transform approx of the Moran Ricker Map	240
C	Figures for Crop Modelling	244
D	Summary of state codes for weather stations	251
D.0.1	Minimum temperature observations	253
D.0.2	Snowfall observations	255
D.1	Flags for Meteorological Observations	255
E	Figures for creating initial conditions for gridded weather	259
	Bibliography	267

List of Figures

- 1.1 The US maize yield time series. The yield has significantly increased between the 1870s and 2012. 36

- 1.2 To calibrate the model linear regressions are fitted to the uncalibrated yield estimates (red dashed line) and yield outcomes (blue line) for a county in Iowa. 44

- 1.3 Moran Ricker Map, the image of \tilde{x}_0 when $\alpha = 3$. Note that it is bounded between 0 and 2.46. 49

- 1.4 Moran Ricker bifurcation diagram. Notice the fixed points at 0 and 1 and that when $\alpha = 3$ the Moran Ricker Map is chaotic and bounded. 50

- 1.5 An illustration of how standard kernel dressing converts a nine member ensemble into a forecast distribution. Note the non Gaussian shape of the forecast distribution. 55

- 2.1 Overview of experimental design. Top: multi-model ensemble forecasts used for seasonal weather forecasting. The system is the weather, modelled by seasonal weather models using ensemble forecasting. Skill from (1) is measured against the skill from the bench mark model, climatology (2). Bottom: the idealised world using the Moran Ricker Map as the system. Models which approximate the Moran Ricker Map are used to create an ensemble forecast. Skill from (1) is measured against the bench mark model, climatology (2). 68
- 2.2 Figures (a), (c) and (e) show how close the models are to the system when integrated forward one step. Each model deviates from the system at different locations. For MR12 (b) the large difference is when \tilde{x}_0 approaches the maximum value. For MRLM (d) the large difference is when \tilde{x}_0 is close to 0. For MRFT (f) the differences oscillate across the entire \tilde{x}_0 range but are largest for \tilde{x}_0 close to 0. 70
- 2.3 How the value of κ (left circle) determines the spread of the ensemble (right circle) at iteration two for a two dimensional system. 72
- 2.4 Creating realistic ensemble members: To generate S_{attr} \tilde{x}_0 is integrated forward 1,000 times through the system so that the initial conditions lie close to the system's attractor. State space is shown in green, model spaces are shown in blue. The uncertainty interval width (κ), shows the maximum perturbation added to the initial condition to create an ensemble of initial conditions. κ is set when the ensemble spread equals the mean square error of the model estimate at step two. . . . 74
- 2.5 κ is set where the mean square error for the model image (blue line) crosses the ensemble spread (blue dashed line) for all models integrated forward two steps. Note that κ is widest for MR12. 76

2.6	Creating ensembles and outcomes: S_{attr} is divided into two sets; one set is used to train parameters (S_{train}) and one set is used to verify the forecast (S_{test}). For each point in S_{train} , N_e ensemble members are selected. These ensembles are integrated forward five steps. Outcomes are from S_{test} integrated forward five steps.	78
2.7	Moran Ricker outcomes (Y_{dim}).	80
2.8	Skill by model and sample size. Note that the confidence intervals do not decrease as the sample size increases.	87
2.9	Histograms of skill by sample size for MR12. For clarity (e) and (f) have different y axes. Note with larger sets it is more likely that the sample mean will be closer to the actual mean.	88
2.10	Skill by model and sample size and 5% and 95% bootstrap resampling intervals. Note that the order of which model has the most skill remains consistent across all the sample sizes.	89
2.11	Comparing kernel dressing parameters of different sized forecast-outcome archives with values estimated from a large archive (S_{arch}) at step three. σ is on the left and u is on the right. Note that the σ and u values tend towards the green line as the sample size increases.	90
2.12	The variability of σ from 1000 sets of S_{sample} when integrated forward five steps. σ from S_{arch} is shown as dashed black lines. MR12 has the widest variability of σ estimates.	91
2.13	The variability of u from 1000 S_{sample} when integrated forward for five steps and two steps. S_{arch} values of u are shown as dashed black lines. The spread of u estimates is wider at step 5 than at step 2.	92
2.14	Histograms comparing the skill from a small archive. Negative values mean the skill of the small archive has been overestimated. Skill is relative to climatology. When integrated forward five steps most of the skill has been overestimated.	93

2.15	Comparison of skill relative to climatology for equally weighted multi-model forecasts (a) without climatology and (b) with climatology at step five. Note how the variation in relative Ignorance scores reduces in (b), as there are no forecast busts (red crosses).	97
2.16	Comparison of skill for equally weighted multi-model forecasts when the models are individually blended with climatology beforehand using a data set of 2000 points at step five. Skill is measured relative to climatology. Note that there are less points above 0 relative Ignorance.	101
2.17	Comparison of skill relative to climatology for equally weighted blended models at (a) step 2, (b) step 3 (c) step 4 and (d) step 5. There were 1000 points in both the training and testing sets. Note that the skill for 0100 relative to the other models changes by step.	103
2.18	Blending parameter α as the number of steps integrated forward increase. As the number of steps increase, α for all the models approaches 0.	104
3.1	Time series of the CERES-Maize estimates (green) and outcomes (red). The CERES-Maize estimates are usually quite accurate at predicting steep falls in maize yield.	110
3.2	The forecast distribution for selected years from the CERES-Maize model. The forecast distribution is shown as a blue line, the CERES-Maize estimates are shown as green circles and the outcome is shown as a red star for the years 1984, 1988 and 1993. Notice how the outcome has a high probability mass for most years.	113

- 3.3 The skill of the CERES-Maize model. Empirical Ignorance is a green line, the standard deviation of the empirical Ignorance is a red dotted line and the mean empirical Ignorance is a red line. The empirical Ignorance all lie within the mean plus one standard deviations with the exception of years 1993 and 2009. 114
- 3.4 The robustness of the forecast distribution is demonstrated by comparing the empirical Ignorance by kernel width. Here four selected years for the CERES-Maize model are shown. The best kernel width changes from year to year, the kernel width for 1993 needs to be wider than the other years. 115
- 3.5 Time series of the persistence model estimate (green line) with one standard deviation of the forecast distribution (dotted green line) and outcomes (red line). By construction the persistence model misses every large fall in maize yield and the recovery the following year. 119
- 3.6 Ignorance by year from (a) the climatology model, (b) the persistence model and (c) the persistence model relative to climatology. Note the relative Ignorance shows that the persistence model has more skill in the latter half of the time series, with the exception of 2012, a drought year. 121
- 3.7 Skill of the CERES-Maize model relative to the persistence model. Negative values show the CERES-Maize model has more skill than the persistence model. 122
- 3.8 Time series of the dynamic climatology, ensemble (green dots) and outcomes (red line). The outcomes are within the ensemble with the exception of 1988 and 1994. 124
- 3.9 Skill of the dynamic climatology model, the ratio model and asymmetric model relative to the persistence model. The ratio model (b) has more skill than the other models. 126

3.10	Time series of the ratio model ensemble (green) and outcomes (red). The ensemble members are not evenly spread across the yield axis, there is a gap between the lowest ensemble members and the others.	127
3.11	Fig 3.11a: The forecast distribution of the asymmetric kernel (green) and sample (blue). Note the second smaller kernel for negative falls. Fig 3.11b. Notice the asymmetric kernel is wider at the bottom than the top.	133
3.12	The estimated probabilistic density function by year for the asymmetric model. The two separate Gaussians are clearly identified and there is an area of low probability between them.	135
3.13	The kernel width by year for the asymmetric model. Unless the chosen kernel width is too narrow, the empirical Ignorance is quite robust.	136
3.14	The skill of the asymmetric model when 3.14a σ_1 from the 1st mixed Gaussian is altered, 3.14b when σ_2 from the 2nd mixed Gaussian is altered and 3.14c the weight (ω) is altered. There are no sudden jumps in empirical Ignorance for the three kernel dressing parameters.	138
3.15	Value of AIC by C_0 . Note that when C_0 is 21 or 22 the AIC is lowest.	140
3.16	The Gamma model forecast (green) and the observations (red). In years with adverse weather such as a drought in 2012 the observation falls outside the standard deviation (dotted green line).	142
3.17	The estimated probabilistic density function by year for the gamma model. As the α parameter is high the kernel shape is quite similar to a Gaussian kernel.	143

- 3.18 The empirical Ignorance by kernel width is shown in blue. The kernel width, from the standard deviation of the differences (or errors), is in red. The kernel width from minimising Ignorance is in green. Minimising Ignorance has the narrowest kernels, and so the most skill, for all the empirical models. 145
- 3.19 Comparison of skill when the kernel widths are chosen by different methods. Ignorance and 5 to 95% bootstrap resampling intervals when kernel width is chosen using the drop one out method by minimising Ignorance (x -axis) and when the kernel width is chosen using drop one out errors (y -axis). Notice that CERES-Maize is the model with the most skill, irrelevant of how the kernel widths are chosen. 147
- 3.20 Comparison of model skill relative to persistence when the kernel widths are chosen by minimising ignorance. The dot is the mean relative Ignorance and the line is the 5% and 95% bootstrap resampling interval. The ratio model has significantly more skill than the other empirical models as the relative ignorance is lower. 149
- 3.21 Comparison of skill between a single model and two equally weighted models for the CERES-Maize, Ratio and Dynamic Climatology model. The top and bottom of the line represent the 5% and 95% bootstrap resampling interval. Notice how the most skillful model is the equally weighted CERES-Maize and asymmetric models in (a). 153
- 3.22 Comparison of skill between a single model and two equally weighted models for the asymmetric, persistence and gamma model. 154
- 3.23 CERES-Maize: the skill is relative to the CERES-Maize model. When CERES-Maize is equally weighted with the ratio model and the asymmetric model it has more skill than the CERES-Maize model 155

3.24	Fig 3.24a: Comparison of skill relative to persistence for the CERES-Maize model for the top eight maize producing states. The more skillful CERES-Maize models are the states of Illinois, Indiana and Ohio. Fig 3.24b: Comparison of the most skillful equally weighted crop models by state. For 7 out of 8 states the most skillful multi-model is CERES-Maize with the ratio model.	157
3.25	Comparison of the top three models relative to the CERES-Maize model. For Illinois and Indiana the CERES-Maize model provides the best forecast.	160
4.1	Location of USHCN weather stations in Iowa. Note that although the weather stations are evenly spread throughout Iowa, there is not a weather station in every county.	167
4.2	Monthly $\bar{t}_{max,m}$ in Iowa. On average higher $\bar{t}_{max,m}$ are during the months June, July and August.	168
4.3	Annual anomalies measured from the $\bar{t}_{max,ay}$ in Iowa. Note that 1981, 1987 and 1988 all have large positive anomalies, so are warmer than an average year.	169
4.4	Highest t_{max} in each year ($t1_{max}$) by weather station in Iowa. The black line is the mean ($\bar{t1}_{max}$) across the weather stations. The highest $\bar{t1}_{max}$ in Iowa was in 1988.	170
4.5	The mean of the highest t_{max} in the years 1970 to 2010 ($\bar{t1}_{max,ay}$) by county in Iowa. Counties located in the West and South have higher temperatures shown by the orange and yellow colours.	171
4.6	Monthly precipitation in Iowa. Note that on average more of the precipitation falls between May and August.	175

4.7	Annual precipitation by year and selected weather stations in Iowa. The black line is the mean precipitation across all the weather stations. The years 1973 and 1993 had high precipitation.	176
4.8	Growing season precipitation by year and month in Iowa. The precipitation in July 1993 (light blue) was much more than July in the other years.	177
4.9	Map of the mean annual precipitation by county in Iowa. The counties in the North West of Iowa have lower annual precipitation than the other counties.	178
4.10	Weather Station 135796 Mount Pleasant Observation Station, Henry, Iowa * [21]. Meteorological observations will be hindered by the overgrown vegetation which blocks precipitation and sunlight.	179
4.11	Weather Station 137147 Rock Rapids Observation Station, Lyon, Iowa * [21]. Meteorological observations will be affected by the surroundings.	180
4.12	Comparison of the size of the residuals, the highest annual daily t_{max} and precipitation for Iowa. The highest values for residuals (red stars) are mainly found in years where the maximum temperatures and precipitation are average.	182
4.13	Comparison of the number of days over 29°C and the yield residual for Iowa. Note the low residual yield for 1993 when the wet weather damaged crops. A linear regression, red dashed line, was fitted to the data (excluding 1993) as the number of days above 29°C increases the yield falls. See text for discussion of 1993.	184
4.14	Yearly maize yield for six counties in Iowa. For most counties there is a large rise in yield between 1970 and 2012.	185

List of Figures

4.15 Histogram of maize yield in 4.15a shows for most counties yield is towards the upper end of the range. The average standard deviation of maize yield in 4.15b is approximately 32. The slope histogram 4.15c shows most slope parameters are approximately 2.	186
4.16 Mean yield by county in Iowa. Southern counties have lower mean yields, coloured blue.	187
4.17 Mean standard deviation of maize yield by county in Iowa. The counties in the South West of Iowa have the highest standard deviations coloured orange.	187
4.18 Technology slope by county in Iowa. The counties in the North West have the higher slopes (orange) and the counties in the South have lower slopes (blue).	188
4.19 For each county in Iowa a blue dot marks the year with largest percentage fall in yield. The years with the largest percentage falls are all found in five years.	189
4.20 Percentage fall in yield by county for Iowa in 1974. The darker the blue, the larger the fall in yield here the largest falls are in the west of Iowa.	190
4.21 Percentage fall in yield by county for Iowa in 1977. Although this year had the largest fall in yield for any county it is more localised than the other years.	190
4.22 Percentage fall in yield by county for Iowa in 1983. The largest falls in yield were for counties in the South East of Iowa. . . .	191
4.23 Percentage fall in yield by county for Iowa in 1988. The largest falls in yield were for counties in the East of Iowa.	191
4.24 Percentage fall in yield by county for Iowa in 1993. All the counties in Iowa have been significantly impacted.	192

- 4.25 Nebraska harvested area (acres) by year for irrigated (blue) and non irrigated (red). The irrigated harvested area has increased from 1946. The green line is the total area. Note the total area is wrong in the early years, this is an error in the NASS figures. 194
- 4.26 Nebraska percentage of harvested area irrigated (blue) and non irrigated (red). The percentage of irrigated area has slightly decreased in the last twenty years. 195
- 4.27 Weighted yield curve for Nebraska, total yield (green), irrigated yield (blue), non-irrigated yield (red). Most of the yield increase is from the irrigated lands. 195
- 4.28 The root mean square error between the fitted trend line with the “elbow” at each year and the observations by year for US maize yield time series. Notice how the bottom of the curve is quite flat. 197
- 4.29 Identifying the year (γ) when technical advances start. The RMSE between the best fitted trend line and data is at a minimum when $\gamma = 1943$. Note there is quite a large gap between the “elbow” of the fitted trend and the yield. . . . 198
- 4.30 Figure 4.30a: Comparison of the state yield (green line) with the maximum prior yield (red line) for Iowa. The blue line marks γ the start of the trend. Notice there is a steep rise in yield between 1961 and 1973. Figure 4.30b: A step function of the years where the prior yield (r_i) was greater than one for all the states in Iowa. Note that between 1961 and 1972 nearly every year had an increase in yield, which is unlikely to be caused just by favourable growing conditions. 200

4.31	Histogram for prior ratios greater than 1 for counties in Iowa by 4.31a early years and 4.31b later years. Notice how the later years have more prior ratios greater than 1 than the early years indicating that there was more technological advancements in the later years.	202
4.32	Iowa percentage of years where the ratio is greater than 1 (blue) and less than 1 (red) for the early years (left) and the later years (right). Note that there are significantly more ratios greater than 1 after 1943.	202
4.33	Comparison of the state level maximum prior yield (red line), the state level yield (green line) and the county level yield (green dots). There is significantly more variability in county level time series than the state level time series.	204
4.34	Prior ratios for all the counties in Iowa by year. The prior ratios less than one have a bigger downward movement than the ratios greater than one have upward movement.	205
4.35	Histogram of the prior ratios greater than one for all the counties in Iowa. Notice how most prior ratios are just greater than one.	206
4.36	Percentage of years where the prior ratio is greater than one (blue) and less than one (red) for all the counties in Iowa. Notice how this is almost identical for years both sides of the technology increase year.	207
4.37	These plots each show two things for the years 2012, 1987, 1962 and 1937 (1) What percentage of the total harvested area each county is to the state of Iowa which is recorded on the right y-axis in blue. (2) The cumulative of harvested area percentage by county is recorded on the left y-axis in grey. Across all the years the amount harvested in each county is fairly consistent	207

-
- 5.1 One grid and the lay out of the cells surrounding cell number 0210
- 5.2 Differences in gridded t_{min} between border cells and cell 0 for one grid in the centre of Iowa. The expected difference is zero. 212
- 5.3 μ for Iowa gridded t_{min} by month for each of the 16 areas. The μ are mainly clustered around 0. 213
- 5.4 σ for Iowa gridded t_{min} by month for each of the 16 areas. The monthly σ is spread out fairly evenly across all the months. . 214
- 5.5 Difference between the ensemble and t_{min} gridded data for the first six months (5.5a) and the last six months (5.5b). The differences centre around 0. 216
- 5.6 Setting precipitation for an ensemble member when cell 0 is dry. Either 0 or 1 is randomly drawn with probabilities conditional on cell 0 being dry. If 0 is selected the ensemble member has no precipitation. If 1 is selected precipitation needs to be calculated for the ensemble member. Historical precipitation is divided into 9 equally likely bins and a number between 1 and 9 is drawn. If the number is less than 9, precipitation is drawn from within that bin number. If the number is 9 precipitation is randomly drawn from the exponential distribution. If cell 0 is wet different probabilities need to be calculated. . 219
- 5.7 The probability of a) a dry day and b) a wet day, given that cell 0 is dry, by area and month. Note the probability for a) and b) are on different scales. The probability of a dry day, given that cell 0 is dry, is high across all the months and areas. 220
- 5.8 Equally likely bins for when cell 0 is dry (left) and when cell 0 is wet (right). Figures are (a) and (b) location of the edges of the first 8 bins and (c) and (d) the width of the last bin. Note that the box edge axis in (a) and (b) is different. When cell 0 is dry the first eight bins are very small as there is less rain in the surrounding cells. 222

5.9	Comparisons of observations in bin 9 for area 1 in January (5.9a) with randomly drawn numbers from the exponential distribution (5.9b). The lowest precipitation values randomly drawn from the truncated exponential distribution are less than the observations.	224
5.10	Jan to Jun: For Iowa the difference between the cells and the randomly drawn precipitation by month. The differences are all centred around 0.	225
B.1	The Moran Ricker Map when α is 3 is bounded between 0 and 2.46.	235
B.2	Figure B.2a: MR4 tends to infinity for large values of x_0 . Figure B.2b: when the n^{th} expansion is odd x_{i+1} is negative.	236
B.3	Comparison between the Moran Ricker Map and the models MR10 and MR12. Model MR12 provides more accurate results than model MR10 for higher values of x_0	237
B.4	MRLM4 at x_1 . Note this is close to the system except when x_0 is close to 0.	239
B.5	Comparison of x_{10} values of Moran Ricker Map B.5a with log models to the 4 th , 6 th and 8 th power. MRLM4 has more points at 0 than the system and MRLM8 has more points at the maximum value than the system.	240
B.6	MRFT4 at one iteration. This differs significantly from the system when x_0 is close to 0.	242
B.7	Comparison of the system and (a) MRFT10 and (b) MRFT12 at x_1 . Notice how the differences close to $x_0 = 0$ decrease for MRFT12.	243

- C.1 The forecast distribution produced from dressing estimate with a Gaussian kernel of width σ . The forecast distribution from the persistence model is shown as a blue line, the estimates are shown as green circles and the outcomes are shown as red stars for selected years 1984, 1988 and 2006. Notice although the outcome does not always lie in the high probability areas of the forecast distribution for years 1984 and 1988 it is captured by the forecast distribution. 245
- C.2 The robustness of forecast distribution is checked by comparing the ignorance by kernel width for four selected years for the persistence model. In 2006 the outcome fell in a very high probability of the forecast distribution, so the narrower the kernel the better. 246
- C.3 The forecast distribution for dynamic climatology. The forecast distribution from the dynamic climatology model is shown as a blue line, the ensemble of estimates are shown as green circles and the outcomes are shown as red stars for the years 1984, 1988 and 2006. The ensemble of estimates is so wide that the forecast distribution is very flat. 247
- C.4 Examining the ignorance by kernel width for selected years for the dynamic climatology model. The ignorance by kernel width for 1984, 1988 and 1993 have very similar shapes. . . . 248
- C.5 The forecast distribution produced from dressing the ensemble of estimates. The forecast distribution from the ratio model is shown as a blue line, the ensemble of estimates are shown as green circles and the outcomes are shown as red stars for the years 1984, 1988 and 2006. Notice how the outcome does not lie in the high probability area of the forecast distribution for 1984. 249

C.6	Examining the robustness of the kernel width by looking at the ignorance by kernel width for four selected years for the persistence model. The kernel width needs to be much wider for years 1988 and 1993	250
D.1	Monthly \bar{t}_{min} in Iowa. Note that on average the higher temperatures are between June and August.	254
D.2	Monthly anomalies for t_{min} (blue) and t_{max} (green) in Iowa. .	255
D.3	Monthly \bar{t}_{min} for July. Note that 2009 was colder than usual in July for most of the weather stations across Iowa.	256
D.4	Monthly snow in Iowa by year. On average more snow falls in the months of December and January.	257
D.5	Annual snowfall by year and month in Iowa. There is quite a lot of snow in April (light blue) for the years 1973, 1982, 1983 and 1997.	257
E.1	μ parameters for t_{max} by month for each of the 16 sections. The differences are all centred around 0.	260
E.2	σ parameters for t_{max} by month for each of the 16 sections. Notice how the parameters between t_{max} and t_{min} are similar. .	261
E.3	Comparison of gridded precipitation and a 9 member ensemble for gridded precipitation from July to December.	262

List of Tables

2.2	Radius (κ) for the uncertainty interval around the observations from within which the perturbations are randomly drawn.	75
2.4	Kernel dressing parameters and skill after integrating forward five steps for a large forecast-outcome archive. rIGN is empirical Ignorance relative to climatology and rIGN std is the sample standard deviation of relative Ignorance.	82
2.5	Kernel dressing parameters and empirical Ignorance relative to climatology after integrating forward five steps where the parameters were trained using leave-one-out cross-validation on S_{sample} , an archive of 22 points.	83
2.6	Comparison of empirical Ignorance relative to climatology (on S_{test}) when the models are integrated forward five steps and the kernel dressing parameters are trained using 1000 points (IGN _{Big}) and 22 points (IGN _{Small}).	84
2.7	Comparison of empirical Ignorance relative to climatology (on S_{test}) when the models are integrated forward three steps and the kernel dressing parameters are trained using 1000 points (IGN _{Big}) and 22 points (IGN _{Small}).	84
2.8	Skill of equally weighted multi-models (without climatology as a separate model) when integrated forward five steps. Skill is measured relative to climatology.	95

2.9	Skill of equally weighted models (including climatology as a separate model). Skill is measured relative to climatology. . .	96
2.10	Comparison of kernel dressing parameters and skill from blending models with climatology at five steps integrated forward (where N_{arch} is 2000.) Skill is measured relative to climatology. . .	99
2.11	Skill of equally weighted multi-models where the models were blended with climatology before being equally weighted together. Skill is measured relative to climatology.	100
2.12	The values of the blending parameter α for each step integrated forward. Notice which model has the highest blending parameter at each step.	102
3.2	Gaussian mixture parameters for the asymmetric model . . .	132
3.3	A summary of kernel width (σ) by model.	144
3.4	Skill relative to persistence for single models (1 model) and two equally weighted models.	150
3.5	Volume of maize produced by the top eight maize producing states in 2013	152
3.6	Skill of the CERES-Maize model relative to persistence for 8 states.	156
3.7	Model with the most skill relative to persistence by state. . .	159
4.1	The USHCN weather stations in Iowa, their county and the symbol used in this thesis to represent them.	166
4.3	Summary of Q Flags for Temperature Data	173
4.4	Flags from t_{max} observations in Iowa	174
5.1	Mean differences in t_{min} from border cells and cell 0 for one grid in Iowa.	211
D.2	Summary of M Flags for Temperature Data	258

D.3 Summary of S Flags for Temperature Data	258
---	-----

Chapter 1

Introduction to dynamical systems and crop modelling

1.1 Introduction

This thesis examines the main issues surrounding crop modelling by detailed studies of (i) multi-model ensemble forecasting using a simple dynamical system as a proxy for seasonal weather forecasting, (ii) probabilistic forecasts for crop models and (iii) an analysis of the changes in US yield. The ability to forecast crop yield accurately on a seasonal time frame would be hugely beneficial to society in particular farmers, governments and the insurance industry. It would also be of use in businesses such as transport, commodity trading and food services. In addition, advance warning of severe weather patterns that could devastate large areas of crops would allow contingency plans to be put in place before the onset of widespread famine, potentially

averting a humanitarian disaster.

The thesis is divided into four main sections. All underlined words are defined in the glossary. In Chapter 2 the experimental design of the state of the art forecasting methods, multi-model ensemble forecasting [31, 60], currently in use by seasonal weather is investigated. The experiments explore the limitations from working with a small forecast-outcome archive and how the skill of a probabilistic forecast is improved using climatology when a large forecast-outcome archive is available. In Chapter 3 probabilistic forecasts for crop modelling are considered using the Crop Environment Resource Synthesis Maize model (CERES-Maize) [19, 37]. A study examines if equally weighting the CERES-Maize model with an empirical model can increase the skill of these probabilistic forecasts. In crop modelling the focus is on maize in the US, where data is freely available. In Chapter 4 the meteorological observations, yield observations and technical advancements are analysed. Finally, in Chapter 5 how to capture the uncertainty in a crop forecast is considered by creating an ensemble of weather realisations. Currently the CERES-Maize model uses just one realisation of weather to make a forecast. In this chapter a nine member ensemble is generated to reflect the sampling uncertainty in the original realisation of weather. In future work this ensemble will be used to create a probabilistic forecast for yield for the CERES-Maize model.

In this chapter an outline is given for topics relevant to thesis. The different types of crop models available to forecast yield and the differences between each type of model are given in Section 1.1.1. The main causes of variability

in maize yield time series are discussed in Section 1.2. A review of the main meteorological observations used by crop models and how these irregularly spaced observations are converted into grids is given in Sections 1.3 and 1.4. The CERES-Maize model [19,37] is discussed in Sections 1.5 and 1.6. One of the problems facing crop modelling is the variability in yield, some of which is caused by weather and some by technical advancements. Section 1.7 provides a literature review on estimating the technical advancements in the crop yield.

Using simple chaotic systems as a starting point to understand more complex chaotic systems is a recognised and useful methodology [46,55]. In this thesis a simple dynamical system is used as a means to explore issues for a more complex dynamical system, seasonal weather. The dynamical system used is the Moran Ricker Map [56,70], as explained in Section 1.8.

The true state of the dynamical system for seasonal weather is obscured by observational uncertainty [41]. Additionally, there is uncertainty from model inadequacy in the forecast state. Every model is imperfect [17,26,93,96]. To provide a lower bound for these uncertainties, an ensemble of initial conditions is used with a collection of models [17,26,93,96]. Why ensemble forecasting is used for chaotic systems is discussed in Section 1.9. An important consideration there is how to convert an ensemble of estimates into a probability distribution function. This is discussed in Section 1.10 [5,79]. The skill is a measure of how accurate a probabilistic forecast is. A skill score $S(p(y), Y)$ evaluates the forecast distribution $p(y)$, against the outcome Y [4]. The skill score used here is empirical Ignorance [25,72] which

is explained in Section 1.11. Seasonal weather models and their multi-model forecast-outcome archives are outlined in Section 1.12.

1.1.1 Different crop model types

There are several distinct model types used to predict crop yields:

1. **Physical simulation models.** These models predict yield by replicating the physical conditions that surround the crops [11, 19, 37, 38]. For example, the weather state, soil type and details about farm management are a sample of the inputs required by a crop model.
2. **Indicator/Teleconnection models.** These models look at the relationships between weather patterns and crop yield to forecast future crop yield [9, 28]. For example, patterns in the El Niño-Southern Oscillation is used to predict crop yields in the United States (US) [28].
3. **Empirical models.** These models use historical crop yields to forecast future crop yields. Empirical models typically do not consider the underlying physical conditions which the crops to be forecast are growing in.

In the next Section the US maize yield times series is examined along with the main causes of variability within this time series.

1.2 Maize yield time series

The amount of maize grown each year in the US is measured in yield, where yield is the number of bushels per harvested acre and a bushel weighs 56

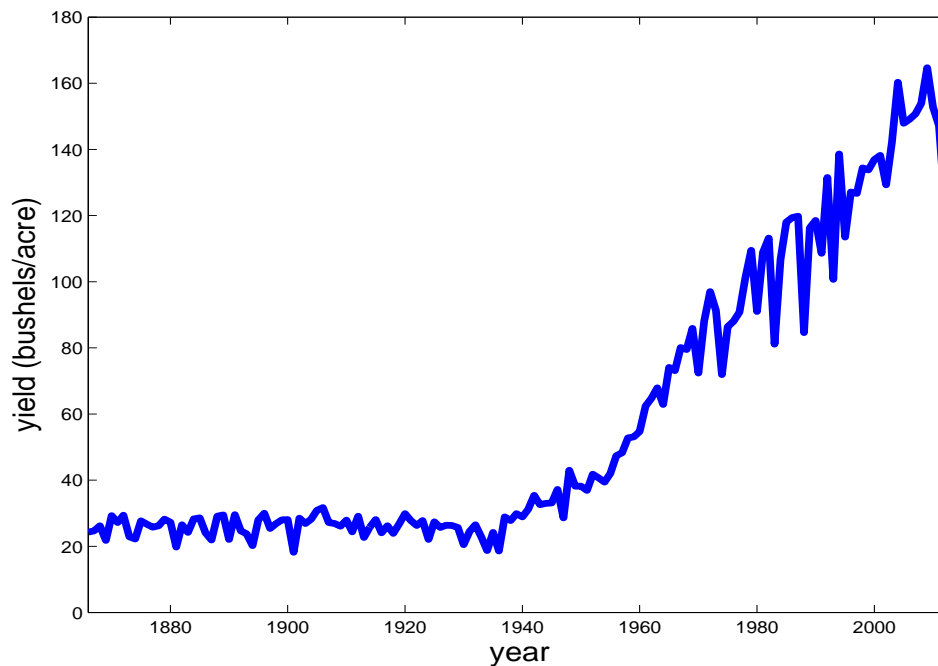


Figure 1.1: The US maize yield time series. The yield has significantly increased between the 1870s and 2012.

pounds. Maize yield is collated yearly at county level, a subdivision of a state. The yield is collected by the United States Department of Agriculture - National Agricultural Statistics Service (USDA-NASS) [78]. To calculate state level yield the county level yields are weighted according to the size of the harvested area for each county with respect to the harvested area for the state.

As maize yield only records bushels per **harvested** acre, it does not capture the full impact of weather on maize [42]. Unfavourable weather sometimes causes maize to be replanted, and this is not reflected in the yield values. Although we are aware of this, in this thesis yield is the measure we use.

The US national maize yield time series [78] (Figure 1.1) shows that maize yield has risen significantly. The amount of maize grown on one acre has increased from 24.3 bushels per acre in 1866 to 158.8 bushels per acre in 2013. This rise in yield is caused by technical advancements in seed genetics, fertilisers, crop management techniques and changes in land use [27, 29, 87]. Despite all these improvements crop yield is still significantly dependent on weather which affects both the quantity and quality of harvested crops; though only quantity is easily measured [43]. In effect the time series of the yield has two causally separate components, a non-linear increase over time caused by the technical advancements and yearly variability caused by fluctuations in weather [101]. Accounting for the technical advancements in yield time series is difficult due to its non-linearity. Presently the increase in US maize yield due to technical advancements is slowing down or stagnating [40, 68].

Examining the maize yield in Figure 1.1 shows that the biggest deviations are large downward swings followed by a retracement in the following year. These large downward swings seen in the years 1983, 1988, 1993 and 2012 are due to the weather creating adverse growing conditions, mainly drought [57]. An exception is 1993 when very wet weather caused the crops to rot in their fields [57]. To understand the limitations of any model, an understanding about the amount of uncertainty in the model inputs is required. In the next section the observational uncertainty found in meteorological observations is examined and the causes of this uncertainty reviewed.

1.3 Meteorological observations

Accurate and widespread meteorological observations are useful to understand weather's impact on crop yield [43]. The United States Historical Climatology Network (USHCN) monitor and manage weather stations throughout the US [52]. All the USHCN weather stations have long time series of meteorological observations that are checked for quality. The daily observations recorded by the USHCN are:

1. **t_{min}** : the lowest temperature (°C) at ground level
2. **t_{max}** : the highest temperature (°C) at ground level
3. **prec**: the amount of precipitation (mm/day) at ground level
4. **snow**: the amount of snow (mm/day) at ground level

Daily meteorological observations are checked by USHCN for obvious errors, for example t_{min} cannot be higher than t_{max} in a day [50]. Any problems found have error flags which show what is wrong with the observation. Error flags attached to meteorological observations are examined in Section 4.1.

Despite these quality checks, time series of daily meteorological observations still contain systematic biases and step changes caused by inconsistent measuring conditions [15, 49, 51]. A gradual change to the time of observation from the afternoon to the morning has introduced systematic positive bias [51]. Ideally the daily meteorological observations should be measured at midnight. As the USHCN network of weather stations is run by volunteers, who historically had to manually read the meter, reading the meter

at midnight was not possible. A manual adjustment is made to account for the time of observation bias. Baker [3] compared the effect of the times of the observations on the monthly means by comparing the actual means (i.e. the monthly mean temperature when the time of observation is at midnight) against the mean temperatures when the time of observations were taken for every hour. Baker demonstrated that the time of observation made a significant difference to the monthly means. Positive deviations from the mean were found when the time of observation was between 10 am and 11 pm, with the largest positive deviation at 3 pm. Negative deviations were found when the time of observation was between midnight and 8 am with the largest negative deviation at 5 am. Additionally changes to the metering equipment and location of the weather stations caused step changes in the time series [51].

Although the monthly time series have been adjusted to account for these inconsistencies, they are still contained in the daily time series [52]. There are further problems with the siting of the weather stations. For example some weather stations are located close to heat sources so the temperature observations are contaminated [21]. The observational uncertainty within the meteorological observations is discussed further in Section 4.2.

Although the USHCN has a wide network of weather stations, it does not have a weather station in every county. For example in the state of Iowa there are 99 counties but only 23 weather stations (see Section 4.1.1). Gridded meteorological data is used by the crop model. The next section examines how meteorological observations from irregularly spaced point locations are

turned in to regularly spaced gridded data.

1.4 Gridded meteorological data sets

The CERES-Maize model provides estimates for maize yield over large areas, for this it requires gridded meteorological data as an input [19]. The gridded data sets are either derived directly from meteorological observations or from reanalysis data.

1. **Gridded precipitation:** This is directly from ground level observations made into a 0.25° by 0.25° grid by the Climate Prediction Centre (CPC) Unified Rain Gauge Database [33].
2. **Gridded temperature:** Gridded t_{min} and t_{max} are from reanalysis data by The National Centres of Environmental Prediction (NCEP) North American Reanalysis (NARR) [53]. The NCEP-NARR supplies gridded t_{min} and t_{max} 2 m above ground on a 0.3° by 0.3° grid.
3. **Gridded solar radiation:** This is reanalysis data from NASA [84], it is on a grid of size of 1° by 1° at ground level.

The meteorological observations are recorded at irregularly spaced weather stations across wide areas. To convert these observations into a regularly spaced grid a data interpolation scheme is applied. These data interpolation schemes give a higher weighting to observations which lie closest to the grid point [53].

Reanalysis data is created from historical meteorological observations (including some from the USHCN weather stations discussed above) and the most recent seasonal weather model, currently the community climate system model 3 from the National Centre for Atmospheric Research (NCAR CCSM3) [13, 53]. The meteorological observations are checked against the physics of the seasonal weather model. Using a combination of both the physics of the model and the meteorological forecast, gridded reanalysis data is created. To improve consistency across the reanalysis data the most recent seasonal weather forecast model is used across the whole time series.

A more detailed description of how these gridded data sets are generated is given in Appendix A. In the next Section the physical simulation crop model used in this thesis, the CERES-Maize model, and the inputs required by this model are discussed.

1.5 The crop model

The CERES-Maize model, one part of the parallel decision support system for agrotechnology transfer (pDSSAT) [37], [19], is a modular system. CERES-Maize is a physical simulation model which produces large scale estimates for maize yield by country, state and county level. To run the CERES-Maize model, modules for weather, soils and farm management are populated with data. The gridded meteorological data sets used by the model are t_{min} , t_{max} , precipitation and solar radiation as discussed in Section 1.4. Further inputs are used by the crop model but these are not examined in this thesis. Fixed soil parameters are taken from the Harmonized World

Soil Database [22]. Information is needed about the farm management of the crops such as planting and harvesting dates, the cultivars planted and (for each cultivar) the specific crop phenology, the volume and timing of the fertiliser application, whether the crops are irrigated and the row spacing.

The gridded meteorological data used by the CERES-Maize model needs to be on a 0.5° by 0.5° scale. Gridded daily t_{min} and t_{max} are rescaled from a 0.3° by 0.3° grid by the CERES-Maize modellers using linear interpolation before being used by the crop model. The precipitation (on a 0.25° by 0.25° grid) is rescaled by the CERES-Maize modellers using resampling. Areas over sea and water are masked.

Specific weather impacts on the development of maize are captured in the CERES-Maize model. One of these is heat stress, a well known cause of crop failure [30]. The model measures heat stress via the accumulation of growing degree days (GDD). For each day d it tracks the maximum temperature $T_{max,d}$ and the minimum temperature $T_{min,d}$ against the base temperature of 10°C , T_{base} .

$$GDD = \sum_{d=1}^n \frac{(T_{max,d} + T_{min,d})}{2} - T_{base} \quad (1.1)$$

where n is the number of days in the growing season. Timing of heat stress in the life cycle is important. Just a few hours of high temperature at a critical time of maize development can cause a large negative impact on the yield. If heat stress occurs at antithesis, when the maize is fully flowered, or when the grain is growing during the reproductive phase it has large negative impacts

on maize yield. The model also tracks any stress caused by the length of the photoperiods, water and nutrients.

The CERES-Maize model provides gridded large scale estimates of maize yield on a $\frac{1}{12}^\circ$ by $\frac{1}{12}^\circ$ grid, where $\frac{1}{12}^\circ$ is about 10 kilometres. The outcomes (the yield) are recorded by USDA-NASS at county level so the gridded yield estimates have to be converted to irregular county shapes to be directly comparable with the outcomes. To convert the gridded yield estimates to county level yield estimates the CERES-Maize modellers weight the output by harvested areas in each county. If a grid cell straddles two counties the area of the grid cell in each county determines the amount of yield to include for each county.

To account for technical advancements, the CERES-Maize yield estimates need to be calibrated to the yield outcomes. To do this the CERES-Maize modellers plot a simple linear regression through yield outcomes (A_{out}). Separately, a simple linear regression is plotted through the yield estimates (A_{est}). The calibrated yield estimate (y_i) for each year i is the uncalibrated yield estimate ($y_{unadj,i}$) plus the difference between $A_{out,i}$ and $A_{est,i}$:

$$y_i = y_{unadj,i} + (A_{out,i} - A_{est,i}) \quad (1.2)$$

An example of how a county in Iowa is calibrated is shown in Figure 1.2. The red dotted line is the linear regression fitted to yield estimates from the CERES-Maize model, the blue line is the linear regression fitted to the yield outcomes. To calibrate the model the differences between the two linear

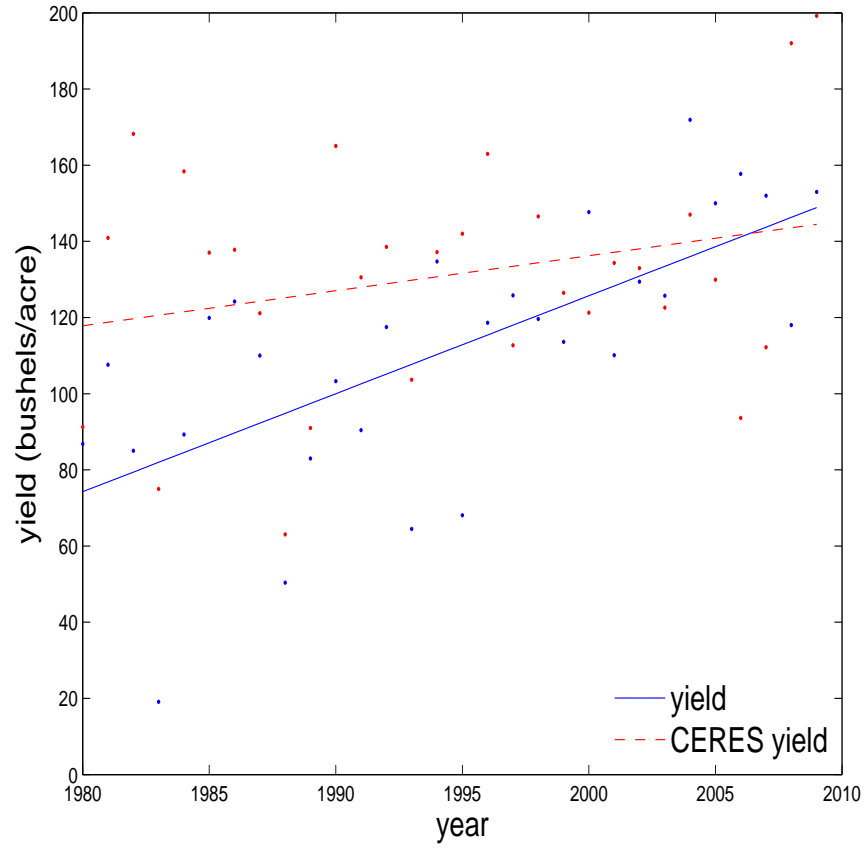


Figure 1.2: To calibrate the model linear regressions are fitted to the uncalibrated yield estimates (red dashed line) and yield outcomes (blue line) for a county in Iowa.

regression lines are added to the model estimates. For example in year 2000 the difference between the two linear regressions, -10.5, is added to the uncalibrated yield 121.26 to create a calibrated yield of 110.8 bushels/acre.

In the next section the CERES-Maize model and its weaknesses are reviewed.

1.6 Primary weaknesses in CERES-Maize model

As with all model inputs, the input data contains uncertainty. This uncertainty is introduced at several places; in initial conditions (see Section 1.3), when irregular meteorological observations are gridded (see Section 1.4) and when gridded data is rescaled for the model (see Section 1.5). The physical location of weather stations can also introduce errors. If a weather station is located in a county where maize is grown, it only measures the observations at that one point location whereas maize is grown over a large area. The meteorological observations from one point within a county may differ significantly from the actual meteorological state across the maize growing area. This is particularly true with respect to precipitation which can be extremely localised, and in regions with large topographical variations. The impact of precipitation may be smoothed by irrigation, this is considered in Section 4.4.3. Additionally the gridded t_{min} and t_{max} are estimated at 2 m above ground, whereas maize is grown near ground level.

There are also uncertainties with the size and location of maize growing areas [67] which can change over time. These changes, however, are not available on a year by year basis [67]. How the model deals with the technical advancements is another issue. The CERES-Maize model is calibrated using linear regressions (as previously discussed in Section 1.5). This assumes that maize yield increases linearly over time, however there have been concerns that the rate of technology increase is stagnating [40,68]. An issue is whether this method will cause the CERES-Maize model to over estimate future yield. In the next section the literature on how to estimate

the technical advancements from the yield time series is reviewed.

1.7 The impact of technology advancements on the yield time series

The main methods to estimate technical advancements from the yield time series are outlined below. One of the simplest ways is to estimate the technical advancements with a linear trend equation [86]:

$$y_i = \alpha_0 + \alpha_1 i + \epsilon \quad (1.3)$$

where y_i is the predicted yield for i^{th} year, α_0 and α_1 are fixed parameters and $\epsilon \sim N(0, \sigma^2)$ is the error term which is independent and identically distributed (i.i.d.). This works well if the increase in the time series is linear, however looking at Figure 1.1 the yield does not appear to be linear. There is a sharp increase in yield from approximately the 1950s, caused by a step change in technical advancements, also the variability seems to be increasing with time. One option is to add an additional parameter (α_2) for the yield increase at this step change, and another parameter (α_3) for a squared term to account for possible non-linear changes in the yield [54]:

$$y_i = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 j^2 + \epsilon, \quad (1.4)$$

$$\text{where } j = \begin{cases} 0 & i < 1950 \\ i & \text{otherwise} \end{cases} \quad (1.5)$$

where the additional terms from Equation 1.4 are j , which is set to zero before the 1950s (although the exact change point would have to be calculated) and $\epsilon \sim N(0, \sigma^2)$ is the error term which is i.i.d.

Yield time series can also account for technical advancements by basing the yield on a certain year, although which method to use depends on whether or not the variance (σ) is constant across all the years [9, 29]. Hansen [29] suggested one way to deal with yield time trend. If the trend (t_i) is a parametric or a smoothing function of the time series and the time series is stationary, an additive adjustment is used:

$$y_i = y_{unadj,i} + t_b - t_i \quad (1.6)$$

where t_i is the value of the trend function at the i^{th} year and t_b is the value at the base year (b) and y_i is the calibrated yield estimate for the i^{th} year. Hansen also suggested for non-stationary time series, a multiplicative adjustment could be made [29]:

$$y_i = y_{unadj,i} \frac{t_b}{t_i} \quad (1.7)$$

Hawkins [30] fitted a cubic regression spline which takes into account the non-linear increase of technical advancements and the stagnation in crop yields over the past decade. Other techniques used include a smoothing function

which assumes technology as a low frequency component and weather as a high frequency anomaly. Fourier analysis [28] and single spectrum analysis, a type of principal component analysis [40] are examples of this method.

To avoid having to separate the technical advancements from the yield time series a progressive-difference method [101] can be used. Yu [101] demonstrated this with multiple regression analysis on rice yields using the consecutive differences between yearly yields and climatic factors (such as sunshine, average air temperature and precipitation).

Most of these methods assume technical advancements cause a linear increase in yield. Yield increase, however, is not always steady over time, sometimes there are steep increases in yield. Additionally, yield increase is slowing down or stagnating in some locations [40, 68], so an assumption of linearity is not true.

A method to account for the technical advancements which considers the non-linearity is proposed in this thesis in Section 4.5. In the next section a simple dynamical system, the Moran Ricker Map is defined [56, 70], this will be used as a tool to explore problems inherent in larger more complex dynamical systems.

1.8 A chaotic system: Moran Ricker Map

A chaotic system is a dynamical system which is sensitive to initial conditions [83]. This means that two almost identical initial conditions will **on average** diverge exponentially over time [47]. The chaotic system used in this thesis

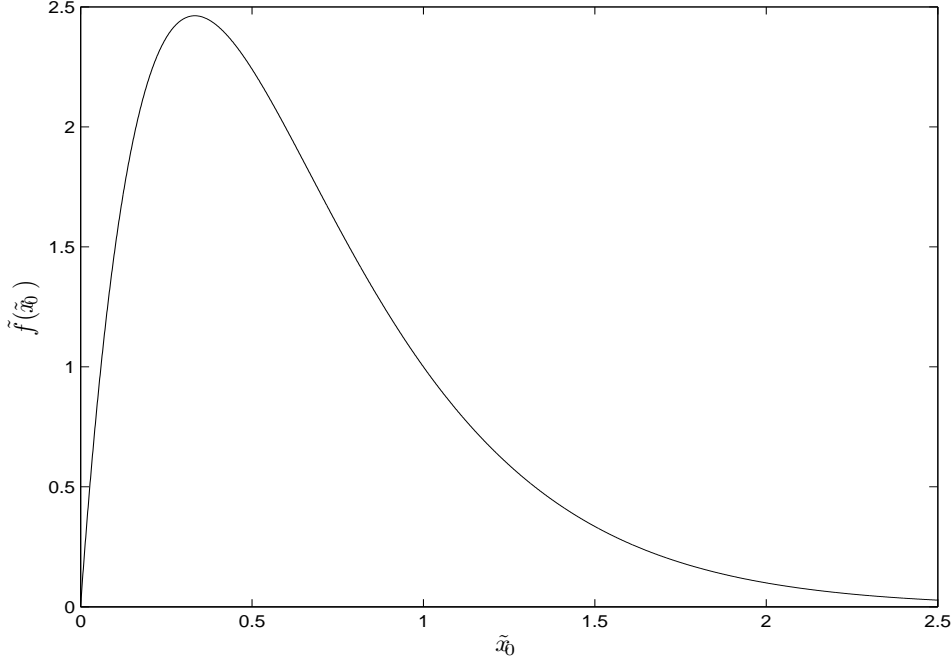


Figure 1.3: Moran Ricker Map, the image of \tilde{x}_0 when $\alpha = 3$. Note that it is bounded between 0 and 2.46.

is the Moran Ricker Map [56, 70]:

$$x_{i+1} = x_i e^{\alpha(1-x_i)} \quad (1.8)$$

The Moran Ricker Map has fixed points at $x = 0$ and $x = 1$ and is bound between 0 and $x_{max} = \frac{e^{\alpha-1}}{\alpha}$. The Moran Ricker Map when α is 3 is in Figure 1.3. For this figure the initial condition, \tilde{x}_0 , is linearly spaced between 0 and 1. The image of the initial condition is $\tilde{f}(\tilde{x}_0)$. Integrated forward means the initial conditions are run though the Moran Ricker Map to generate outcomes. For each step integrated forward, the outcomes from the Moran

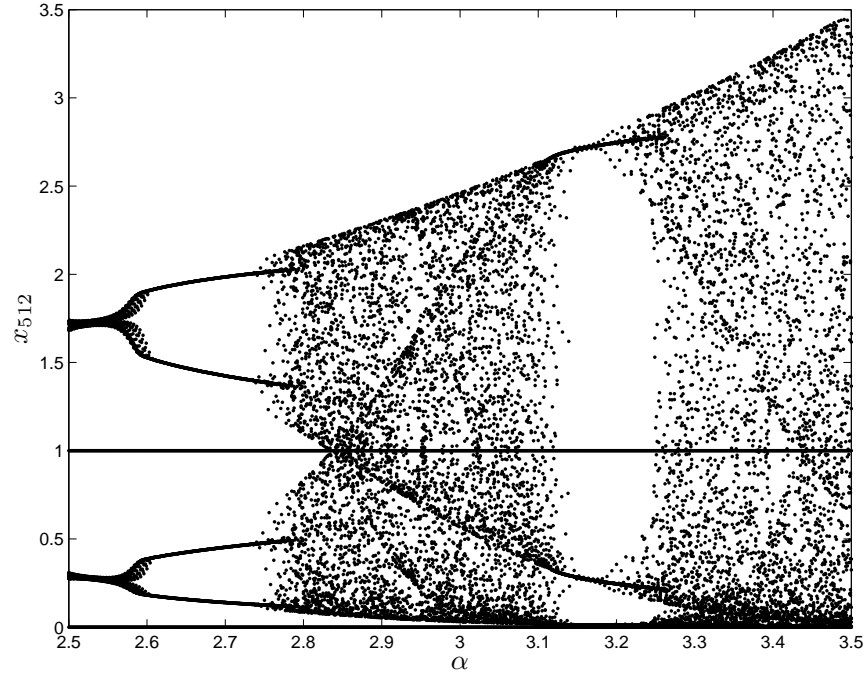


Figure 1.4: Moran Ricker bifurcation diagram. Notice the fixed points at 0 and 1 and that when $\alpha = 3$ the Moran Ricker Map is chaotic and bounded.

Ricker Map are used as initial conditions to be input again into the Moran Ricker Map.

The bifurcation diagram in Figure 1.4 shows Moran Ricker Map values when the parameter α varies between 2.5 and 3.5. The initial conditions, \tilde{x}_0 , are integrated forward 512 steps through the Moran Ricker Map. The fixed points at $x = 0$ and $x = 1$ are shown as straight horizontal lines, in this parameter space they remain unchanged across all the values of α . When α approaches ~ 2.5 a period 2 orbit starts where the values alternate between ~ 0.25 and ~ 1.75 , when α increases a period 4 orbit starts. As the period of the orbit always increases by a factor of 2, it is called period doubling

bifurcation [81]. As the value of α increases the period doubling bifurcations occur closer and closer together until eventually infinite period doubling bifurcations occur. There is however an area of stability (shown by the white space) for higher values of α when the Moran Ricker Map returns to a period 3 orbit before the period doubling bifurcations begin again.

The sensitivity to the initial conditions is measured as:

$$\left\langle \frac{\| \delta x(t) \|}{\| \delta x_0 \|} \right\rangle_{X_0} \approx e^{\lambda t} \quad (1.9)$$

where the average is over points X_0 on the attractor, $x(t)$ is the value at time t and λ is the separation of the trajectories of the system known as the Lyapunov exponent [47]. Consider a one dimensional map (such as the Moran Ricker Map) using two initial conditions at \tilde{x}_0 and $\tilde{x}_0 + \Delta \tilde{x}_0$, after being integrated forward one step the separation would be:

$$\Delta x_1 = \tilde{f}(\tilde{x}_0 + \Delta \tilde{x}_0) - \tilde{f}(\tilde{x}_0) \approx \Delta x_0 \tilde{f}'(\tilde{x}_0) \quad (1.10)$$

where $\tilde{f}' = \frac{df}{dx}$. As the number of steps integrated forward through the Moran Ricker Map tends towards infinity, the Lyapunov exponent (λ) measures the average rate at which very close trajectories diverge [81]. A positive Lyapunov exponent would mean the Moran Ricker Map has growing uncertainty [83]. The global Lyapunov exponent is:

$$\lambda = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \ln \left| \tilde{f}'(X_n) \right| \quad (1.11)$$

where N is the number of iterations.

The Lyapunov exponent for the Moran Ricker Map, when α is 3, is calculated by integrating forward the initial conditions 10^9 steps using Equation 1.11. Estimates of the Lyapunov exponent range from 0.386 to 0.387, these positive numbers show that the Moran Ricker Map is chaotic. An attractor is a set of points which x_{i+1} moves towards over time [83]. The Moran Ricker Map, like many dissipative systems, has an attractor [81]. If the initial conditions do not lie on the attractor they will make their way towards the attractor in a transitory phase [83]. In this thesis the Moran Ricker Map is used to explore skill from multi-model ensemble forecasts for a chaotic system as a proxy for a seasonal weather system.

1.9 Ensemble forecasting

Ensemble forecasts are created by adding small perturbations around the initial condition [26,59,88] to create an ensemble of initial conditions. This ensemble of initial conditions is then integrated forward through the model to create an ensemble of estimates [26,59,88]. The ensemble of estimates is converted into a probabilistic forecast by standard kernel dressing [5].

Chaotic non-linear dynamical systems are found throughout the natural world, for example in weather systems and population dynamics. In chaotic systems variations grow on average exponentially [47]. An additional problem is that in the natural world the exact initial conditions are unknown. As Lorenz has demonstrated, a minuscule error in the initial condition for a chaotic system, can cause a significantly different output [46]. Initial con-

ditions that are almost identical can soon take very different trajectories through the dynamical system. An issue is how to account for uncertainty in the initial condition.

This is a problem that has been addressed in state of the art seasonal weather models using ensemble forecasting [6, 12, 26, 31, 60]. In meteorological offices from UK, France, Germany and Italy instead of integrating forward just one initial condition through the system, an ensemble of initial conditions are used. Given a perfect model and very many observations, the initial conditions would be selected from a known probability distribution. In reality, this probability distribution is unknown. Integrating forward the trajectories of the entire initial state would use too much computer resources. Instead a finite sample is randomly selected. This ensemble of initial conditions is then integrated forward through the dynamical system to provide a measure of how the full initial probability distribution would have been transformed.

How to select the initial conditions is another issue for seasonal weather modellers. Only a finite number of initial conditions are used in operational forecasting centres due to the size of the computing required [99]. For seasonal weather forecasts, ensembles are created by adding small perturbations around the initial conditions [26, 59, 88]. The selection of these initial conditions is an area of active research; just adding random numbers to the best estimate of the initial condition does not produce the most accurate ensemble forecast [61]. These ensembles are then integrated forwards through time using a coupled ocean-atmosphere model [60, 96].

An ensemble of estimates contains information about the distribution of the

transformed initial state. A probabilistic forecast from the ensemble provides a measure of the uncertainty in the forecast. For example, if you were told there was a 90% chance of heavy rain you might take your umbrella whereas for a 5% chance you might decide not to. There are many approaches to transform the ensemble into a probabilistic forecast [24,66,73,79,94,98]. In this thesis standard kernel dressing [5] is used.

1.10 Converting an ensemble into a probabilistic forecast

Standard kernel dressing converts an ensemble into a probabilistic forecast [5,79]. Standard kernel dressing is illustrated in Figure 1.5 where individual ensemble members (black circles) are converted into probabilistic forecasts by replacing each ensemble member with a Gaussian kernels (dotted black lines). The Gaussian kernels are summed across all the ensemble members and normalised to create a probability density function (red line). An N_e member ensemble is $X_i = [x_i^1, \dots, x_i^{N_e}]$ at time i . The kernel dressing parameters used are $\Theta = [\sigma, u]$ where σ is the kernel width and u is the offset. If there are systematic errors in the model a fixed offset is used to correct the bias in the ensemble. If the width of σ is too wide, or too narrow, the skill of the probabilistic forecast will be impacted [18]. The probability density function for an ensemble is:

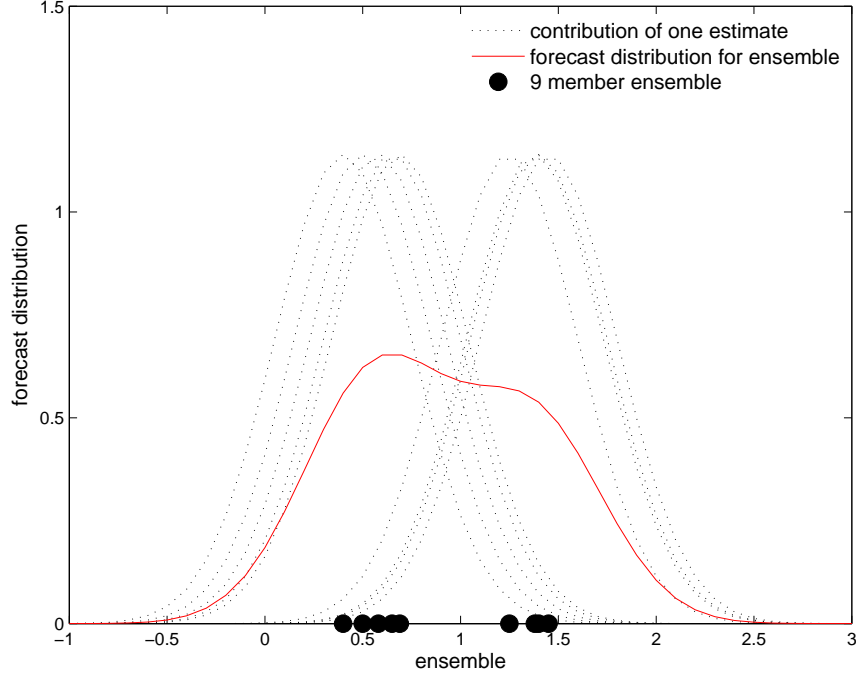


Figure 1.5: An illustration of how standard kernel dressing converts a nine member ensemble into a forecast distribution. Note the non Gaussian shape of the forecast distribution.

$$p(y : X, \sigma, u) = \frac{1}{N_e \sigma} \sum_{j=1}^{N_e} K \left(\frac{y - (x^j - u)}{\sigma} \right) \quad (1.12)$$

where p is the forecast distribution, y is the outcome and the j^{th} ensemble member x^j is replaced by a kernel centred on $(x^j - u)$. The Gaussian kernel $K(\cdot)$ is:

$$K(\zeta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\zeta^2} \quad (1.13)$$

In this thesis kernel dressing parameters were chosen by minimising Ignorance [18] as explained in Section 1.11.

1.11 Measuring skill from a probabilistic forecast

To measure the skill of a probabilistic forecast, the model's forecasts need to be evaluated against the independent outcomes. There is a set of N forecast-outcomes pairs (p_i, Y_i) , where p_i is the forecast distribution, from Equation 1.12, and Y_i is the outcome for point i . Each point i , where $i = 1, \dots, N$, has one forecast and one outcome. The skill score used to evaluate them is empirical Ignorance; the sum of the negative log probabilities of the outcomes [25, 72]:

$$S_{EI}(p(y), Y) = \frac{1}{N} \sum_{i=1}^N -\log_2[p_i(Y_i)] \quad (1.14)$$

The more skill the probabilistic forecast has, the lower on average the value of the empirical Ignorance. As the value of the probabilistic forecast can fall between a minimum of 0 and a maximum of 1, correspondingly empirical Ignorance values can range between infinity and 0. A useful way to evaluate a model's skill is to consider relative Ignorance, where the model's empirical Ignorance is measured relative to a bench mark model [5]. In seasonal

weather forecasting a standard bench mark model is climatology [36], obtained by standard kernel dressing historical outcomes, this is explained in Section 2.2.4. For climatology the skill score is:

$$S_C(p(y), Y) = \frac{1}{N} \sum_{i=1}^N -\log_2[p_c(Y_i)] \quad (1.15)$$

The equation for relative Ignorance then becomes:

$$S_{Rel}(p(y), Y) = \frac{1}{N} \sum_{i=1}^N -\log_2[(p_i(Y_i)) + \log_2[p_c(Y_i)]] \quad (1.16)$$

Negative relative Ignorance means the model's forecast has on average more probability mass on the outcome than the bench mark model forecast. Ideally the model's skill should be evaluated out of sample using cross-validation [2]. In large forecast-outcome archives, the archive is divided into two sets with one set used to train the parameters and the other independent set used to evaluate the skill.

Unfortunately only small forecast-outcome archives are available for seasonal weather models, so all their data is precious and a more careful approach must be used for training and evaluating. In this situation a compromise needs to be reached between using as much of the data as possible to obtain the most accurate results without over fitting so leave-one-out cross-validation is used [82]. The forecast distribution is $p(X_i, \Theta)$, where the ensembles are X_i , the parameter vector is Θ , the outcomes are Y_i for $i = 1, \dots, N$ and where N is the number of forecast-outcome pairs. To fit the parameter vector one forecast-outcome pair $(p(X_j), Y_j)$ is omitted and Θ is chosen

by minimising empirical Ignorance over the remaining pairs from Equation 1.16 [18]. This is repeated until every forecast-outcome pair has been omitted once. The median value $\bar{\Theta}$ from the set of N estimated Θ s is used as the kernel dressing parameters by the model. Taking the median is unlikely to allow significant information contamination [82].

1.12 Seasonal weather models

In seasonal weather forecasting not only are there are uncertainties in the initial conditions, but also from the seasonal weather models themselves. Although weather is similar to a complex chaotic system, the physical processes can only be approximated by the models. State of the the art forecasting mitigates for this by using not one model forecast, but a multi-model forecast [60, 96]. The probabilistic forecast from each model is equally weighted together. Multi-model forecasts have been shown to improve the skill of the forecast [17].

Seasonal weather forecasts are made using multi-model ensemble forecasts [60, 96]. An ensemble is made by integrating forward an ensemble of initial conditions through each model to account for observational uncertainty [32, 60]. More than one model is used to account for model inadequacy [17, 26, 93, 96]. Development of the European Multimodel Ensemble system for seasonal to interannual prediction (DEMETER) uses seven global coupled ocean-atmosphere models [60] from European Centre for Research and Advanced Training in Scientific Computation, France (CERFACS), European Centre for Medium-Range Weather Forecasts International Orga-

nization (ECMWF), Istituto Nazionale de Geofisica e Vulcanologia, Italy (INGV), Laboratoire d’Océanographie Dynamique et de Climatologie, France (LODYC), Centre National de Recherches Météorologiques, France (Météo-France), the UK Met Office, UK (Met Office) and Max-Planck Institut für Meteorologie, Germany (MPI).

So that these seasonal weather models were directly comparable with each other, the initial conditions run through each model were as similar as the individual model’s constraints allowed. A set of hindcasts, historical weather initial conditions, was used from ECMWF 40-year Re-analysis (ERA40) from 1980 to 2001 [89]. The models were launched four times a year (on the 1st February, 1st May, 1st August and 1st November) creating a six month seasonal forecast each with nine ensemble members. Due to the seasonal nature of weather a month ahead forecast for March is not comparable to a month ahead forecast for September. In the [forecast-outcome archive](#) there are only 22 March one month ahead forecasts, hence 22 points. The ensemble from each model was equally weighted to produce a multi-model forecast [17, 60]. There is a high computational cost from running multiple seasonal models, so only a small number of ensemble members, nine, were run for each model in DEMETER [8, 60]. There are more recent international multi-model ensemble forecasts for seasonal weather such as ENSEMBLES [31]. Nevertheless the experimental design for the multi-model ensemble forecasts used in this thesis is based on DEMETER.

1.13 Overview of introduction

This chapter has provided an introduction to crop models, a simple dynamical system and meteorological observations. The background places the thesis in context, but aside from the presentation there is no new contribution in this Chapter. Original work is introduced starting in Chapter 2. The questions considered in this thesis include:

1. “How to interpret the information contained within an ensemble?” is in Section 1.10.
2. “How to combine and weight estimates from different models to produce a forecast with the most skill?” is in Sections 2.4 and 2.6.
3. The limitations from using small data sets is in Section 2.3.
4. “Whether including an empirical crop model in the forecast can add skill?” is in Section 3.7.
5. “How to account for technical advancements in the yield curve?” is in Section 4.5.

1.14 Contributions

The main contributions from this thesis are:

1. Creation of models for a simple one-dimensional dynamical system, as shown in Figure 2.2, as a proxy for seasonal weather forecasts. Using these models to investigate multi-model ensemble forecasting.

2. Demonstrating the limitations from making probabilistic forecasts using a small forecast-outcome archive in Section 2.3.
3. An investigation proving the skill of multi-model ensemble forecasting can be improved by including climatology as a separate model in an equally weighted multi-model forecast is in Section 2.5.
4. An investigation proving individual model skill is improved by blending climatology with models of the Moran Ricker Map is in Section 2.6. Blending with climatology is not new, but blending with the Moran Ricker Map is.
5. An investigation proving the skill of multi-model ensemble forecasts is increased by blending Moran Ricker Map models with climatology before equally weighting the forecasts is in Section 2.6. Again blending is not new but using blending with multi-model forecasts is.
6. Generating a probabilistic forecast from the CERES-Maize model in Section 3.1 by standard kernel dressing the singleton ensembles with a Gaussian kernel.
7. Selection of a suitable bench mark model to measure the skill of crop models is discussed in Section 3.3.
8. Identifying which method to select the kernel width for probabilistic forecasts provides the most skill is illustrated in Figure 3.19 from Section 3.5 for crop yields.
9. An investigation proving equally weighting the CERES-Maize model

with an empirical model increases the skill of the forecast. This is explored at US level in Section 3.7 and at state level in Section 3.8. The state level results are shown in Figure 3.25.

10. Graphical analysis of the variability of meteorological observations and yield across different counties in Iowa are in Chapter 4.
11. A new method for identifying technology increase in the yield curve using maximum prior yield is explored in Section 4.5.
12. A method to produce an ensemble of gridded weather realisations which capture the sampling uncertainty in the gridded data is in Chapter 5. This method will be used in the future work to create a probabilistic forecast for yield.

Chapter 2

Illustrating challenges to multi-model forecasting when data are precious

2.1 Introduction

Many practical forecasting systems are challenged by the fact that the forecast-outcome archive is small, this is due to the short duration that the system has been observed as well as the slow pace of adding new outcomes. In seasonal weather forecasting, for example, there are roughly 50 well observed years. The outcomes start at the advent of high quality satellite monitoring, and can only increase by one additional year every year. Given a collection of models, each producing an ensemble of simulations, challenges arise ranging from how to establish the skill (in-sample with cross-validation) to determin-

ing how to weight and combine (or discard) ensembles or simulations from different models.

This chapter studies these challenges by constructing an analogous system. Within this system the number of models and their quality, along with the parameters that define their ensemble simulations and the size of the forecast-outcome archive are varied, so that the impact of this variation is quantified. The analogue is found in an “idealised world” based on the Moran Ricker Map. The aim here is not to solve the problems of seasonal forecasting with a one-dimensional chaotic map, but rather to illustrate these challenges, and clarify which are unavoidable given multiple models and a small archive, and which might be resolved. It is hoped insights regarding the nature of the challenges are generalised to actual seasonal forecasting.

Section 2.2 is used to explain the background information for the numerical experiments, such as how the initial conditions for the ensemble are selected (Section 2.2.1) and the bench mark model (Section 2.2.4). The bench mark model uses a naïve prior distribution called climatology [5]. This is a standard bench mark model used in seasonal weather forecasting [36].

A series of experiments are conducted in the “idealised world” designed to resemble the multi-model ensemble forecasting methods used by seasonal weather forecasting. Given the complexity of the physical weather systems, these experiments are conducted on the simple one-dimensional Moran Ricker Map [56, 70] (Equation 1.8), a chaotic non-linear dynamical system that acts as a proxy for a weather system. Using three imperfect models of this system allows us make an “idealised world”, unencumbered with most

of the problems faced by seasonal weather models, but still able to demonstrate the important features of forecasting a chaotic dynamical system. An overview of the experimental design is given in the flow chart in Figure 2.1. In particular, the only uncertainty is from model inadequacy, as the perfect model is known and noise free observations are generated.

Contrasts between a case with a large archive, and a case with a small archive are made in Section 2.3. How forecasts from multiple models are combined is examined in the context of a large forecast-outcome archive. The skill from equally weighting the forecasts from multiple models is examined in Section 2.5, in addition whether using all the models in the multi-model forecast is the best option is explored. In seasonal weather modelling, blending with climatology improves the skill of the forecast [5, 80] and this is generally believed to be true for chaotic systems [5]. Blending with climatology is explored in an idealised chaotic system with unlimited data. In particular whether using climatology as a separate model in an equally weighted forecast is beneficial is explored in Section 2.5. Additionally, the skill of a multi-model forecast if the individual model's forecasts are blended with climatology before being equally weighted together is examined in Section 2.6.

symbol	meaning of symbol
α	blending parameter that weights model with climatology
ϵ	random noise
ϵ_j	perturbation for the j^{th} ensemble member
κ	size of the radius of the uncertainty circle

symbol	meaning of symbol
K	Gaussian kernel used for kernel dressing
m	number of models in the multi-model forecast
N_{arch}	the number of points in the large forecast-outcome archive S_{arch}
N_{attr}	number of points that lie on the systems attractor
N_{clim}	number of points in very large set of outcomes Y_{clim}
N_e	number of ensemble members
N_{iter}	number of steps integrated forward through the model
N_{train}	number of points in the training set S_{train}
N_{test}	number of points in the testing set S_{test}
$p_c(y)$	probability density function of climatology
$p_m(y)$	density function of the forecast distribution
s	observation
$S(p(y), Y)$	skill score
S_{arch}	large forecast-outcome archive
S_{attr}	set of initial conditions which lie close to the dynamical system's attractor
S_{CL}	empirical Ignorance for climatology
S_{EI}	empirical Ignorance
S_{Rel}	empirical Ignorance relative to climatology
S_{sample}	small set of forecast-outcome pairs
S_{test}	large set of forecast-outcome pairs used for testing
S_{train}	large set of forecast-outcome pairs used for training

symbol	meaning of symbol
σ	kernel width, a kernel dressing parameter
σ_{cl}	kernel width for climatology
u	offset, a kernel dressing parameter
v	state space
\tilde{x}_0	a set of evenly spaced points between 0 and 2.5
x_i^j	the j^{th} ensemble member from X_i
X_i	an N_e member ensemble of estimates at time i
Y	outcome
Y_{clim}	a very large set of outcomes

In the next section models for the chaotic dynamical system are explained.

2.2 Models for the non-linear dynamical system

In the “idealised world”, the observations generated by the Moran Ricker Map are noise free. Points integrated forward through the system are defined as $\tilde{f}(x)$. To examine the effect of model error the same points are integrated forward through the model $f(x)$. Three models of the Moran Ricker Map are created, these are:

MR12: A truncated Taylor series expansion of the $e^{3(1-x_i)}$ section of the Moran Ricker Map (Equation 1.8) to the 12th term (see Appendix B.1):

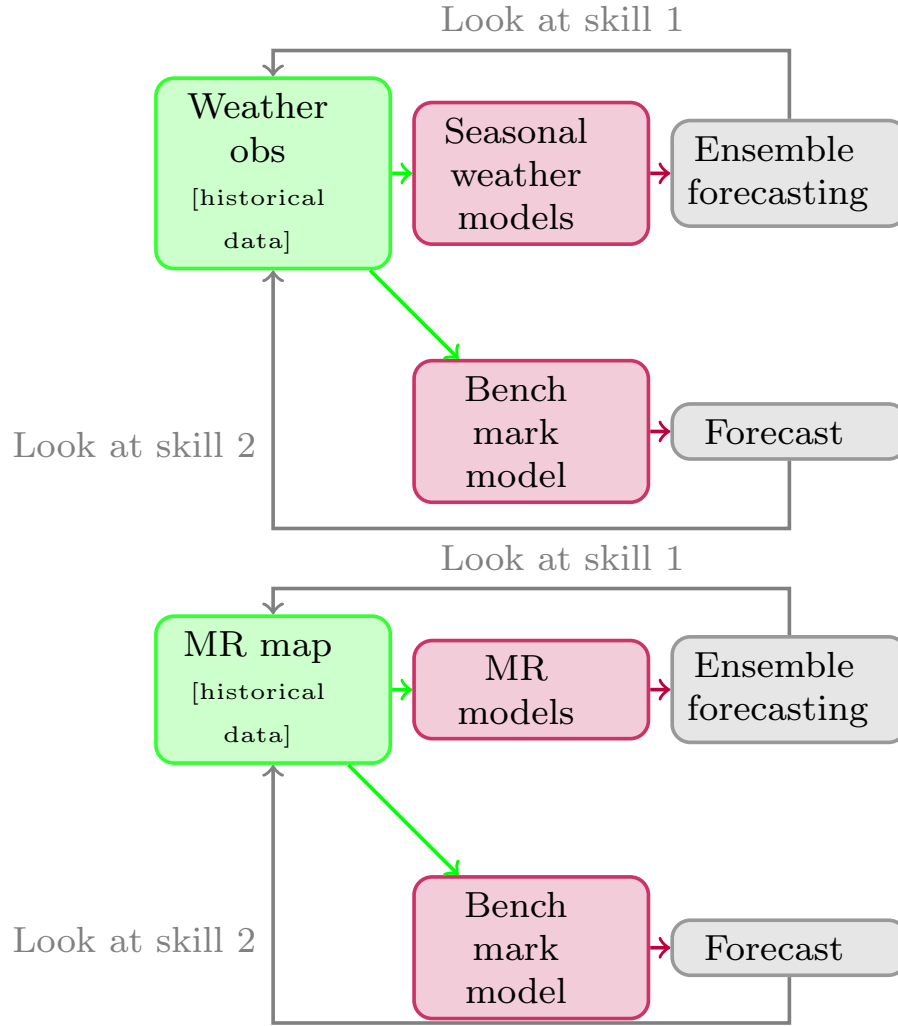


Figure 2.1: Overview of experimental design. Top: multi-model ensemble forecasts used for seasonal weather forecasting. The system is the weather, modelled by seasonal weather models using ensemble forecasting. Skill from (1) is measured against the skill from the bench mark model, climatology (2). Bottom: the idealised world using the Moran Ricker Map as the system. Models which approximate the Moran Ricker Map are used to create an ensemble forecast. Skill from (1) is measured against the bench mark model, climatology (2).

$$x_{i+1} = x_i \left[1 + 3(1 - x_i) + \frac{1}{2!}(3(1 - x_i))^2 + \frac{1}{3!}(3(1 - x_i))^3 + \dots + \frac{1}{12!}(3(1 - x_i))^{12} \right] \quad (2.1)$$

MRLM: A truncated log model expansion of the Moran Ricker Map to the 8th term (see Appendix B.2):

$$\log(xe^{3(1-x)}) = \log(x) + \log(e^{3(1-x)})$$

$$\log(xe^{3(1-x)}) = \log(x) + 3 - 3x$$

$$\log(xe^{3(1-x)}) = \log(x) + 3 - 3(e^{\log(x)})$$

$$\log x_{i+1} = \log x_i - 3 \left(\log x_i + \frac{(\log x_i)^2}{2!} + \frac{(\log x_i)^3}{3!} + \dots + \frac{(\log x_i)^8}{8!} \right) \quad (2.2)$$

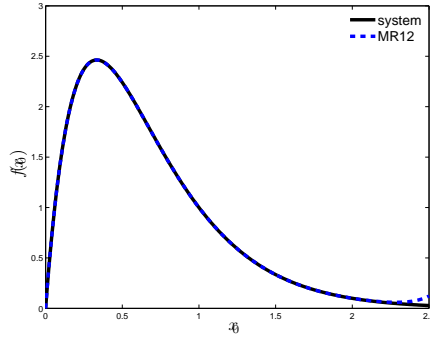
At the fixed point 0, x_{i+1} is set to 0.

MRFT: A truncated Fourier model of the Moran Ricker Map to the 12th term (see Appendix B.3):

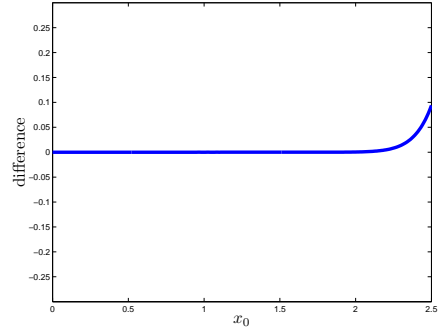
$$y = \frac{2.8391}{\sqrt{(2)}} + 0.7551 \cos\left(\frac{2\pi x}{2.46}\right) - 0.2872 \cos\left(\frac{4\pi x}{2.46}\right) \quad (2.3)$$

The initial condition \tilde{x}_0 contains 1000 regularly spaced points between 0 and

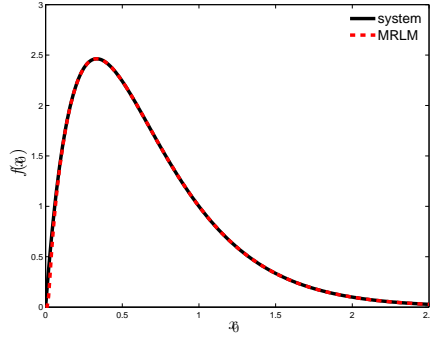
[†]see Appendix B.3 for full values



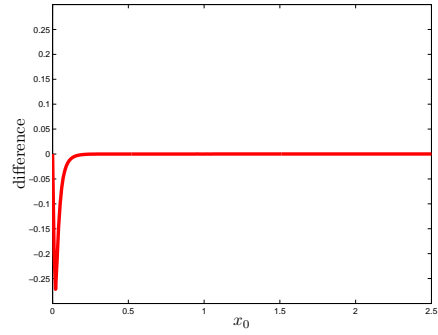
(a) System and MR12.



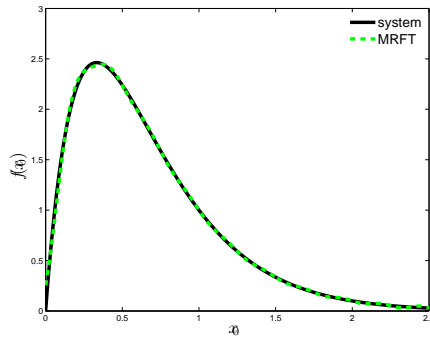
(b) Differences using MR12.



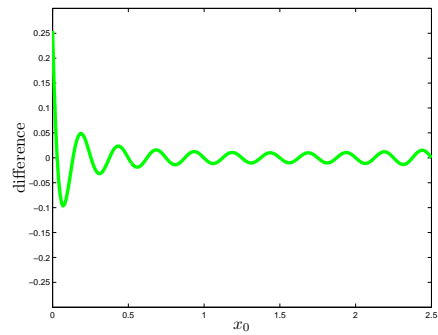
(c) System and MRLM.



(d) Differences using MRLM.



(e) System and MRFT.



(f) Differences using MRFT.

Figure 2.2: Figures (a), (c) and (e) show how close the models are to the system when integrated forward one step. Each model deviates from the system at different locations. For MR12 (b) the large difference is when \tilde{x}_0 is approaches the maximum value. For MRLM (d) the large difference is when \tilde{x}_0 is close to 0. For MRFT (f) the differences oscillate across the entire \tilde{x}_0 range but are largest for \tilde{x}_0 close to 0.

2.5. \tilde{x}_0 is integrated forward one step through the model $f(\tilde{x}_0)$ and the system $\tilde{f}(\tilde{x}_0)$. Comparisons between the model (in colour) and the system (in black) in Figure 2.2 show the similarities between the models and the system. Considering the differences, the locations of the deviations from the system by each model are clear. Model MR12 in Figure 2.2b has a large deviation from the system when \tilde{x}_0 approaches 2.5. Model MRLM in Figure 2.2d has a large deviation when \tilde{x}_0 approaches 0. Lastly model MRFT in Figure 2.2f oscillates around 0 across all the values of \tilde{x}_0 but the larger deviations are when \tilde{x}_0 approaches 0.

2.2.1 Initial conditions

Initial conditions should be a realistic sample from the dynamical system. Making random perturbations in all directions could result in an unrealistic ensemble of states including some states which the system would never actually reach. To ensure initial conditions used by the models are representative of initial conditions from the dynamical system, the initial conditions must be close to the dynamical system's attractor [83]. For clarity there may well be an attractor in each model, but in this experiment only initial conditions close to the attractor from the system are used. To ensure this \tilde{x}_0 is integrated forward 1000 steps through the Moran Ricker Map to make S_{attr} . Points generated from earlier integrations are not used in these experiments. S_{attr} is defined as a set of initial conditions on or close to the system's attractor.

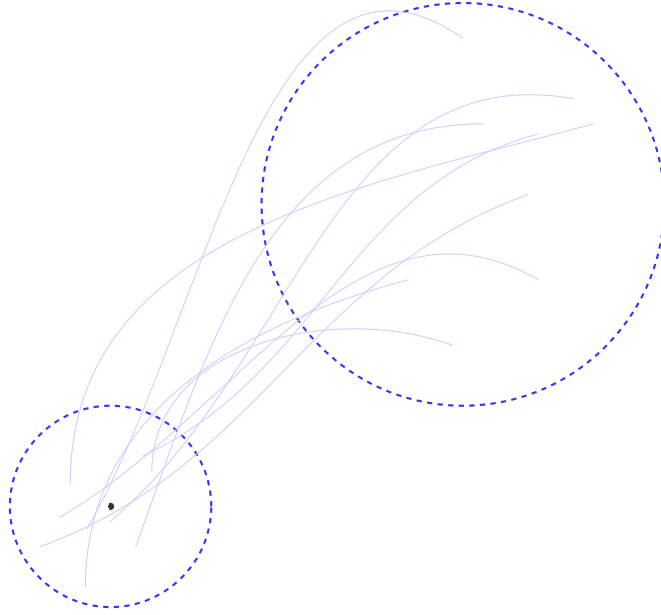


Figure 2.3: How the value of κ (left circle) determines the spread of the ensemble (right circle) at iteration two for a two dimensional system.

2.2.2 Creating the ensemble

Following work by [23] κ is defined as the spread such that the mean square error of the model estimate when integrated forward two steps is equal to the ensemble spread.

To create an ensemble of initial conditions, perturbations are drawn from inside an uncertainty circle surrounding the observation. Initial conditions for the Moran Ricker Map are drawn from a one dimensional line, so the width of an uncertainty interval needs to be estimated, rather than the radius of an uncertainty circle. As the observations are noise free, in this experiment

the uncertainty captured by the ensemble needs to be representative of the model uncertainty. As all the models vary in different ways from one another they each need intervals of different widths. The interval κ needs to be wide enough so that once the ensemble of initial conditions is integrated forward two steps through the model the observed outcome should be contained within the spread of the ensemble [7]. Even a perfect ensemble however has a $\frac{2}{N_e}$ chance that the outcome lies outside the range of the ensemble [63]. The steps to calculate κ are summarised in the flow diagram in Figure 2.4. S_{attr} was integrated forward two steps through both the system and the models. To create an ensemble member a perturbation was added to the observation s_i . The perturbation was randomly drawn from a uniform distribution and scaled to match the uncertainty interval, applicable for each model, by tuning by κ as shown in Equation 2.4.

$$x_i^j = s_i + \epsilon_j \quad (2.4)$$

Where the j^{th} perturbation is ϵ_j , $\epsilon_j \sim U(-\frac{\kappa}{2}, \frac{\kappa}{2})$ where $0 \leq \kappa \leq 1$ and $i = 1, \dots, N_{attr}$ and $j = 1, \dots, N_e$. To calculate the value of κ for each model the following calculations were made after integrating forward two steps using the set of initial conditions S_{attr} :

- 1. Mean square error between model image and system image:**

The differences between the images of the initial conditions from the system $\tilde{f}(x_i)$ and each model $f(x_i)$ was quantified using the mean square error.

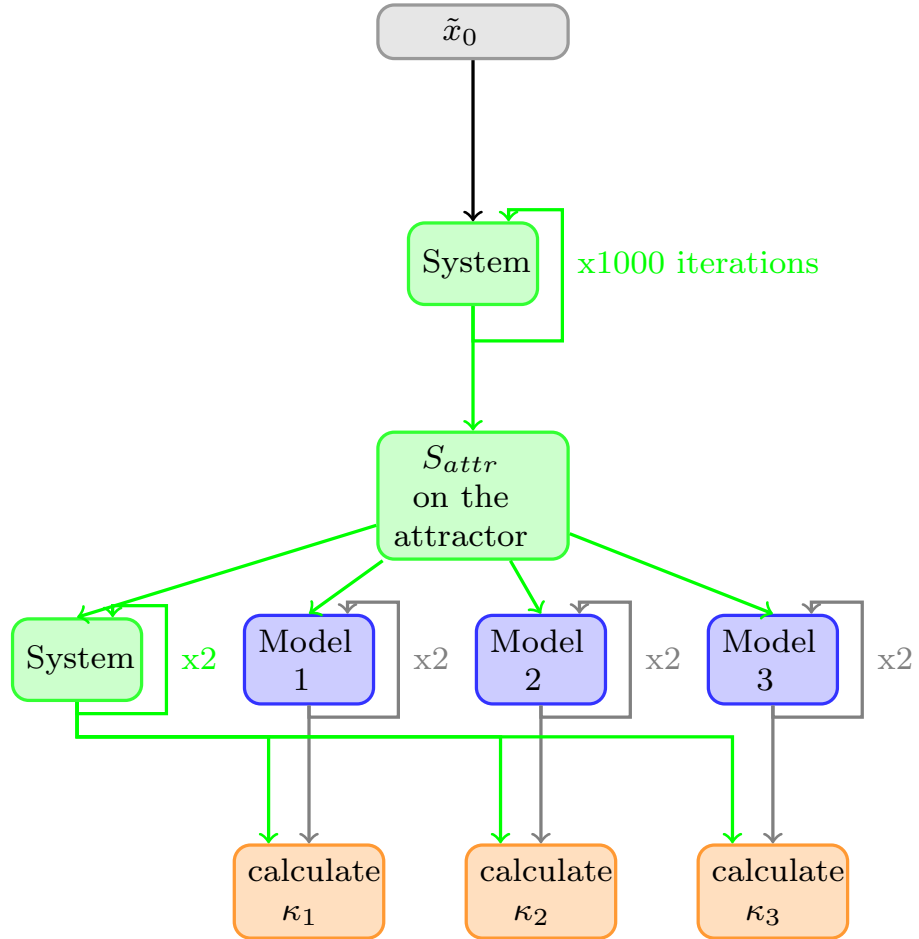


Figure 2.4: Creating realistic ensemble members: To generate S_{attr} \tilde{x}_0 is integrated forward 1,000 times through the system so that the initial conditions lie close to the system's attractor. State space is shown in green, model spaces are shown in blue. The uncertainty interval width (κ), shows the maximum perturbation added to the initial condition to create an ensemble of initial conditions. κ is set when the ensemble spread equals the mean square error of the model estimate at step two.

Model	κ
MR12	0.059
MRLM	0.030
MRFT	0.039

Table 2.2: Radius (κ) for the uncertainty interval around the observations from within which the perturbations are randomly drawn.

$$e = \frac{1}{N_{attr}} \sum_{i=1}^{N_{attr}} (f(x_i) - \tilde{f}(x_i))^2 \quad (2.5)$$

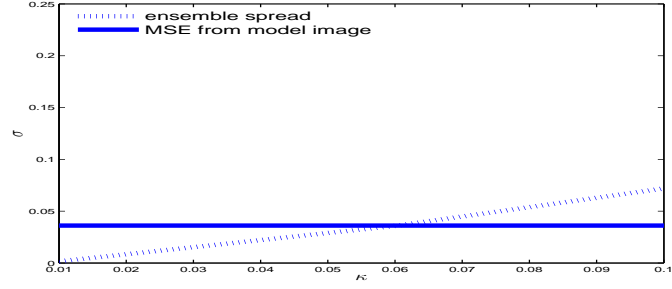
where $i = 1, \dots, N_{attr}$.

2. **Ensemble spread:** The mean square error between the individual ensemble members (x_i^j) and the ensemble mean (\bar{X}_i) .

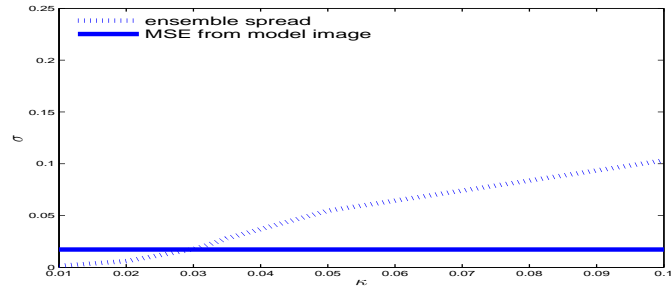
$$e = \frac{1}{N_e N_{attr}} \sum_{i=1}^{N_{attr}} \sum_{j=1}^{N_e} (x_i^j - \bar{X}_i)^2 \quad (2.6)$$

Figure 2.5 shows how κ was selected for each model, as the value where the mean square error of the model estimate (blue line) crosses with the ensemble spread (dotted blue line). The ensemble spread is widest for model MR12 in 2.5a.

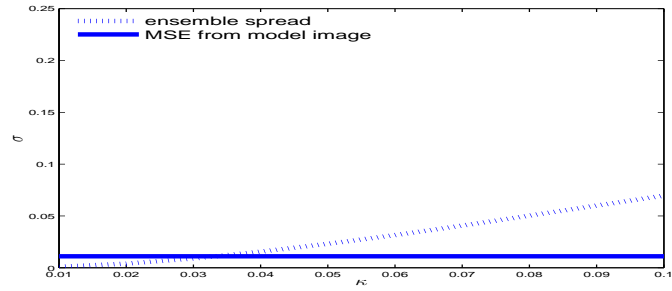
The width of the uncertainty interval (κ) for each model is shown in Table 2.2. Models MRLM and MRFT have smaller uncertainty intervals than model MR12.



(a) MR12



(b) MRLM



(c) MRFT

Figure 2.5: κ is set where the mean square error for the model image (blue line) crosses the ensemble spread (blue dashed line) for all models integrated forward two steps. Note that κ is widest for MR12.

2.2.3 Creating an idealised forecast-outcome archive

To collate data for international seasonal weather models requires considerable collaboration between different countries to compile the same initial conditions, the same file formats, etc. In addition satellite readings, which are now necessary as model inputs, are only available from the 1980s onwards. This means that seasonal weather forecast-outcome archives are only small, and grow very slowly, so their data is “precious” [80].

All the tools necessary to create a multi-model ensemble forecast from an idealised world which mirrors the state of the art seasonal forecasts have been created. The flow chart in Figure 2.6 explains the steps required to create an archive when integrated forward five steps. Each model has a set of N_{arch} pairs of ensembles and outcomes $[X_i, Y_i]$ at time i , where X_i has N_e ensemble members. In the idealised world a large number of ensembles from the models and outcomes from the system are generated, a luxury not available in the real world.

2.2.4 How to form a naïve probability distribution using priors

A naïve probability distribution is used as a bench mark model. Called climatology, it is a static distribution based on historical observations [5]. To convert the static climatological distribution into a probability distribution each observation is dressed with a Gaussian kernel where the kernel width (σ_{cl}) is set by minimising Ignorance [5, 18] defined in Section 1.11. If the historical set of outcomes is Y_{clim} , the probability density function is:

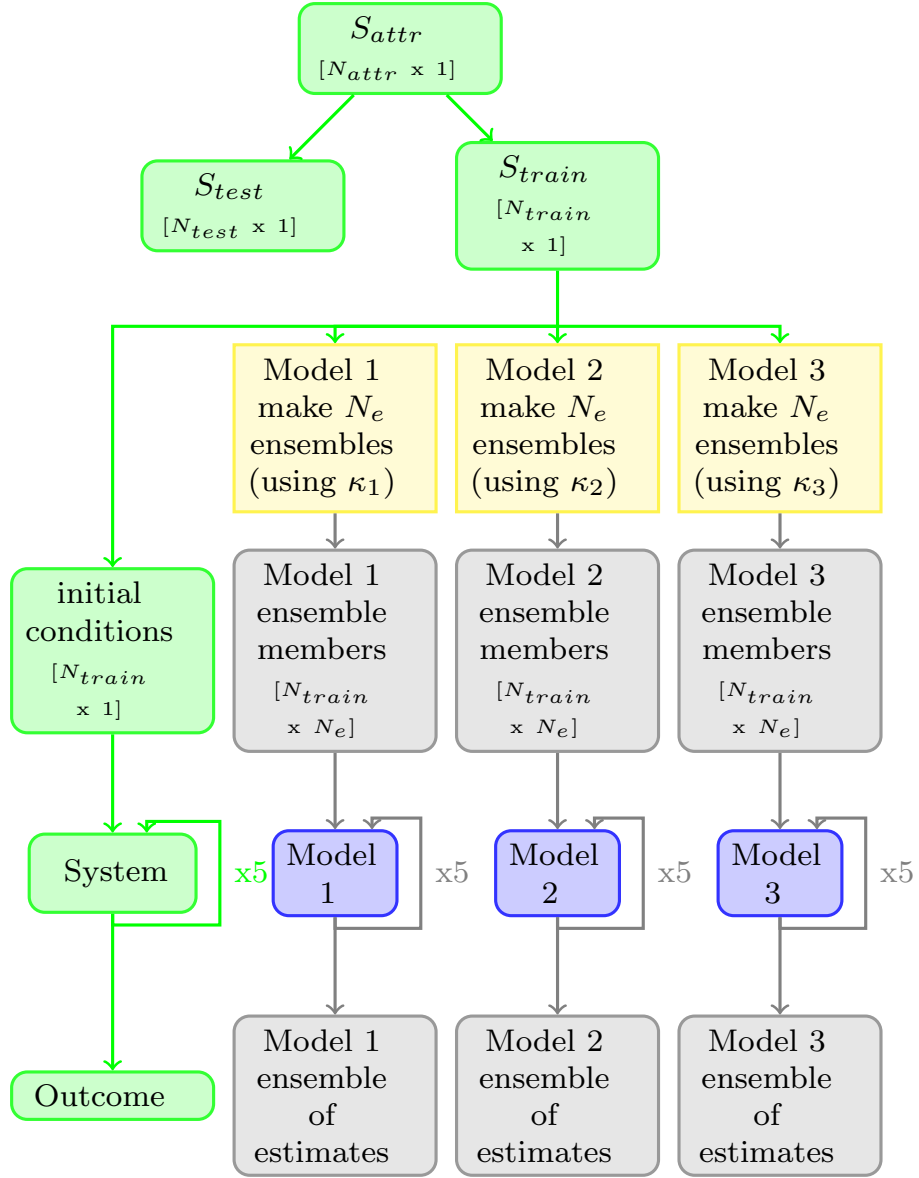


Figure 2.6: Creating ensembles and outcomes: S_{attr} is divided into two sets; one set is used to train parameters (S_{train}) and one set is used to verify the forecast (S_{test}). For each point in S_{train} , N_e ensemble members are selected. These ensembles are integrated forward five steps. Outcomes are from S_{test} integrated forward five steps.

$$p_{cl}(y : Y_{clim}, \sigma_{cl}) = \frac{1}{N_{clim}\sigma_{cl}} \sum_{i=1}^{N_{clim}} K\left(\frac{y-y_i}{\sigma_{cl}}\right) \quad (2.7)$$

In the idealised world based on a simple dynamical system, computational output is cheap, so a very large set of outcomes are generated from the Moran Ricker Map. Here Y_{clim} is a set of 10,000 outcomes. The kernel width (σ_{cl}), chosen by minimising ignorance, is 0.02. Although the Moran Ricker Map is used as a proxy for a seasonal weather system, it has no seasonality. [†] A histogram of the historical outcomes is shown in Figure 2.7, most cluster at a number just greater than 0.

symbol	value	description
N_e	9	number of ensemble members
N_{arch}	2000	points in the large <u>forecast-outcome archive</u>
N_{clim}	10000	large set of historical outcomes
N_{iter}	5	number of steps through the model
N_{sample}	22	points in small <u>forecast-outcome archive</u>
N_{test}	1000	points in the testing <u>forecast-outcome archive</u>
N_{train}	1000	points in the training <u>forecast-outcome archive</u>

[†]To estimate climatology when weather patterns differ in different seasons of the year, the kernel width (σ_{cl}) is calculated separately for each month by separating out the historical observations by month.

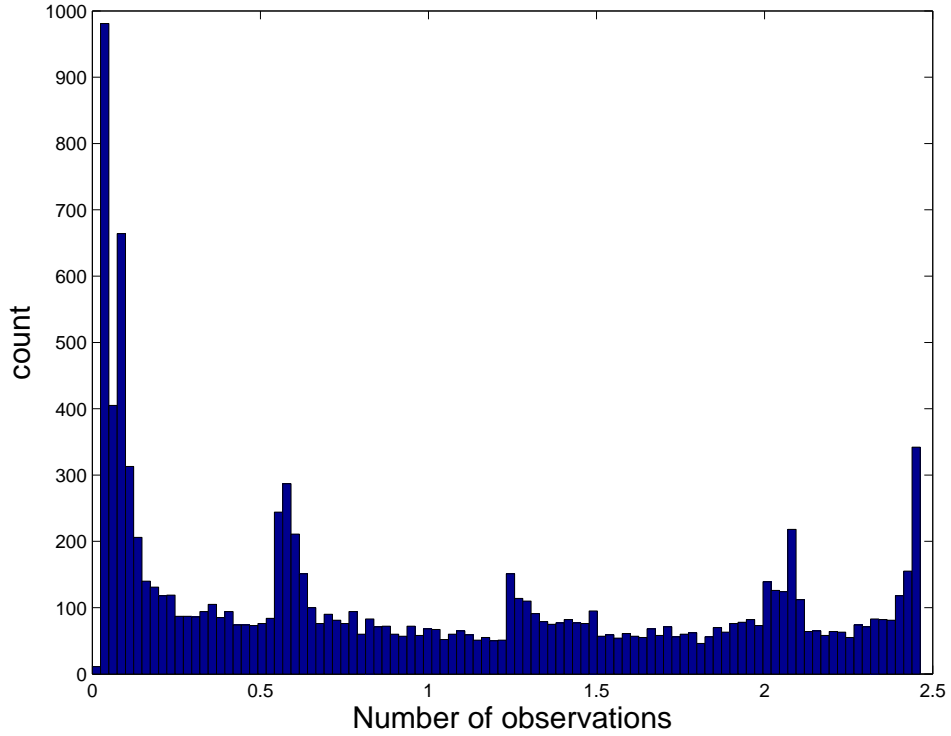


Figure 2.7: Moran Ricker outcomes (Y_{clim}).

2.3 Illustrating the restrictions of a small forecast - outcome archive

How kernel dressing parameters and skill are impacted by the size of the forecast-outcome archive is explored in this section using data from the idealised world where the only uncertainty is from model inadequacy. The dynamical system is the Moran Ricker and the three models of this system are MR12, MRLM and MRFT.

A number of original experiments are conducted on forecast-outcome pairs $(p(X_i), Y_i)$ where $p(X_i)$ is the forecast from kernel dressing the nine ensemble

2.3. Illustrating the restrictions of a small forecast - outcome archive

members X_i , and Y_i is the outcome at time i .

In these experiments the large set (S_{arch}) contains 2000 forecast-outcome pairs. The forecasts from the large set are made by using half of this set (S_{train}) to train the kernel dressing parameters, with the independent second half (S_{test}) used to evaluate the skill. The small set (S_{sample}) contains just 22 forecast-outcome pairs. The kernel dressing parameters and skill are trained and evaluated using leave-one-out cross-validation on S_{sample} . The skill of the probabilistic forecasts from the models is measured relative to climatology.

Experiment 1 compares the kernel dressing parameters and skill between different models trained and evaluated on large set S_{arch} after being integrated forward five steps.

Experiment 2 compares the skill between models trained and evaluated on a small set S_{sample} integrated forward five steps. It also compares the probabilistic forecast skill from S_{arch} and S_{sample} .

Experiment 3 examines the uncertainty in skill from probabilistic forecasts from a small set (S_{sample}) when integrated forward five steps and three steps by comparing:

1. IGN_{Big} : where kernel dressing parameters are trained on the large set S_{train} and the skill is evaluated on the large set S_{test} .
2. IGN_{Small} : where kernel dressing parameters are trained on the small set S_{sample} and the skill is evaluated on the large set S_{test} .

Experiment 4 examines kernel dressing parameters and skill from different

sized forecast-outcome archives integrated forward three steps. Step three is selected as differences between the skill at step five are hard to distinguish. Five different sized sets of forecast-outcome pairs are created with 22, 44, 88, 176, 352 and 704 points respectively. Comparisons are made between kernel dressing parameters and skill from these sets and the large set S_{arch} . Experiment 5 examines the variability in the kernel dressing parameters and skill from 1000 sets of 22 pairs (S_{sample}). For clarity kernel dressing parameters and skill are calculated separately for each S_{sample} .

The model's forecast with the most skill when a large forecast-outcome archive is used is investigated. The kernel dressing parameters and skill from Experiment 1 are in Table 2.4. The kernel width σ is wider for model MR12 at 0.84 and narrower for the other models at 0.33 for MRLM and 0.35 for MRFT. All the models have offsets around 0. None of these models significantly out perform climatology, relative Ignorance shows that only one model MRLM, with a negative score of -0.04, has slightly more skill than climatology. Model MR12 with relative Ignorance of 1.14, much higher than the other two models, has the least skill. The standard deviation which measures the variability of the relative Ignorance shows model MR12 has the widest spread of values.

Model	σ	u	rIGN	rIGN std
MR12	0.84	0.05	1.14	2.14
MRLM	0.33	0.07	-0.04	1.78
MRFT	0.35	-0.05	0.09	1.77

Table 2.4: Kernel dressing parameters and skill after integrating forward five steps for a large forecast-outcome archive. rIGN is empirical Ignorance relative to climatology and rIGN std is the sample standard deviation of relative Ignorance.

2.3. Illustrating the restrictions of a small forecast - outcome archive

Which model has the most skill using a small forecast-outcome archive is examined in Experiment 2. The results in Table 2.5 show that the wider kernel is again for model MR12 and the narrower kernel is for model MRFT. For all the models the value of u has moved further away from 0. From relative Ignorance, two of the three models MRLM and MRFT have more skill than climatology, a result inconsistent with the results from the large forecast-outcome archive (S_{arch}) in Table 2.4. If these results are to be believed, forecasting using a small sample has more skill than forecasting from a large sample. With S_{sample} from Experiment 2 used for both the training and evaluation of skill it seems that the skill may have been overestimated, a premise that is explored next. The numbers in this table are calculated from one small forecast-outcome archive, and are expected to vary significantly if a different small forecast-outcome archive is used.

Model	σ	u	IGN	IGN std
MR12	0.79	-0.15	0.94	1.96
MRLM	0.32	0.05	-0.44	1.95
MRFT	0.30	-0.15	-0.27	1.48

Table 2.5: Kernel dressing parameters and empirical Ignorance relative to climatology after integrating forward five steps where the parameters were trained using leave-one-out cross-validation on S_{sample} , an archive of 22 points.

The skill when the kernel dressing parameters are trained on large and small archives are compared in Experiment 3. The kernel dressing parameters for IGN_{Big} are trained on the large set S_{train} and for IGN_{Small} they are trained on the small set S_{sample} . They are both evaluated on the same set, S_{test} . The results, from a forecast for five steps integrated forward, is in Table 2.6. Negative differences show that the IGN_{Big} column always has more skill than the

IGN_{Small} column. Evaluating the skill using leave-one-out cross-validation on a small sample of 22 points, has caused the skill to be overestimated. Is this still the case when the models have more skill than climatology? A forecast when there is a smaller number of steps will increase the skill of the forecast. This experiment is repeated with the number of steps integrated forward reduced from five to three. The results in Table 2.7 show that models MRLM and MRFT now have more skill than climatology. Even at step three, the ensemble forecast using a small sample has overestimated the skill in this example.

Model	IGN _{Big}	IGN _{Small}	diff
MR12	1.14	1.25	-0.11
MRLM	-0.04	0.16	-0.20
MRFT	0.09	0.11	-0.02

Table 2.6: Comparison of empirical Ignorance relative to climatology (on S_{test}) when the models are integrated forward five steps and the kernel dressing parameters are trained using 1000 points (IGN_{Big}) and 22 points (IGN_{Small}).

Model	IGN _{Big}	IGN _{Small}	diff
MR12	0.20	0.55	-0.34
MRLM	-1.32	-1.08	-0.24
MRFT	-1.06	-0.99	-0.06

Table 2.7: Comparison of empirical Ignorance relative to climatology (on S_{test}) when the models are integrated forward three steps and the kernel dressing parameters are trained using 1000 points (IGN_{Big}) and 22 points (IGN_{Small}).

How do kernel dressing parameters and skill change with the size of the forecast-outcome archive? This is considered in Experiment 4. The skill from different sized forecast-outcome archives is shown in Figure 2.8 for the MR12 model (red), MRLM model (green) and MRFT model (blue). The dots are the mean empirical Ignorance relative to climatology and the bars

show the 5% and 95% confidence intervals. The confidence intervals do not decrease with sample size, which is a surprising result. This was examined further for model MR12 using histograms of the skill for the different sample sizes shown in Figure 2.9. Although the 5% and 95% confidence intervals remained a similar size across the different sized samples, the likelihood of an individual sample having a mean relative Ignorance score matching the overall mean increased with the number of points in the sample. The 5% and 95% confidence intervals were replaced with bootstrap re-sampling. For this 1000 bootstrap re-samples were taken with replacement and the means were calculated independently for each set, the results are shown in Figure 2.10. In this figure the 5% and 95% bootstrap re-sampling intervals decrease as the sample size increases. The problem, that the confidence intervals did not decrease with sample size, lay with how the confidence interval was measured. Bootstrap re-sampling provides a more accurate measure of uncertainty as it measures the distribution. The order of skill of the models remains consistent across all the different sample sizes.

Although MRLM and MRFT have quite similar skill the mean relative Ignorance of model MRLM is lower than MRFT. Comparisons between σ and u by sample size and model are in Figure 2.11. The blue line is the respective kernel dressing parameter. The bootstrap resampling line, set as 5% and 95% of 1000 sorted bootstrap values, is the dotted blue line. The green line is σ (or μ) estimated from the large set S_{arch} . For all the models σ tends towards the S_{arch} σ as the sample size increases. For model MR12, the model with the least skill, there is some distance between σ and it's S_{arch}

σ value when the sample size is 704. For the other models the estimate is closer to the S_{arch} value at smaller sample sizes. Similar patterns are seen with u estimates over the different sample sizes.

Experiment 5 investigates whether a small forecast-outcome archive with 22 points normally over estimates the skill using 1000 sets of S_{sample} . The variability for the 1000 different kernel dressing parameters is shown in Figure 2.12 for σ and in Figure 2.13 for u . In both Figures kernel dressing parameters are shown separately by model and by steps integrated forward, with five steps on the left side and two steps on the right side. For models MRLM and MRFT most of the kernel dressing parameters have more variability at five steps than at two steps as for each integration forward through the model more uncertainty is introduced. For model MR12 (blue) σ has more variability than the other models at lead time two. There are some discrepancies on how well the kernel dressing parameters are estimated using small archives. In particular both u and σ at step two for model MR12 tend to be over estimated and σ at step two for model MRFT tends to be under estimated when compared to their values estimated using the large set S_{arch} (black dashed line).

Histograms showing the difference in skill between the large set (S_{arch}) and 1000 small sets of 22 points are shown in Figure 2.14. Negative values show skill calculated using a small forecast-outcome archive has been over estimated. At five steps integrated forward the majority of samples over estimate skill. These results show that the uncertainty of the skill is large when the sample size is small. The amount of uncertainty seems dependent on

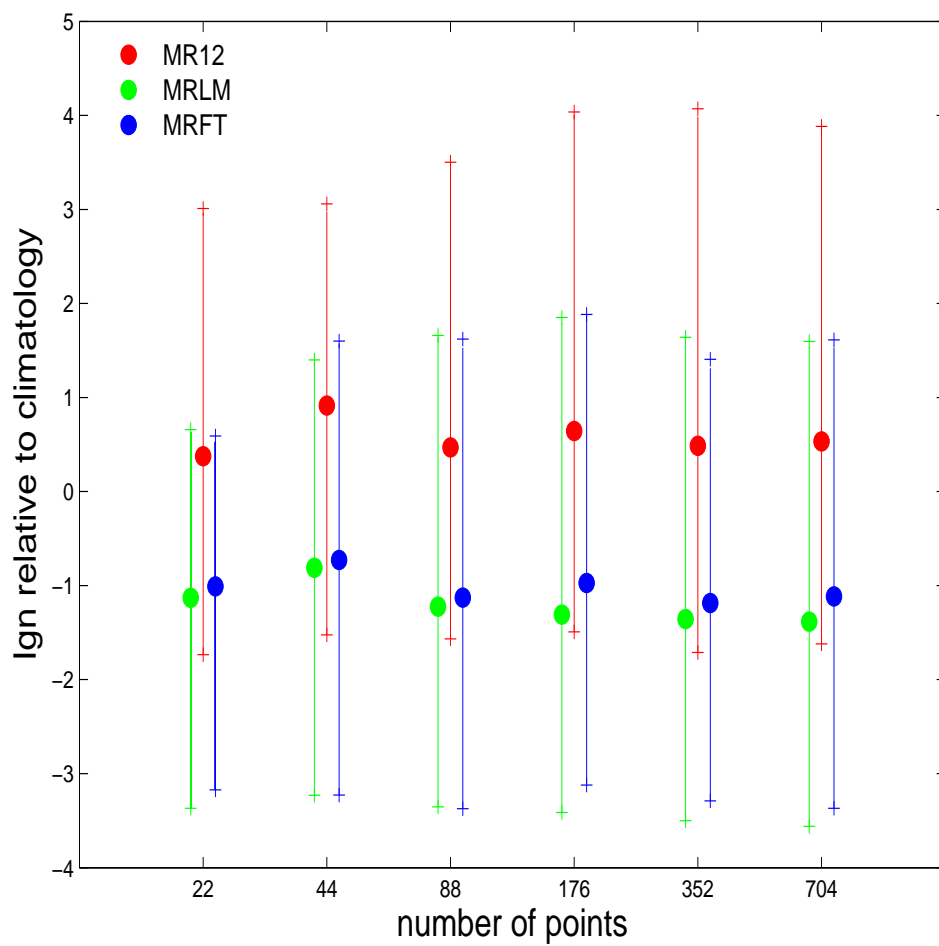


Figure 2.8: Skill by model and sample size. Note that the confidence intervals do not decrease as the sample size increases.

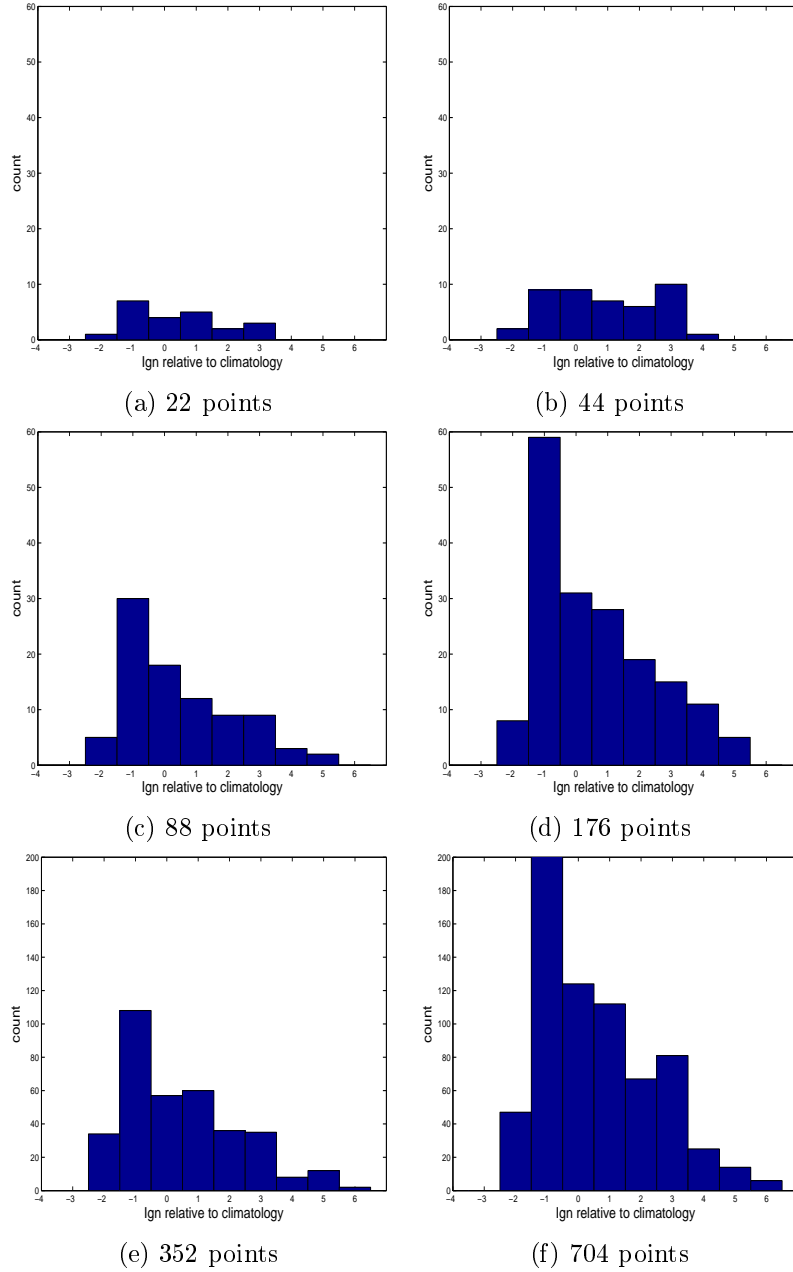


Figure 2.9: Histograms of skill by sample size for MR12. For clarity (e) and (f) have different y axes. Note with larger sets it is more likely that the sample mean will be closer to the actual mean.

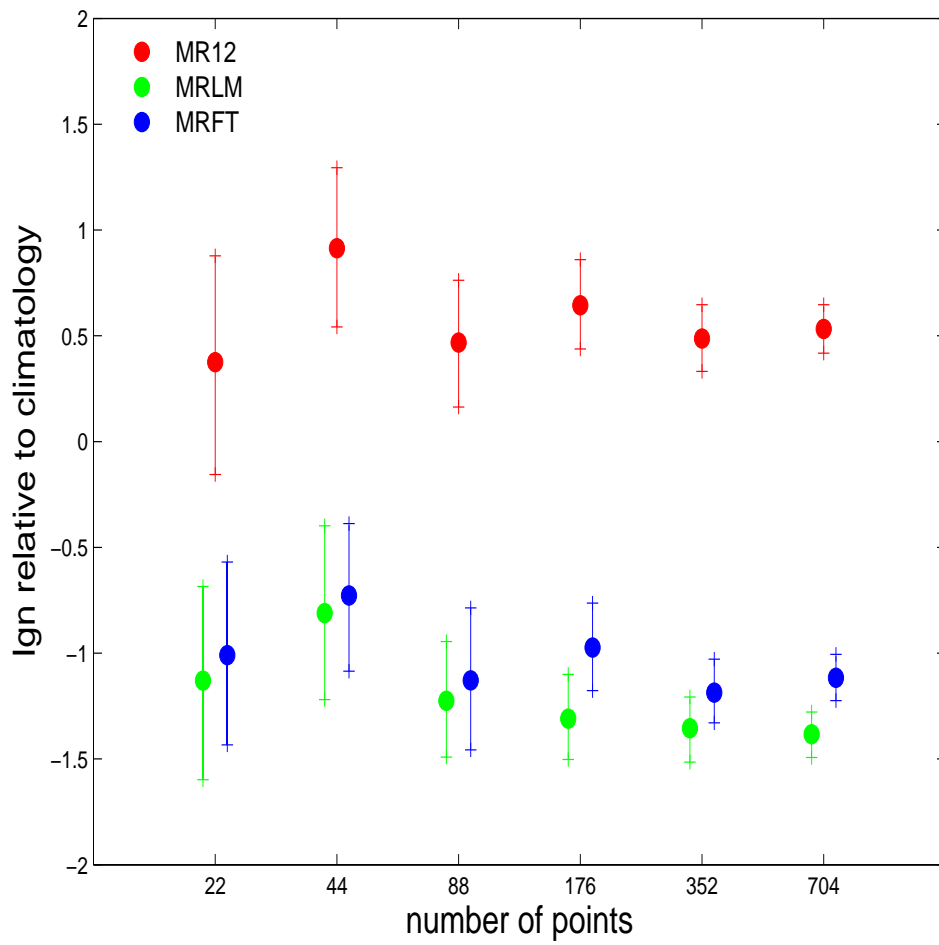
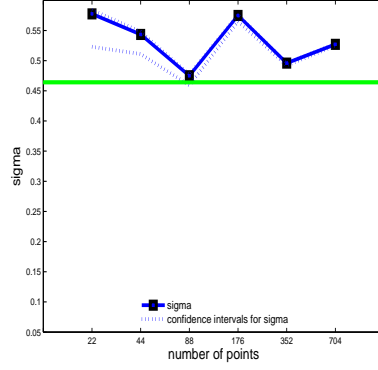
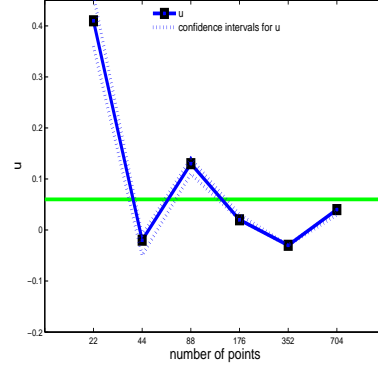


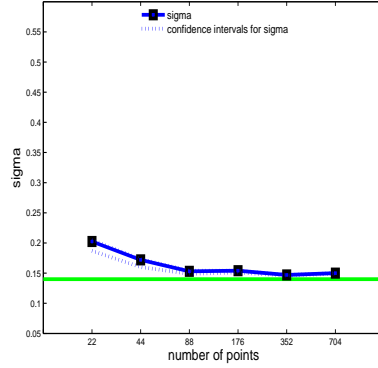
Figure 2.10: Skill by model and sample size and 5% and 95% bootstrap resampling intervals. Note that the order of which model has the most skill remains consistent across all the sample sizes.



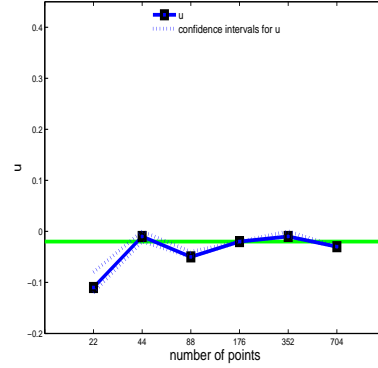
(a) MR12 σ



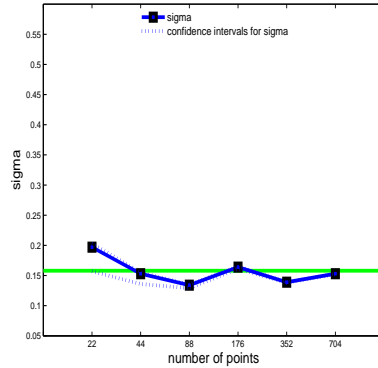
(b) MR12 u



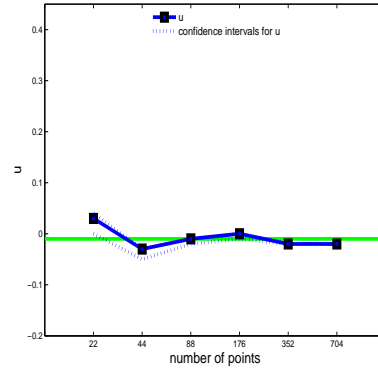
(c) MRLM σ



(d) MRLM u



(e) MRFT σ



(f) MRFT u

Figure 2.11: Comparing kernel dressing parameters of different sized forecast-outcome archives with values estimated from a large archive (S_{arch}) at step three. σ is on the left and u is on the right. Note that the σ and u values tend towards the green line as the sample size increases.

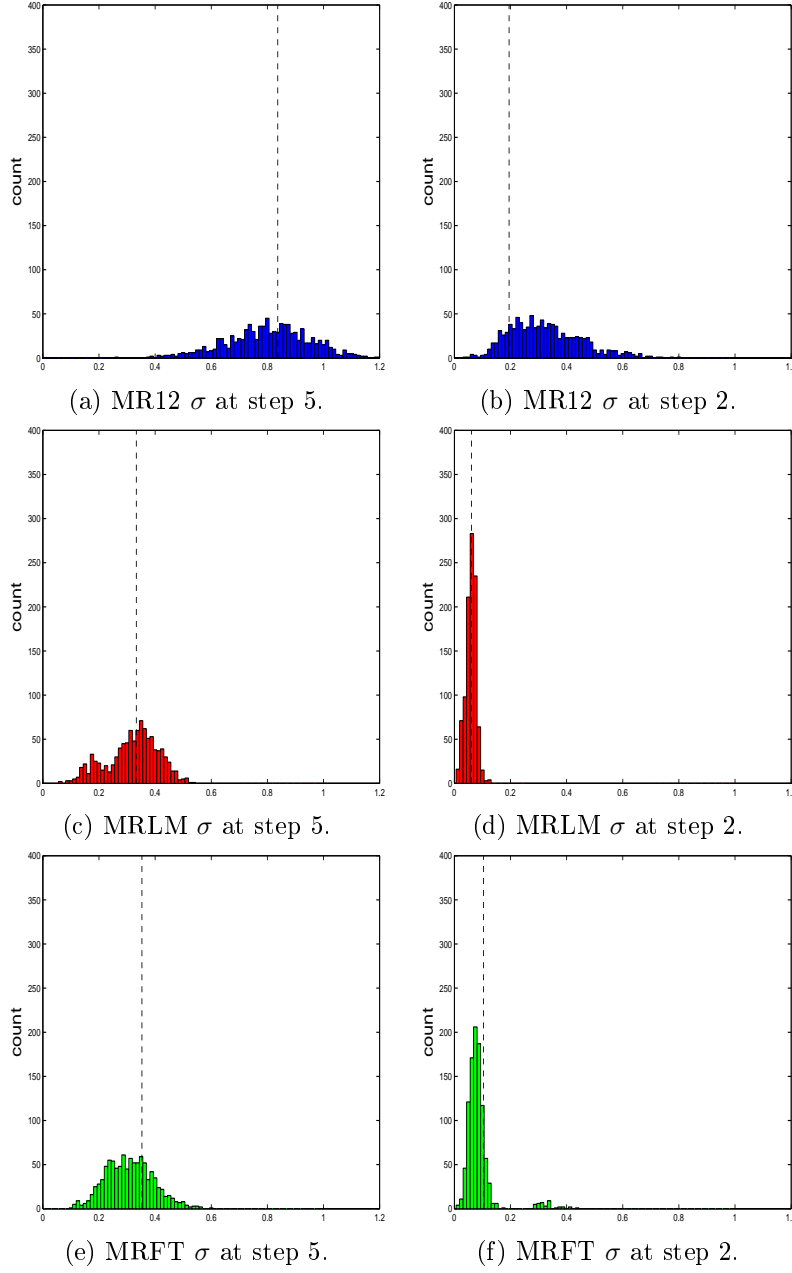


Figure 2.12: The variability of σ from 1000 sets of S_{sample} when integrated forward five steps. σ from S_{arch} is shown as dashed black lines. MR12 has the widest variability of σ estimates.

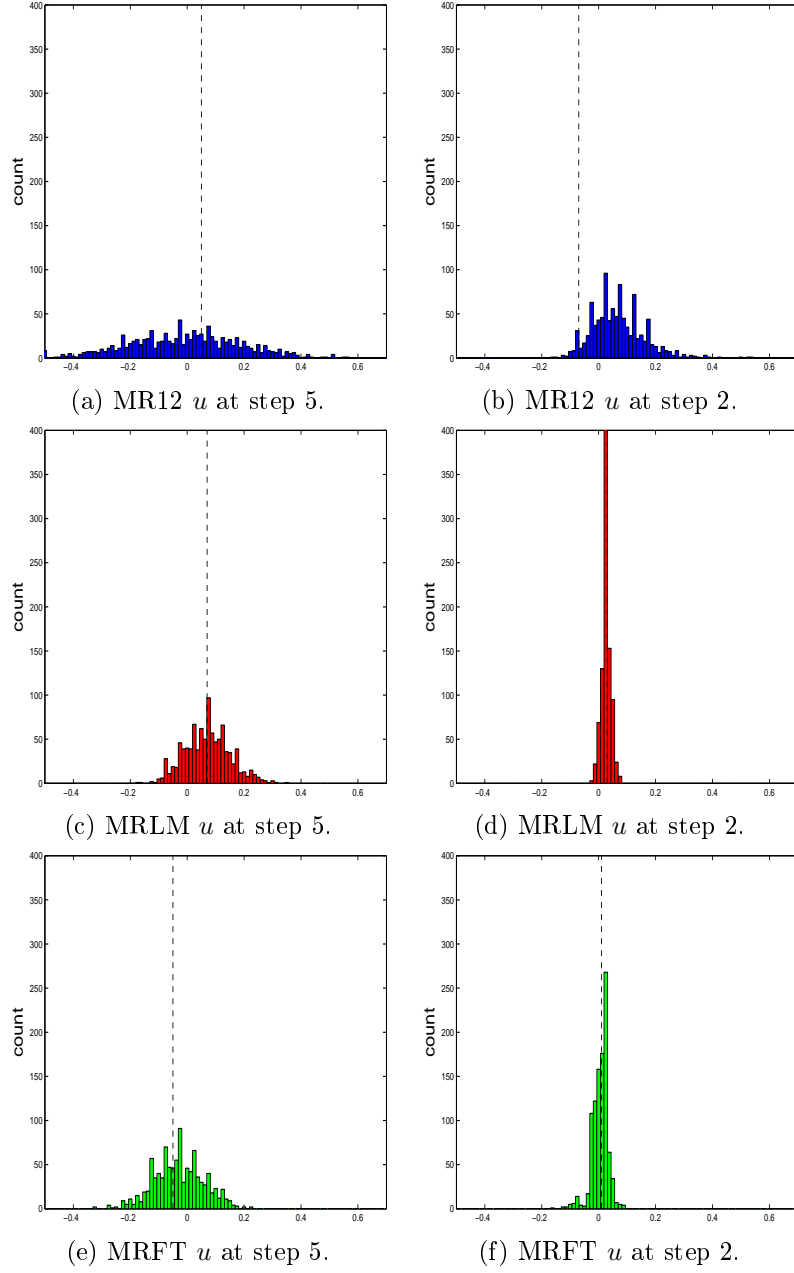


Figure 2.13: The variability of u from 1000 S_{sample} when integrated forward for five steps and two steps. S_{arch} values of u are shown as dashed black lines. The spread of u estimates is wider at step 5 than at step 2.

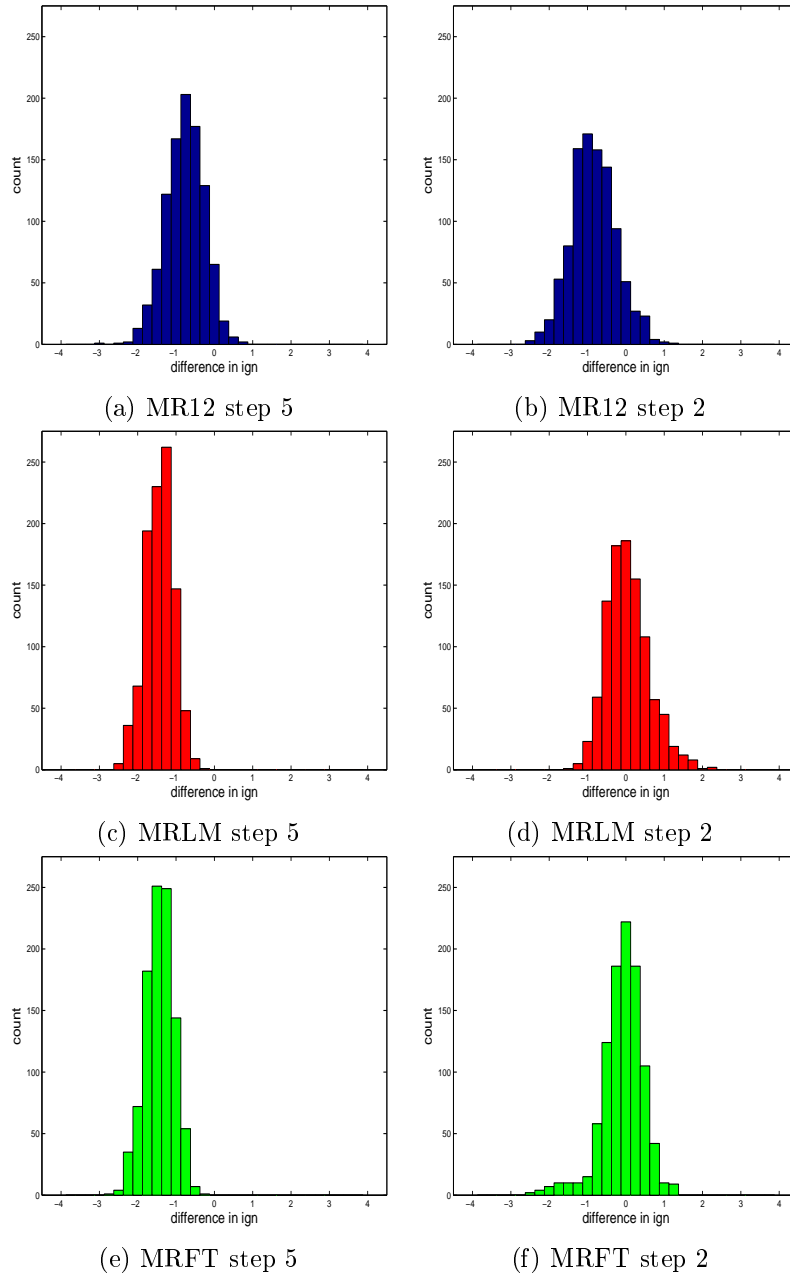


Figure 2.14: Histograms comparing the skill from a small archive. Negative values mean the skill of the small archive has been over estimated. Skill is relative to climatology. When integrated forward five steps most of the skill has been overestimated.

both the level of model inadequacy and number of steps integrated forward.

symbol	value	description
m		number of models included in multi-model forecast
N_{arch}	2000	number of points in large forecast-outcome archive S_{arch}
N_{train}	1000	points in training set S_{train}
N_{test}	1000	points in testing set S_{test}
N_{clim}	10000	points in Y_{clim}

2.4 How to combine multiple models

In DEMETER, R Hagedorn, F J Doblas-Reyes and T N Palmer [26] combine the individual model forecasts using equal weights. In this section the skill of multi-model ensemble forecasting is explored using three models of a chaotic dynamical system, the Moran Ricker Map. By design the only uncertainty in this idealised world is from model inadequacy. For these experiments a large forecast-outcome archive with 2000 points is used, a luxury not available in the real world and means the skill can be measured out of sample. The skill of equally weighted models is again evaluated relative to climatology.

2.5 Skill from equally weighted models

Sometimes the outcome lies a long way outside the ensemble, so the probability mass of the outcome occurring is approximately zero, this gives the negative log-form of empirical Ignorance a very large penalty, known as a

m	MRFT	MRLM	MR12	Clim	Relative IGN	Std
3	1	1	1	0	0.13	1.49
2	0	1	1	0	0.29	1.56
2	1	1	0	0	-0.10	1.54
2	1	0	1	0	0.35	1.54
1	1	0	0	0	0.08	1.77
1	0	1	0	0	-0.02	1.79
1	0	0	1	0	1.14	2.13
1	0	0	0	1	0	0

Table 2.8: Skill of equally weighted multi-models (without climatology as a separate model) when integrated forward five steps. Skill is measured relative to climatology.

forecast bust. To assist cases where the outcome is a long way from the ensemble members, climatology is added as a separate model to an equally weighted multi-model forecast.

Including climatology as a separate model in an equally weighted multi-model forecast is examined to see if it improves the skill. First the skill of multi-model forecasts without climatology as a separate model is calculated. The forecasts from the models are equally weighted together using $\frac{1}{m}$, where m is the number of models in that particular multi-model combination. The relative Ignorance (Equation 1.16) is calculated for every possible combination of the three models (MRFT, MRLM and MR12). Table 2.8 shows the mean relative Ignorance and the standard deviation of these relative Ignorance values.

The multi-model forecasts do not significantly outperform climatology. The forecast with the most skill is not from the three models equally weighted, but from 1100 and 0100 with relative Ignorances of -0.10 and -0.02 respec-

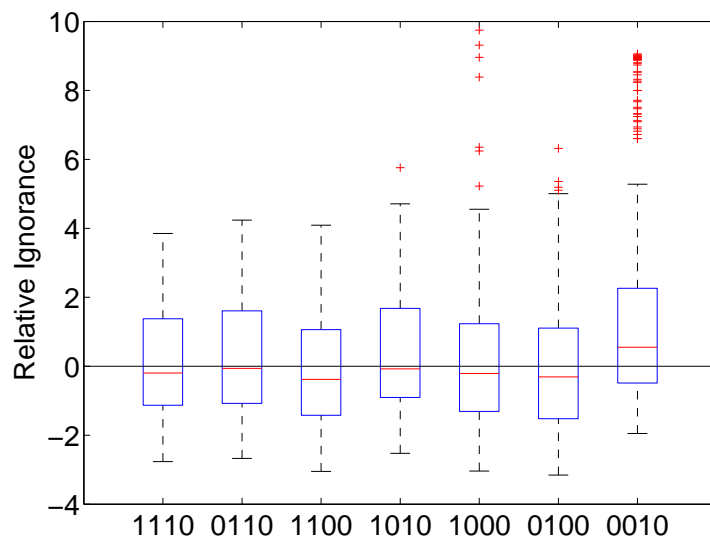
m	MRFT	MRLM	MR12	Clim	Relative IGN	Std
4	1	1	1	1	-0.06	1.01
3	0	1	1	1	-0.01	0.90
3	1	1	0	1	-0.24	0.95
3	1	0	1	1	0.04	0.85
2	1	0	0	1	-0.18	0.75
2	0	1	0	1	-0.25	0.80
2	0	0	1	1	0.21	0.55

Table 2.9: Skill of equally weighted models (including climatology as a separate model). Skill is measured relative to climatology.

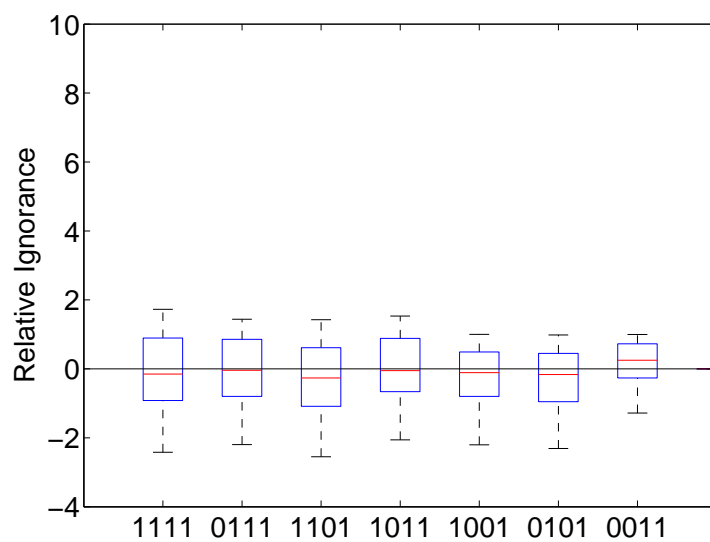
tively, though given the large standard deviation, this is not significant. The multi-model forecast with the largest relative Ignorance, and so the least skill at step five, is 0010 (or MR12) with 1.14. By definition the relative Ignorance of climatology is zero.

The skill from including climatology as a separate model in an equally weighted multi-model forecast is considered. The results are in Table 2.9. The skill for all the multi-model forecasts has increased and the variability of relative Ignorance, measured by standard deviation, has decreased. Including climatology as a separate model in the multi-model forecast produces forecasts with more skill. Only two multi-model forecasts 1011 and 0011, now have less skill than climatology.

Box plots where equally weighted multi-model forecast excludes climatology (Figure 2.15a) and includes climatology (Figure 2.15b) reveal more details about why including climatology improves the skill. With the box plots the centre of the box plot is the median, the 25th and 75th percentiles are the edges of the boxes and the whiskers mark 1.5 times the inter quartile range,



(a) Without climatology as a model.



(b) With climatology as a model.

Figure 2.15: Comparison of skill relative to climatology for equally weighted multi-model forecasts (a) without climatology and (b) with climatology at step five. Note how the variation in relative Ignorance scores reduces in (b), as there are no forecast busts (red crosses).

any points that falls outside the whiskers are shown as red crosses.

Comparing the box plots of the relative Ignorance values in Figures 2.15a and 2.15b show that if climatology is not included in the multi-model forecast the relative Ignorance scores have a far wider spread and there are more outliers, specifically points greater than 1.5 times the inter quartile range marked as red crosses, this is especially so when only single models are used in the multi-model forecast (1000, 0100 and 0010). The red crosses mark points that fall outside the whiskers, forecast busts.

When climatology is included as a separate model in the multi-model forecast, Figure 2.15b, there are no red crosses beyond the inter quartile range and the whiskers have reduced in length. This shows that including climatology in a multi-model forecast mitigates the effect of any forecast busts, reducing the variability of relative Ignorance and improving the skill. It would be interesting to repeat both of these experiments integrating forward fewer steps so the models have more skill, however this is beyond the scope of this thesis [34].

2.6 Skill from blending models with climatology

With multi-model forecasts in the previous section no differentiation is made between the models in terms of how skilful their forecasts are. If the most skilful model's forecast is given the highest weighting would the out of sample skill of the multi-model forecast improve? To explore this each model's forecast is individually blended with climatology [80] **before** being equally weighted in a multi-model forecast. The blending parameter (α) is set by

minimising Ignorance [18] so the more skill a model has, the less climatology is used in the blend:

$$p(\cdot) = \alpha p_m(\cdot) + (1 - \alpha) p_c(\cdot) \quad (2.8)$$

Where p_m is the forecast distribution from the model and p_c is the forecast distribution from climatology. The kernel dressing parameters are fitted simultaneously to avoid unnecessarily wide kernels [5].

Model	α	σ	u	rel IGN	std
MR12	0.29	0.03	-0.02	-0.24	1.70
MRLM	0.58	0.06	0.02	-0.51	1.91
MRFT	0.45	0.12	-0.01	-0.25	1.41

Table 2.10: Comparison of kernel dressing parameters and skill from blending models with climatology at five steps integrated forward (where N_{arch} is 2000.) Skill is measured relative to climatology.

The results are shown in Table 2.10. Here α is the weight used for the model’s forecast and thus $(1 - \alpha)$ is the weight for the climatology forecast. As the α values are all less than 1, all the models are more skilful when blended with climatology. The MRLM model has the highest value of α which means that blended it has the lowest weighting for climatology. MR12 model has the lowest value for α and so the largest blend of climatology. Blended MRLM and blended MRFT models have higher skill than blended MR12.

A comparison between the kernel dressing parameters and relative Ignorance scores when there is no blending with climatology and when there is blending

N_{model}	MRFT	MRLM	MR12	Relative IGN	Std
3	1	1	1	-0.43	0.96
2	0	1	1	-0.50	1.04
2	1	1	0	-0.43	1.14
2	1	0	1	-0.25	0.82
1	1	0	0	-0.13	1.00
1	0	1	0	-0.52	1.42
1	0	0	1	-0.20	0.84

Table 2.11: Skill of equally weighted multi-models where the models were blended with climatology before being equally weighted together. Skill is measured relative to climatology.

with climatology is made. Comparing the results between Tables 2.4 and 2.10 shows that the kernel width σ for all the models is much narrower when a blending parameter is used. For example, for model MR12 σ reduces by about a factor of 3, from 0.84 to 0.28. The offset parameter u for all the models becomes even closer to zero when blending with climatology. The relative Ignorance scores are now all significantly lower, so for this example blending the model with climatology produces more skilful forecasts than just using the model.

The skill is measured when individually blended forecasts are then combined as an equally weighted multi-model forecast. The results are shown in Table 2.11. All the multi-model forecasts have more skill than climatology. Is this method more skilful than including climatology as a separate model (i.e. without first blending the models with climatology)? A comparison of the results from these two methods in Tables 2.11 and 2.9 show that blending the models before equally weighting the forecasts provides the most skill in this example.

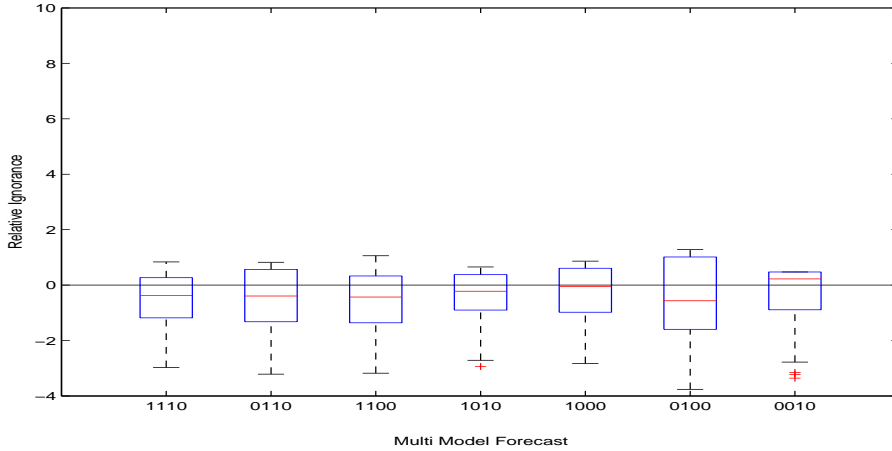


Figure 2.16: Comparison of skill for equally weighted multi-model forecasts when the models are individually blended with climatology beforehand using a data set of 2000 points at step five. Skill is measured relative to climatology. Note that there are less points above 0 relative Ignorance.

The box plots in Figure 2.16 show the spread (when $N_{test} = 1000$) of individual relative Ignorance. Almost all of the multi-model combinations have the majority of their relative Ignorance scores falling below zero, so they have more skill than climatology. The whiskers marking 1.5 times the inter quartile range extend further below the zero line than Figure 2.15b showing that more outcomes had a higher probability mass on the outcome when the models were individually blended with climatology before being equally weighted together.

2.6.1 Blending parameters at different steps

An interesting question is whether one model has more skill at a particular step integrated forward than other models. This is investigated by examining the blending parameters by model and step.

	Blending parameter (α) at step:				
model	1	2	3	4	5
MR12	0.94	0.79	0.66	0.53	0.29
MRFT	0.96	0.99	0.86	0.69	0.45
MRLM	0.99	0.79	0.70	0.80	0.58

Table 2.12: The values of the blending parameter α for each step integrated forward. Notice which model has the highest blending parameter at each step.

Table 2.12 shows the values of the blending parameter (α) for each model by step. For the majority of models the value of α decreases as the number of steps forward increases. The exception to this is model MRLM at step 4 where the value for α increases from 0.7 to 0.8.

The equally weighted blended forecasts are examined by step in Figure 2.17. Model MRLM (0100) at different steps forward is interesting. At step 2 (Figure 2.17a) it is the most skilful model with the lowest median relative Ignorance (red bar in the box plot) and with the 25th to 75th percentile range (blue box) incorporating the lowest relative Ignorance values of all the multi-model forecasts. By step 3 (Figure 2.17b) however the 25th to 75th percentile range has now widened to include both the highest values and lowest values of relative Ignorance. By step 4 (Figure 2.17c) it has again returned to the one of the most skilful models. For model MRLM the skill relative to the other models changes by the number of steps integrated forward.

The value of blending parameter α as the number of steps integrated forward increases is shown in Figure 2.18 for each model. At step 1 all the models have α close to 1. At step 2 only MRFT still maintains the high α , the α

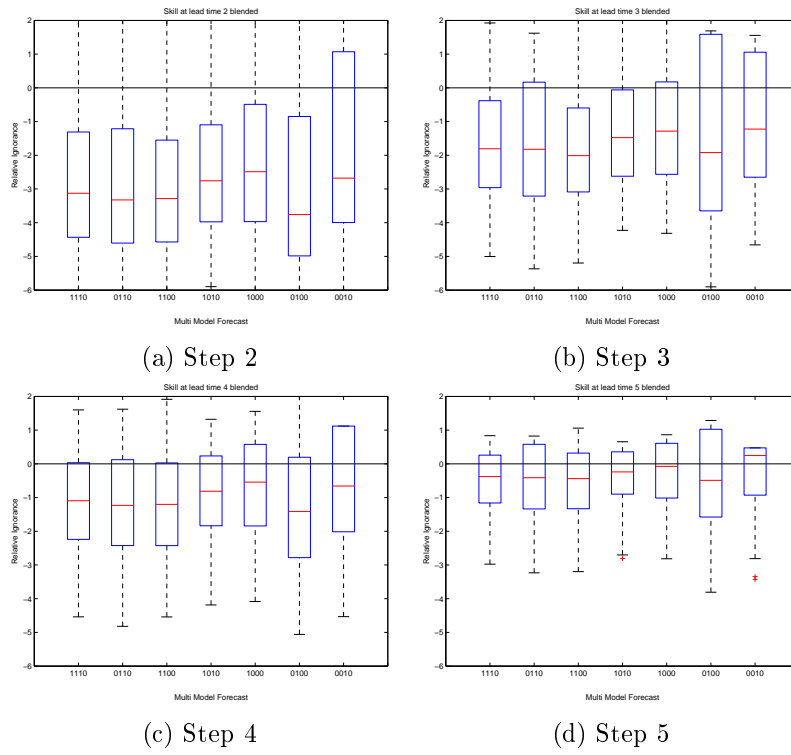


Figure 2.17: Comparison of skill relative to climatology for equally weighted blended models at (a) step 2, (b) step 3 (c) step 4 and (d) step 5. There were 1000 points in both the training and testing sets. Note that the skill for 0100 relative to the other models changes by step.

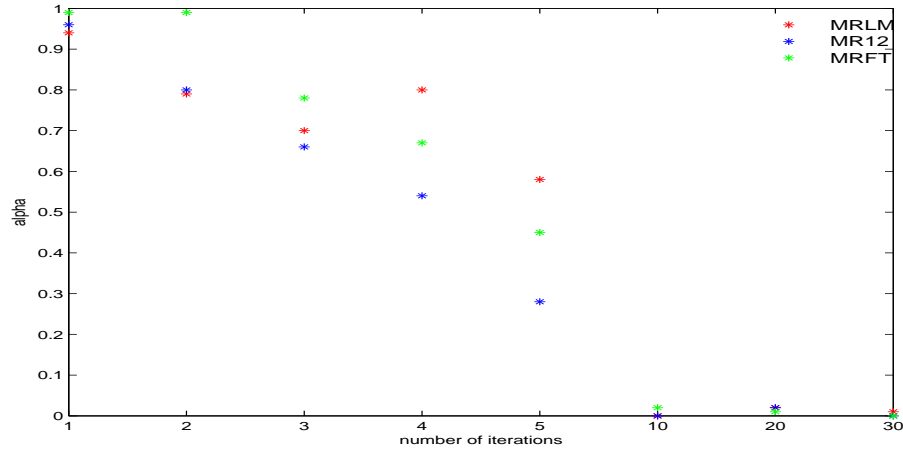


Figure 2.18: Blending parameter α as the number of steps integrated forward increase. As the number of steps increase, α for all the models approaches 0.

for the other models falls slightly. However by step 4 model MRLM has the highest blending parameter. By step 10 all models have α approaching 0.

It would be interesting to develop this further by considering the skill of each model for different sections of the systems attractor at different steps, this is an area of future research beyond the scope of this thesis.

In this chapter only equally weighting the models together is considered, as the size of seasonal weather model's archives are too small to accurately estimate weighting according to the skill of the model [71]. Noise has not been added to these models but this is also an area worthy of further investigation.

2.7 Conclusions

The basic conversion of ensemble of estimates into a forecast distribution is not new, neither is the formation of naïve probability distribution. The new

things in this section are:

1. Exploration of the state of the art forecasting methods currently in use by seasonal weather centres using idealised data.
2. Demonstrations of the restrictions from working with a small archive, the case for all seasonal weather models, is explored using an idealised world in Section 2.3. The limitations are demonstrated by using relative Ignorance to compare between forecasts from large and small archives. It is found in our Moran Ricker example that small archives tend to overestimate the skill.
3. Exploration of the kernel dressing parameters and skill by size of the forecast-outcome archive is in Section 2.3. The more skill the model has the smaller the archive has to be for the kernel dressing parameters to be closer to their asymptotic values. The order for which model has the most skill is identical across the different sized archives as illustrated in Figure 2.8.
4. An experiment examines if an equally weighted multi-model forecast provides the most skill relative to climatology with a large archive of 2000 points. In Section 2.5 the forecast with the most skill is not an equally weighted combination of all three models but a combination of the two models.
5. The increase in skill when climatology is included as a separate model in an equally weighted multi-model forecast is demonstrated for a large

archive in Section 2.5. Including climatology smooths out the forecast busts, thus providing forecasts with more skill.

6. Blending individual models with climatology, where the weight of the blending parameter is set by minimising Ignorance, is in Section 2.6. Blending isn't new, what is new is blending in the context of an idealised world based around the Moran Ricker Map. Blending models with climatology improves the skill of the individual model's forecast.
7. Blended forecasts from models equally weighted together are considered in Section 2.6. Skill scores show that, for these models, the most skilful forecasts are from this method.
8. A demonstration that different models have different skill at different steps integrated forward is in Section 2.6.1.

Chapter 3

Crop Modelling

The skill from probabilistic forecasts made by multi-model crop models is investigated using US maize yield data. In this chapter the ensemble converted into a probabilistic forecast has one member rather than the nine members used in Chapter 2. Initially, Section 3.1 explains how to convert a crop model estimate into a forecast distribution [5, 79]. Although there have been other probabilistic crop forecasts [10], using standard kernel dressing with Gaussians is unique as far as we are aware.

The skill of probabilistic forecasts from the CERES-Maize model is examined using empirical Ignorance [25, 72] in Section 3.2. To compare skill between different models relative Ignorance is used, this considers the crop models relative to a bench mark model [85]. In Section 3.3 persistence is proposed as a bench mark model instead of climatology.

Various empirical models which forecast crop yield are created in Section 3.4, these include dynamic climatology [85] and ratio models which both provide

an ensemble of estimates. Although the dynamic climatology model [85] is not new, the ratio model appears here for the first time. Models which are kernel dressed using asymmetric kernels are also proposed. The asymmetric kernels are either from a Gaussian mixture model [62] or a Gamma distribution. As far as we are aware, kernel dressing using asymmetrical kernels is a new method for crop modelling.

Comparisons between different methods to set the kernel widths are explored in Section 3.5. The kernel widths are set using (i) standard deviation of the errors (or differences) and (ii) by minimising Ignorance [18]. Due to the limited archive available for seasonal crop modelling, σ for both methods was calculated using leave-one-out cross-validation [85].

The skill of probabilistic forecasts when the CERES-Maize model is equally weighted with an empirical crop model is explored at country level in Section 3.7. Including empirical models is shown to increase the skill. This experiment is then repeated at state level in Section 3.8, here again including empirical models increases the skill.

symbol	meaning of symbol
C	number of outcomes in climatology time series
\bar{e}	sample mean
N_e	number of ensemble members
K	Gaussian kernel
n	number of points in forecast-outcome archive
μ_1, μ_2	mixed Gaussian offsets

symbol	meaning of symbol
ω_1, ω_2	mixed Gaussian weights
$p(y_i)$	probability for the i^{th} outcome
r_i	prior ratio of the i^{th} year
σ	kernel width
σ_{cl}	kernel width for climatology
u	kernel offset
X_i	is the N_e member ensemble at time i
x_i^j	the j^{th} ensemble member of X_i
y_i	the i^{th} observed outcome
z_i	the i^{th} rolling maximum outcome

3.1 Creating a probabilistic forecast from a crop model

The CERES-Maize model produces an annual estimate for US maize yield. The archive contains one estimate for the years 1979 to 2012, and one outcome, a total of 35 pairs. The CERES-Maize estimates (red) and the outcomes (green) are in Figure 3.1. The CERES-Maize model provides quite accurate yield estimates even when there are large falls in maize yield, such as the drought year of 1988 [90]. Note that there are also years where there is a large discrepancy from the model estimate, in particular in 1993, where the CERES-Maize model failed to predict the steep fall in yield. The

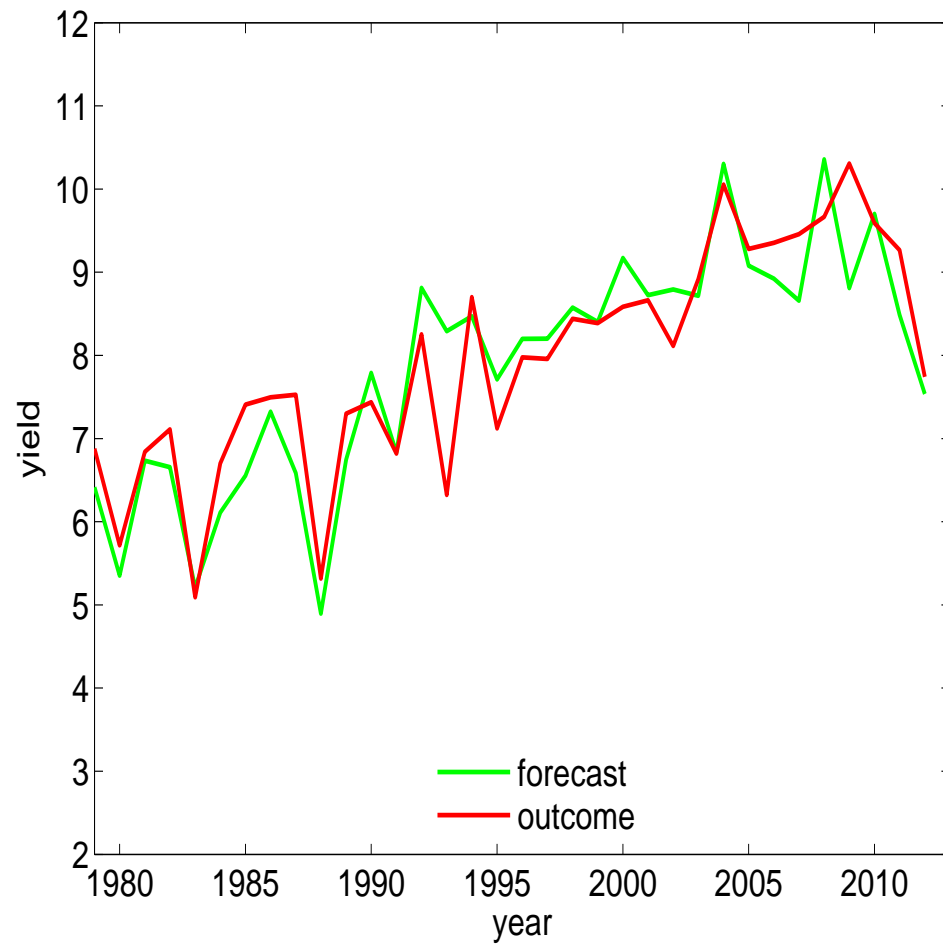


Figure 3.1: Time series of the CERES-Maize estimates (green) and outcomes (red). The CERES-Maize estimates are usually quite accurate at predicting steep falls in maize yield.

drop in yield in 1993 was caused by wet rot, conditions under which the CERES-Maize, a physical simulation model, is known not to provide an accurate yield estimate. In 1993 wet rot was known to exist 3 months before the forecast target date, so the model forecast was known to be irrelevant well before it was evaluated.

An alternative to providing one estimate is to run an ensemble of weather approximations through the crop model. This methodology is used by seasonal weather models to try and mitigate for the observational uncertainty found in initial conditions [41]. Standard kernel dressing the ensemble creates a forecast distribution [4]. Given even a singleton ensemble (x_1^i) , this method can still be used. The archive contains forecast-outcome pairs (p_i, Y_i) where p_i is the forecast distribution from a singleton ensemble and Y_i is the observed outcome at time i . The forecast distribution is made by standard kernel dressing [4]:

$$p(y : \sigma_i) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma_i} K \left(\frac{y_i - x_i^1}{\sigma_i} \right), \quad (3.1)$$

where $n = 35$. The singleton ensemble is replaced with a kernel centred on x_i^1 . The Gaussian kernel $K(\cdot)$ is:

$$K(\zeta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\zeta^2} \quad (3.2)$$

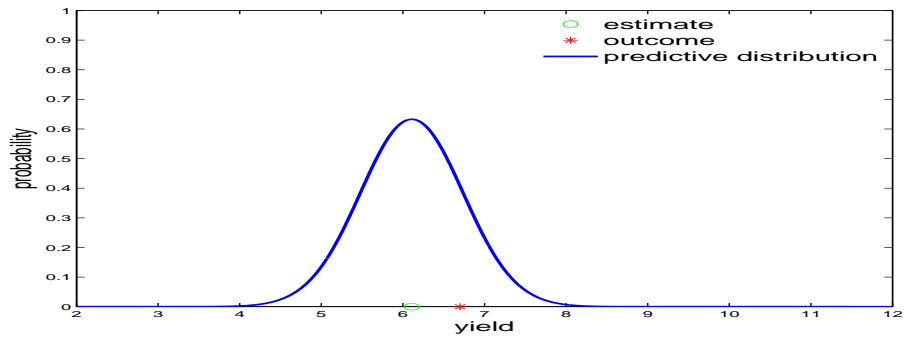
Using leave-one-out cross-validation [82] on the errors, the kernel width (σ_i) is chosen individually by year, this is discussed more fully in Section 3.5.

To calibrate the CERES-Maize estimates, the CERES-Maize modellers fit a linear regression through the estimates and the outcomes [19,37] (see Section 1.5) no offset u is used in kernel dressing. The values of σ_i , calculated from the standard deviation of the errors, are 0.6305 ± 0.0001 .

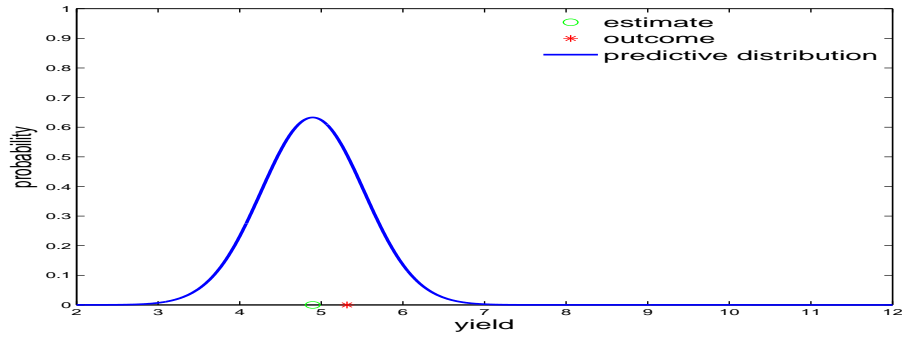
To determine if the kernel width is wide enough to capture the majority of outcomes, the forecast distribution is examined for every year. Selected years are shown in Figure 3.2, these include an average year (1984), a drought year (1988) and a year with wet rot (1993) [57]. For years 1984 and 1988 the outcome (red star) has a high probability mass (blue line). For year 1993, the outcome has a low probability mass. Outcomes should not regularly fall in very low probability areas, as this would imply that the kernel width is too narrow. Visual examination of these figures shows that this is not the case. In the next section skill from probabilistic forecasts from the CERES-Maize model are evaluated.

3.2 Crop model skill

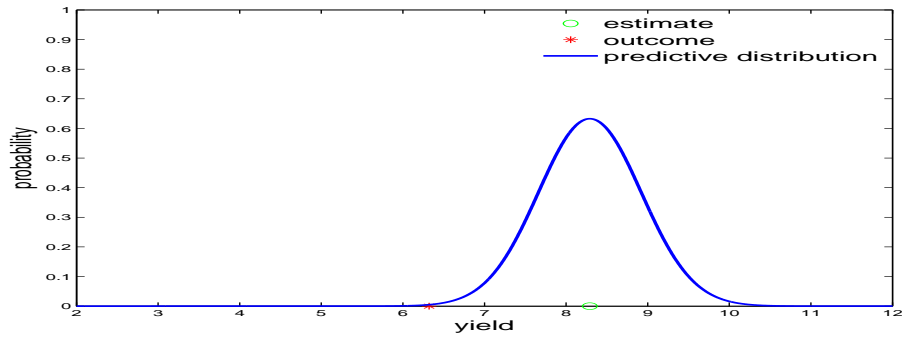
To evaluate the skill of probabilistic forecasts from the CERES-Maize model, the skill score empirical Ignorance is used, as defined in Equation 1.14. Empirical Ignorance for the CERES-Maize model by year is shown in Figure 3.3. Here, with the exception of a few forecast busts, Ignorance is quite consistent. Large empirical Ignorance scores are in 1993 when the CERES-Maize model failed to estimate the large fall in maize yield and 2009 when the CERES-Maize model underestimated the yield. The mean empirical Ignorance is 1.36 and the spread of these empirical Ignorance values,



(a) 1984



(b) 1988



(c) 1993

Figure 3.2: The forecast distribution for selected years from the CERES-Maize model. The forecast distribution is shown as a blue line, the CERES-Maize estimates are shown as green circles and the outcome is shown as a red star for the years 1984, 1988 and 1993. Notice how the outcome has a high probability mass for most years.

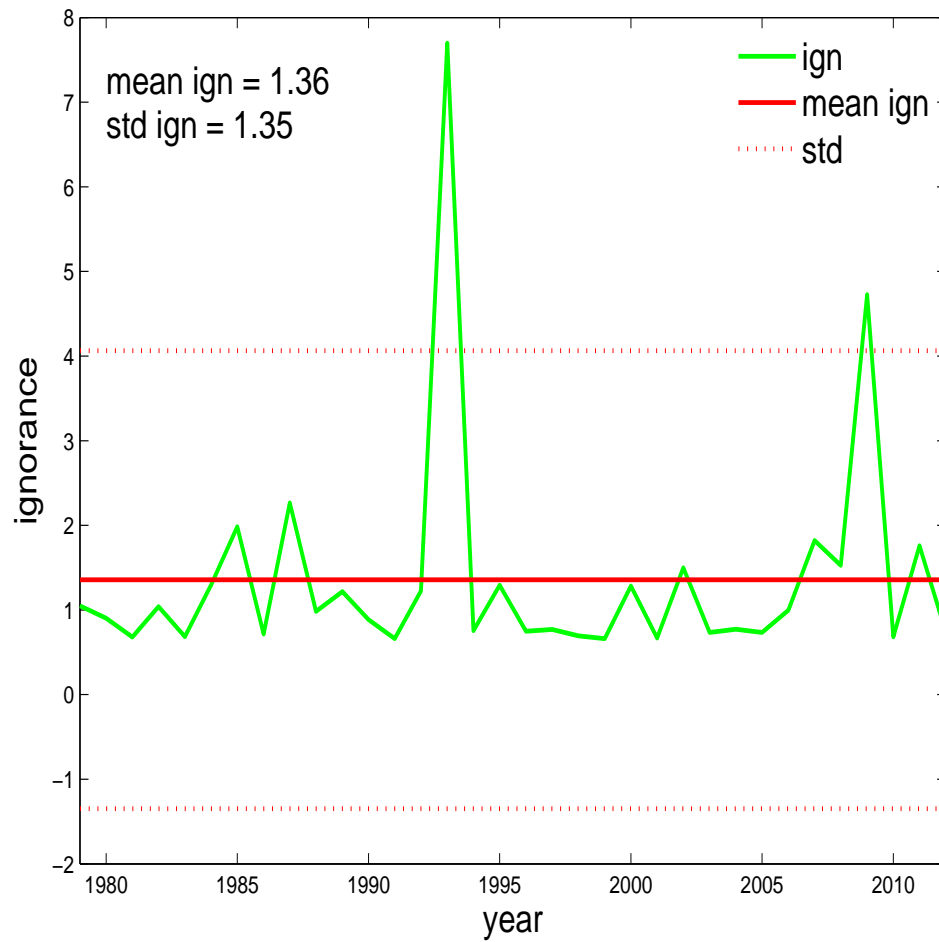


Figure 3.3: The skill of the CERES-Maize model. Empirical Ignorance is a green line, the standard deviation of the empirical Ignorance is a red dotted line and the mean empirical Ignorance is a red line. The empirical Ignorance all lie within the mean plus one standard deviations with the exception of years 1993 and 2009.

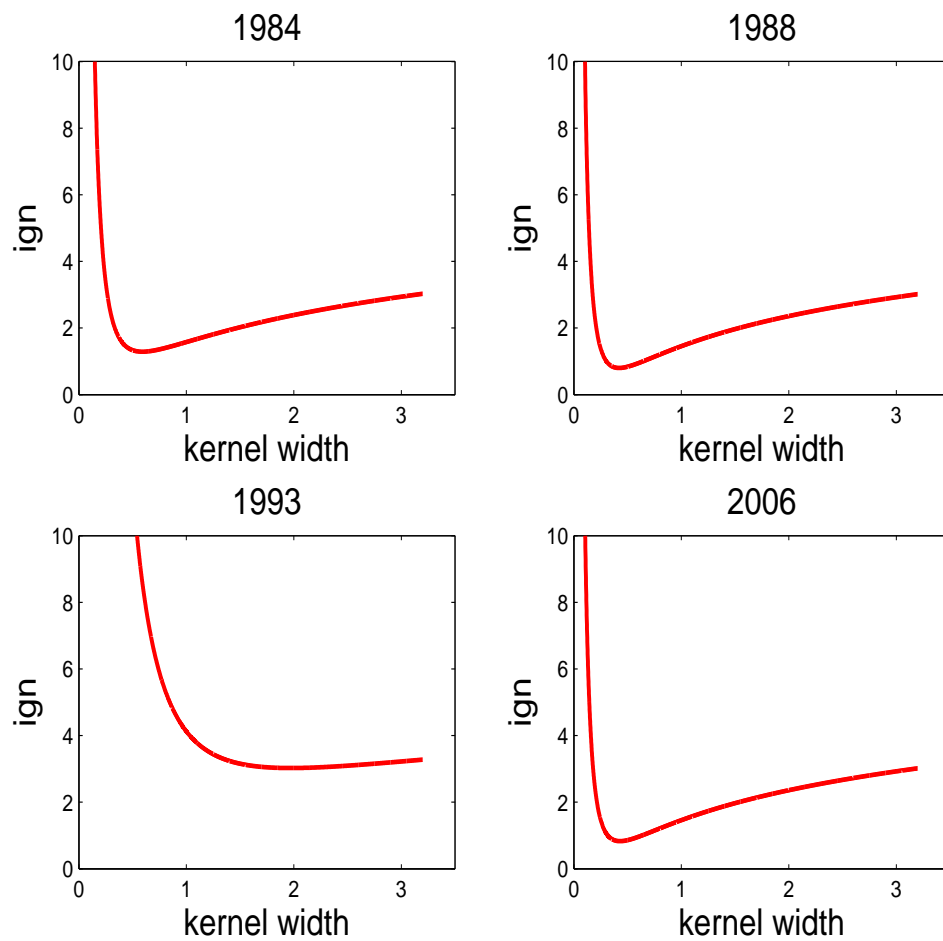


Figure 3.4: The robustness of the forecast distribution is demonstrated by comparing the empirical Ignorance by kernel width. Here four selected years for the CERES-Maize model are shown. The best kernel width changes from year to year, the kernel width for 1993 needs to be wider than the other years.

measured by standard deviation, is of a similar size 1.35.

To investigate how robust empirical Ignorance is to small changes in the kernel width, kernel width against empirical Ignorance is compared for selected years in Figure 3.4. Empirical Ignorance changes gradually unless the kernel width becomes too narrow, so small changes in the kernel width will not significantly alter the value of empirical Ignorance. To be able to compare models empirical Ignorance is measured against a bench-mark model. The selection of the bench-mark model is explained in the next section.

3.3 Creating a bench-mark model for crop modelling

In seasonal weather forecasting the standard bench mark model to use is climatology, from kernel dressing a large set of historical outcomes [24, 36]. For crop modelling climatology is not a suitable bench mark model as the yield has increased by over 500% between 1866 and 2012 [78]. This means yield levels at the start of the time series are unlikely to recur towards the end of the time series, even under adverse weather conditions. Therefore, using climatology as a bench mark model is not a suitable test of skill. To demonstrate this climatology is created by kernel dressing the historical outcomes of US maize. As the maize yield has changed so much only a short time series from 1970 to 2012 is selected. The static climatological distribution is converted into a forecast distribution by standard kernel dressing:

$$p(y : \sigma_{cl}) = \frac{1}{C\sigma_{cl}} \sum_{j=1, j \neq i}^C K\left(\frac{y - y_i}{\sigma_{cl}}\right) \quad (3.3)$$

Where C is the number of observed outcomes, here 34 and σ_{cl} is the kernel width. The Gaussian kernels, centred on y_i , are summed. The mean empirical Ignorance for climatology is 2.42 so, as expected, climatology has significantly less skill than the CERES-Maize model. The standard deviation of yield is 0.61 so the variation in empirical Ignorance is less than for the CERES-Maize model. Skill by year is shown in Figure 3.6a, empirical Ignorance is fairly flat except for the years 1980, 1983, 1988, 1993, 2004 and 2009. Instead of using climatology as the bench mark model, for crop modelling persistence is proposed, where the model's estimate is the previous year's outcome.

3.3.1 Persistence Model

The persistence model uses the previous years outcome as its estimate, so for the i^{th} year $x_i = y_{i-1}$. The singleton ensemble (x_i^1) is converted into a forecast distribution by standard kernel dressing. The kernel width chosen is the median of the standard deviation of the differences using leave-one-out cross-validation [84]. As the yield level changes over time, differences are only considered between consecutive years. The forecast distribution is:

$$p(y : \sigma) = \frac{1}{n\sigma} \sum_{i=1}^n K\left(\frac{y_i - x_i^1}{\sigma}\right), \quad (3.4)$$

where n is 34. The Gaussian kernel $K(\cdot)$ is described in Equation 3.2. The kernel width (σ) is 1.13, nearly twice the size of the CERES-Maize kernel width. The persistence model's estimate and standard deviation from the forecast distribution are shown in Figure 3.5. The one year lag from the persistence model estimates (green) to the outcomes (red) is clearly visible. The persistence model generally does well unless there is a large drop in maize yield, in these years the estimate falls outside the standard deviations. This can be seen in years 1983, 1988, 1993 and 2012. The year following a large drop is also not captured very well, as the estimate is too low.

Forecast distributions for selected years are in Figure C.1. The shape of the forecast distribution for the persistence model is significantly flatter than for the CERES-Maize as the kernel width is significantly wider. Empirical Ignorance for the persistence model is quite robust to small changes in the kernel width, as shown by the smooth changes in Figure C.2. The ideal σ should not be too sensitive to error, for example if it was slightly underestimated empirical Ignorance should not quickly approach infinity. In 2006 the estimate and the outcome were almost identical, so for this year the narrower the kernel width the better.

The skill of the persistence model by year is given in Figure 3.6b. The mean Ignorance, 2.19, shows it has less skill than the CERES-Maize model. The skill measured relative to climatology is -0.23, the negative value means the persistence model has more skill than climatology. Of note is that from 1997 onwards the persistence model has more skill than climatology for every year except 2012, a drought year. As the results show, the persistence model

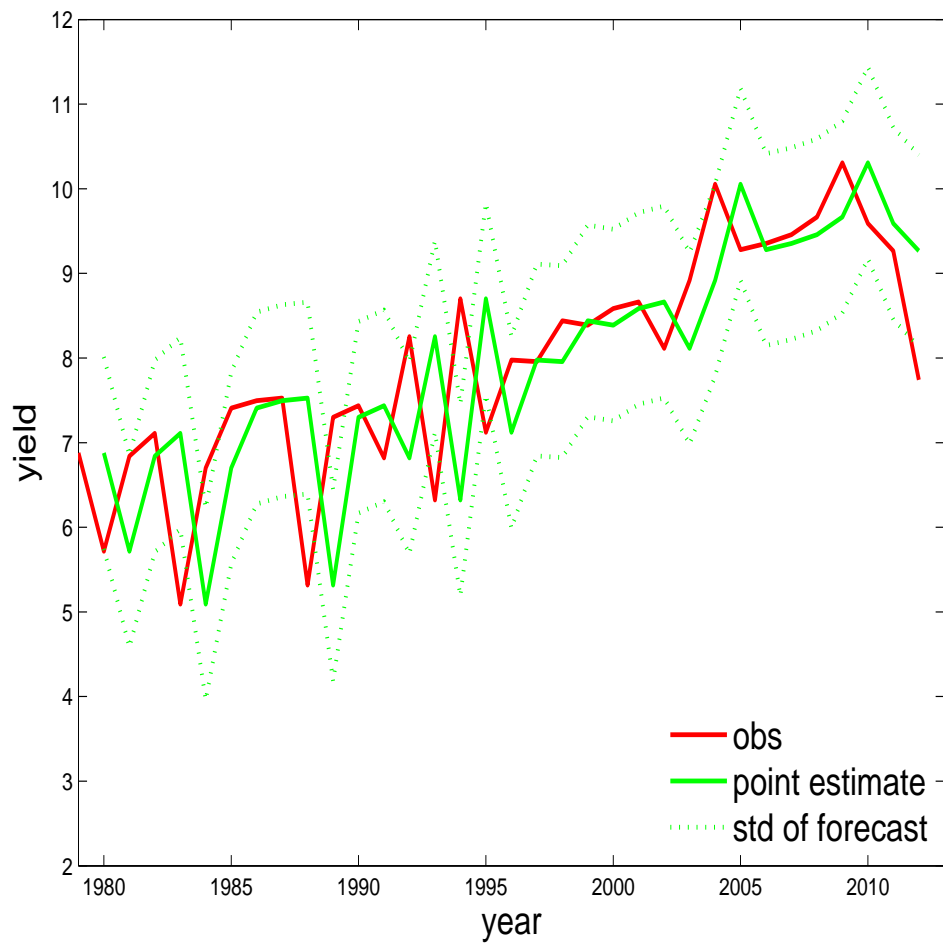


Figure 3.5: Time series of the persistence model estimate (green line) with one standard deviation of the forecast distribution (dotted green line) and outcomes (red line). By construction the persistence model misses every large fall in maize yield and the recovery the following year.

makes a better bench mark model than climatology.

3.3.2 Skill of Crop Models

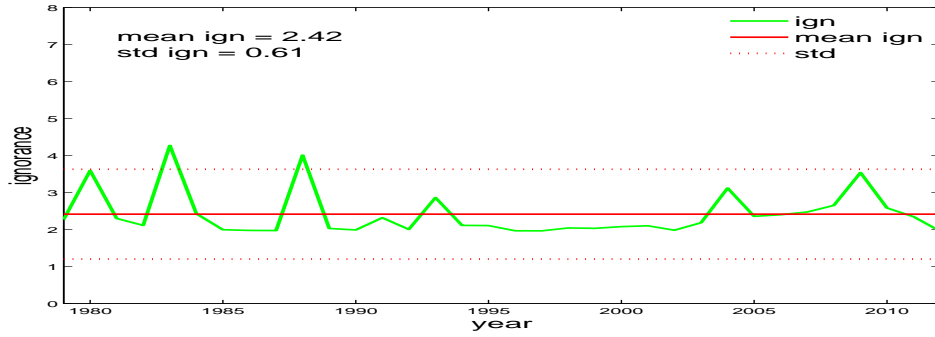
The skill of the CERES-Maize model relative to the persistence model is -0.83 and the standard deviation is 1.53, so the CERES-Maize model has more skill than persistence. The skill by year in Figure 3.7 shows that there are a few years where persistence has more skill, in particular in 1993 when the CERES-Maize model does not take into account the impact of wet rot on maize yield. In the next section three empirical models for predicting maize yield are developed.

3.4 Empirical Crop Models

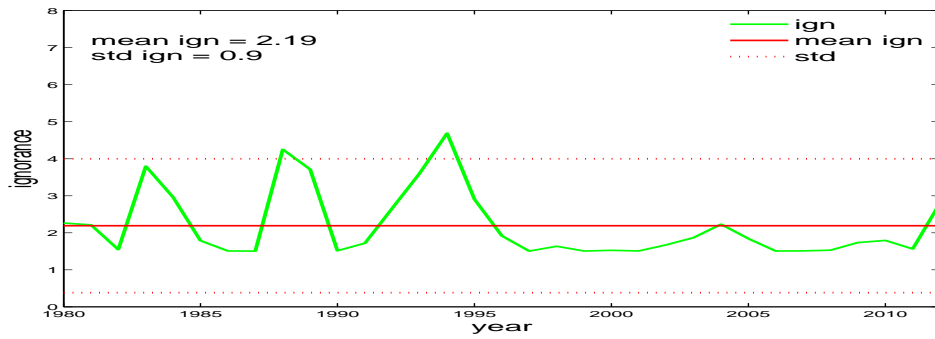
Several empirical crop models were built to provide a probabilistic forecast for maize yield for the year ahead. These empirical crop models were built solely from historical US maize yields between 1979 and 2012, their skill is measured relative to the persistence model. As all the empirical crop models are built using small outcome sets of approximately 35 points, the skill and kernel widths are estimated using leave-one-out cross-validation.

3.4.1 The Dynamic Climatology Model

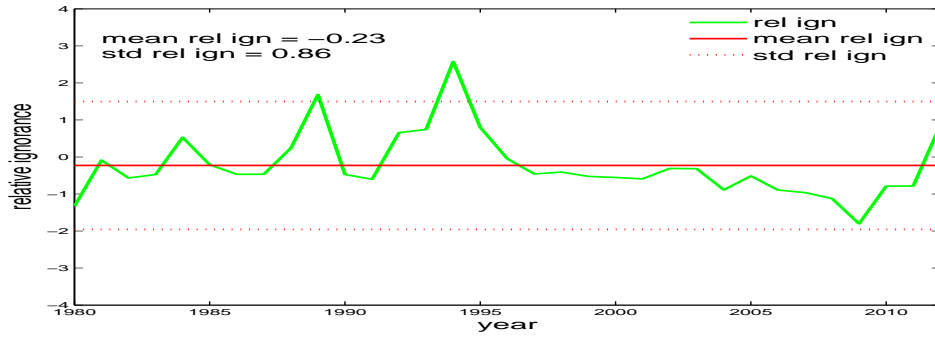
The dynamic climatology model [85] builds an ensemble of estimates using the differences between consecutive maize yield outcomes. The j^{th} ensemble member for year i , x_i^j , is made from the previous years' outcome plus the differences between outcomes:



(a) Climatology



(b) Persistence



(c) Persistence relative to climatology

Figure 3.6: Ignorance by year from (a) the climatology model, (b) the persistence model and (c) the persistence model relative to climatology. Note the relative Ignorance shows that the persistence model has more skill in the latter half of the time series, with the exception of 2012, a drought year.

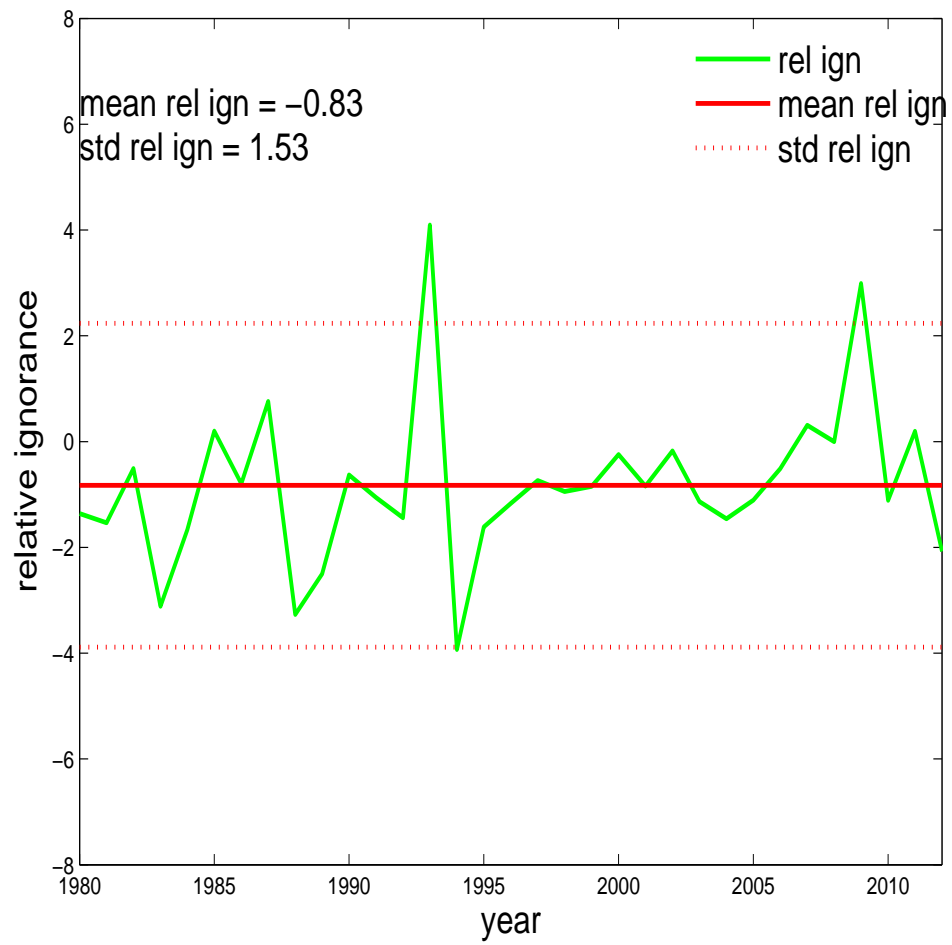


Figure 3.7: Skill of the CERES-Maize model relative to the persistence model. Negative values show the CERES-Maize model has more skill than the persistence model.

$$x_i^j = y_{i-1} + [y_{j+1} - y_j], \text{ where } j = 1, \dots, N_e, i \neq j. \quad (3.5)$$

where N_e is 33. The forecast-outcome pairs are (p_i, y_i) where p_i is the forecast distribution from kernel dressing X_i , a N_e member ensemble, $X_i = [x_i^1, \dots, x_i^{N_e}]$ at time i . The forecast distribution from standard kernel dressing is:

$$p(y : \sigma) = \frac{1}{N_e \sigma} \sum_{j=1}^{N_e} K \left(\frac{y - x^j}{\sigma} \right) \quad (3.6)$$

where the j^{th} ensemble member is replaced with a Gaussian kernel $K(\cdot)$. The kernel width, estimated from the standard deviation of the differences, is 1.13.

The ensemble of estimates is compared against the outcomes in Figure 3.8. The outcomes tend to fall in the middle of the ensemble unless there is a large downwards movement. For example in 1988, a drought year, the outcome fell just beneath the ensemble. In the year following a large drop the estimate (that the ensemble members are centred around) is based on the previous year's low outcome. The following year the outcome usually returns to previous higher yield levels, so once again the outcome falls in a low probability area. An example for this is 1989, the year after a drought, where the outcome fell in a low probability area towards the top of the ensemble.

The forecast distribution for the dynamic climatology model is shown in Figure C.3 for selected years. As the kernel width σ is significantly wider the

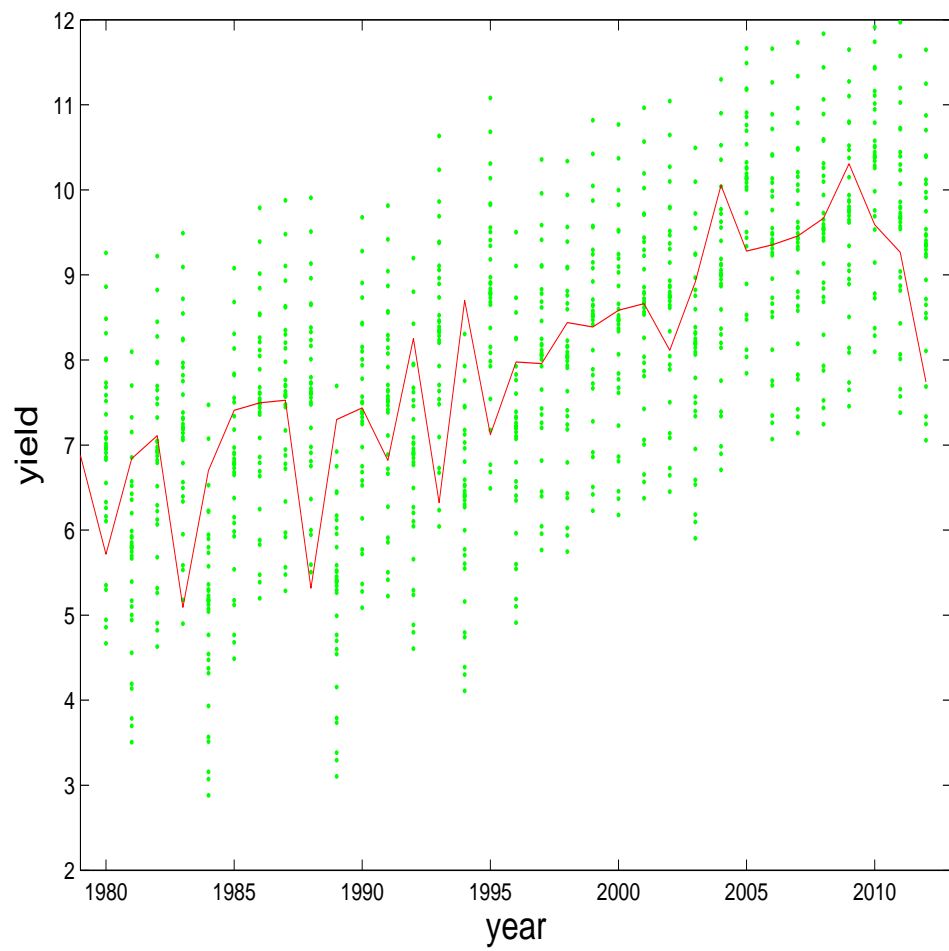
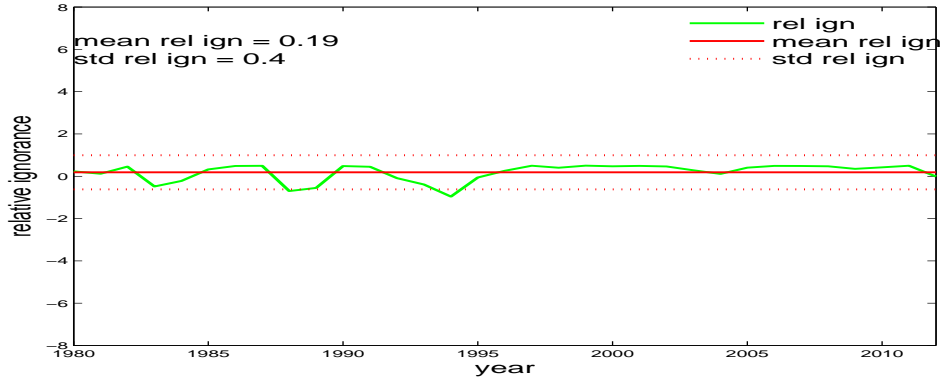


Figure 3.8: Time series of the dynamic climatology, ensemble (green dots) and outcomes (red line). The outcomes are within the ensemble with the exception of 1988 and 1994.

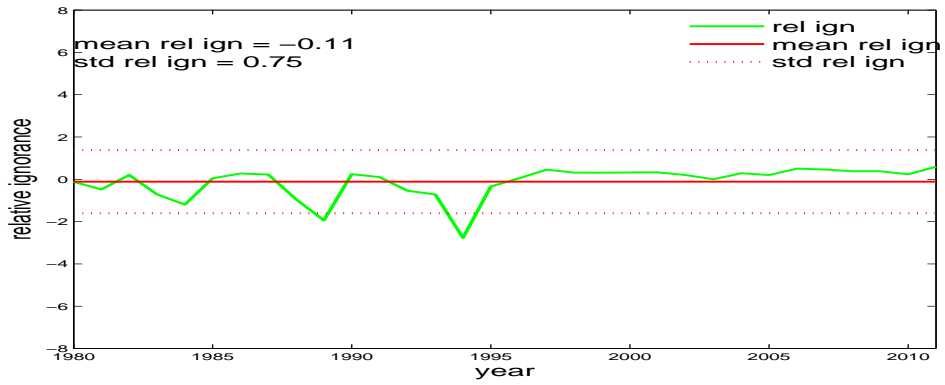
y -axis differs from Figure C.1. Compared to the forecast distribution from the CERES-Maize model in Figure 3.2, the forecast distribution has lower values and is significantly flatter. This means that even if the outcome falls in a high probability area of the forecast distribution, the probability mass on the outcome will be low. Although the outcomes in the figures always fall relatively near to the ensemble members, in 1984 and particularly in 1988 they fall in very low probability areas of the forecast distributions meaning that for these years empirical Ignorance will be high (corresponding to a poor score).

The robustness of empirical Ignorance to small changes in the kernel widths for the dynamic climatology model is shown for selected years in Figure C.4. The empirical Ignorance changes gradually with the kernel width, there are no sudden changes unless the kernel width approaches zero. The shape of empirical Ignorance by kernel width in 2006 is different from the others. From the forecast distribution plot in Figure C.3 it is seen that the outcome falls almost exactly in the highest probability area, so for this particular year the smaller the kernel width the better. Kernel width, however, is not set independently for each year, but as a best fit across all the years.

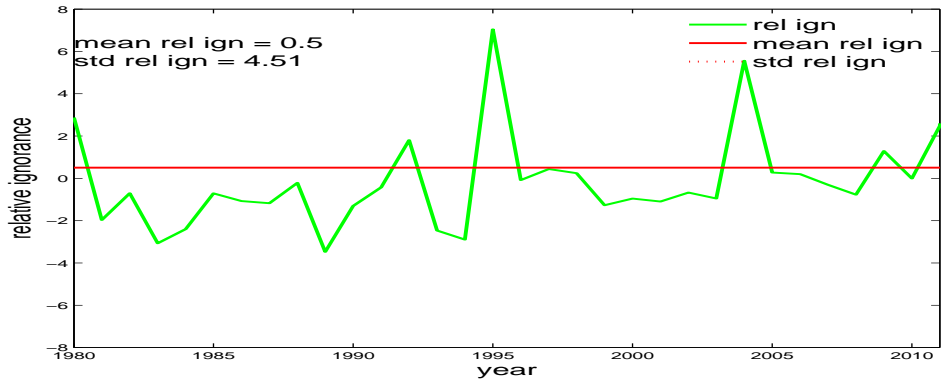
The skill of the dynamic climatology model is shown in Figure 3.9a where the mean relative Ignorance at 0.19 (in red) is higher than the mean relative Ignorance from the CERES-Maize model in Figure 3.7, but the standard deviation is significantly lower. The dynamic climatology model has less skill than the CERES-Maize model but lower standard deviation, so there is less variability in relative Ignorance.



(a) Dynamic Climatology



(b) Ratio



(c) Asymmetric

Figure 3.9: Skill of the dynamic climatology model, the ratio model and asymmetric model relative to the persistence model. The ratio model (b) has more skill than the other models.

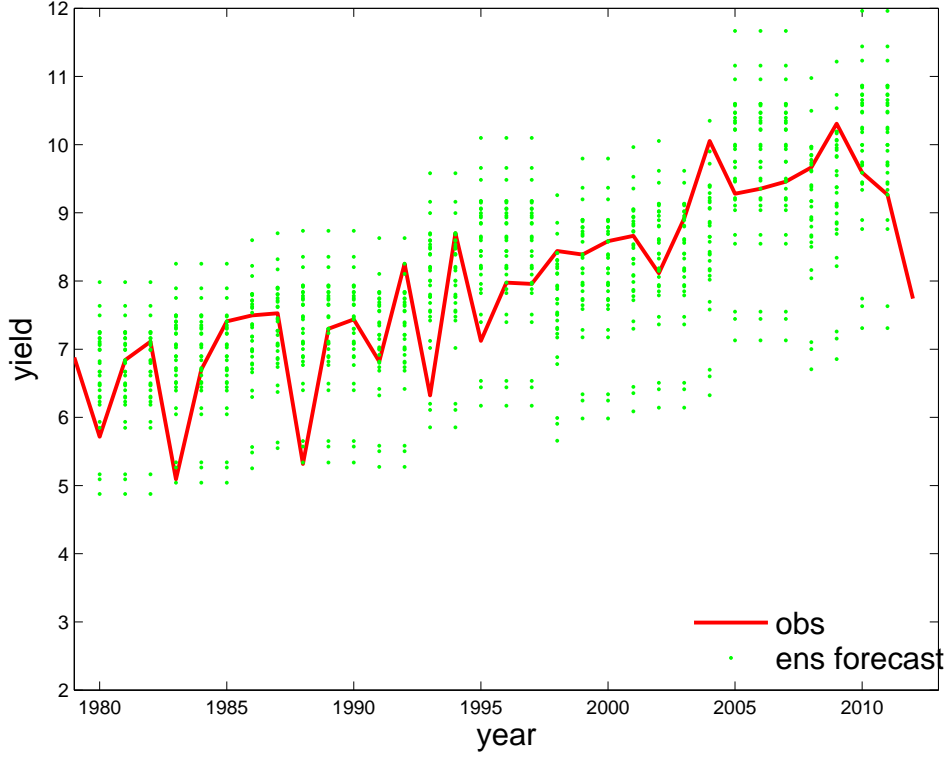


Figure 3.10: Time series of the ratio model ensemble (green) and outcomes (red). The ensemble members are not evenly spread across the yield axis, there is a gap between the lowest ensemble members and the others.

3.4.2 Ratio Model

The ratio model creates an ensemble of estimates based around the rolling maximum of the three previous years outcomes z_i :

$$z_i = \max[y_p], \text{ where } p = i - 1, i - 2, i - 3 \quad (3.7)$$

When $i \leq 3$ the maximum is taken over the number of years available. As an example for 1987 the ratio model estimate is 7.32 which is the maximum of

the three previous outcomes 6.11, 6.55 and 7.32. The ensemble is created by multiplying the rolling maximum by a series of ratios r_i , using the outcome from the year ahead:

$$r_i = \frac{y_{i+1}}{z_i} \quad (3.8)$$

$$x_i^j = [z_i r_j], \text{ where } j = 1, \dots, N_e, i \neq j, \quad (3.9)$$

where N_e , the number of ensembles, is 33. There is no ensemble for 2012 as there is no outcome for 2013 to calculate the ratio from (Equation 3.8). Several years have very similar probabilistic forecasts as their ensembles are so similar, for example the years 1983, 1984 and 1985. The outcomes usually fall within the ensemble as shown in Figure 3.10, unless there is a year with adverse weather. The year after adverse weather the outcomes are again contained within the ensemble, unlike the dynamic climatology model.

The forecast distribution of the ratio model in Figure C.5 is much flatter than the CERES-Maize model as the kernel width is much wider. The ensemble is not as wide as the ensemble from the dynamic climatology model in Figure C.3. Although both the ratio model and the dynamic climatology model have identical kernel widths (1.13) the ensemble members of the ratio model are not as spread out, so the forecast distribution is not as flat. Small changes in the kernel width do not significantly effect the value of empirical Ignorance except in the case where the kernel width is approaching zero (Figure C.6.) Years such as 1988 and 1993 where severe adverse weather

negatively impacted the yield are expected to require wider kernel widths than average. The kernel width is selected as a best fit across all the years rather than one individual year.

The skill of the ratio model by year is shown in Figure 3.9b. Compared to dynamic climatology there are larger negative values and a higher standard deviation. The skill is mean relative Ignorance (with standard deviation of 0.75), so the ratio model has more skill than the persistence and dynamic climatology models, but less skill than the CERES-Maize model.

3.4.3 Asymmetric Model

In the maize yield time series (red) in Figure 3.11b downward movements, caused by adverse weather conditions, are much greater than upward movements. The exception is when the upward movement is a retracement from a large fall the previous year. Although adverse weather has a large negative impact on yield, this normally only impacts the yield for one year. The asymmetric model estimate is the rolling maximum of the three previous years outcomes (Equation 3.7). This is converted into a forecast distribution by kernel dressing using an asymmetric kernel, where the kernel is larger for negative falls in yield than for positive increases.

To construct the asymmetric kernel, errors between the model estimates and outcomes are calculated. A Gaussian mixture model [14] is fitted to these errors using leave-one-out cross-validation. The parameters of the Gaussian mixture model are chosen using the EM algorithm [16]. The Gaussian mixture model is a combination of two Gaussians:

$$\begin{aligned} Y_1 &\sim N(\mu_1, \sigma_1^2), \\ Y_2 &\sim N(\mu_2, \sigma_2^2), \\ Y &= (1 - \Delta)Y_1 + \Delta Y_2, \end{aligned}$$

where Δ is either 0 or 1 and $P(\Delta = 1) = \omega$. Let $\phi_\theta(y)$ be the Gaussian density with parameters $\theta = (\mu, \sigma^2)$. The parameters are fitted using maximum likelihood:

$$\theta = (\omega, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \quad (3.10)$$

Given the training set $Z = y_1, \dots, y_n$. The log-likelihood is:

$$\ell(\theta; Z) = \sum_{i=1}^n \log[(1 - \omega)\phi_{\theta_1}(y_i) + \omega\phi_{\theta_2}(y_i)] \quad (3.11)$$

As Δ are unknown, they are considered latent variables Δ_i taking values of either a 0 or 1. If $\Delta_i = 1$ then Y_i was from model 2, otherwise it came from model 1. The log-likelihood then becomes:

$$\begin{aligned} \ell_0(\theta; Z, \Delta) &= \sum_{i=1}^n [(1 - \Delta_i) \log \phi_{\theta_1}(y_i) + \Delta_i \log \phi_{\theta_2}(y_i)] \\ &\quad + \sum_{i=1}^n [(1 - \Delta_i) \log \omega + \Delta_i \log(1 - \omega)] \end{aligned} \quad (3.12)$$

where the maximum likelihood estimates of μ_1 and σ_1^2 are the sample mean

and variance when $\Delta_i = 0$. For μ_2 and σ_2^2 the sample mean and variance are when $\Delta_i = 1$. As the values are unknown the EM Algorithm is used:

1. An initial guess for the parameters $\hat{\mu}_1, \hat{\sigma}_1^2, \hat{\mu}_2, \hat{\sigma}_2^2$ and $\hat{\omega}$
2. Expectation Step: calculate the responsibility of model 2 for observation i $E(\Delta_i|\theta, Z) = P(\Delta_i = 1|\theta, Z)$.

$$\hat{\gamma}_i = \frac{\hat{\omega}\phi_{\hat{\theta}_2}(y_i)}{(1 - \hat{\omega})\phi_{\hat{\theta}_1}(y_i) + \hat{\omega}\phi_{\hat{\theta}_2}(y_i)}, i = 1, \dots, n. \quad (3.13)$$

3. Maximization Step: weighted means and variances are calculated

$$\begin{aligned} \hat{\mu}_1 &= \frac{\sum_{i=1}^n (1 - \hat{\gamma}_i)y_i}{\sum_{i=1}^n (1 - \hat{\gamma}_i)}, \\ \hat{\sigma}_1^2 &= \frac{\sum_{i=1}^n (1 - \hat{\gamma}_i)(y_i - \hat{\mu}_1)^2}{\sum_{i=1}^n (1 - \hat{\gamma}_i)}, \\ \hat{\mu}_2 &= \frac{\sum_{i=1}^n \hat{\gamma}_i y_i}{\sum_{i=1}^n \hat{\gamma}_i}, \\ \hat{\sigma}_2^2 &= \frac{\sum_{i=1}^n \hat{\gamma}_i (y_i - \hat{\mu}_2)^2}{\sum_{i=1}^n \hat{\gamma}_i} \end{aligned}$$

where the weight is $\hat{\omega} = \sum_{i=1}^n \frac{\hat{\gamma}_i}{n}$.

4. Iterate until convergence

The asymmetric kernel dressing parameters are chosen as the median of the n kernel dressing parameters calculated using leave-one-out cross-validation. The EM algorithm used to estimate the mixed Gaussian parameters is implemented using Matlab 2011a [48]. The results are shown in Figure 3.11a,

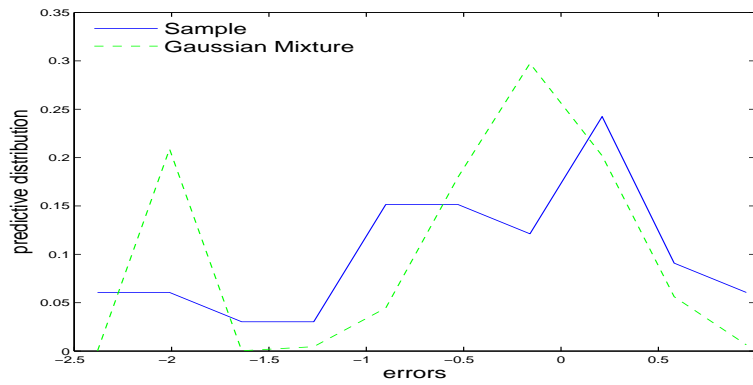
Table 3.2: Gaussian mixture parameters for the asymmetric model

Gaussian mixture	μ	kernel width	weight
Kernel 1	-0.16	0.36	0.87
Kernel 2	-2.1	0.09	0.14

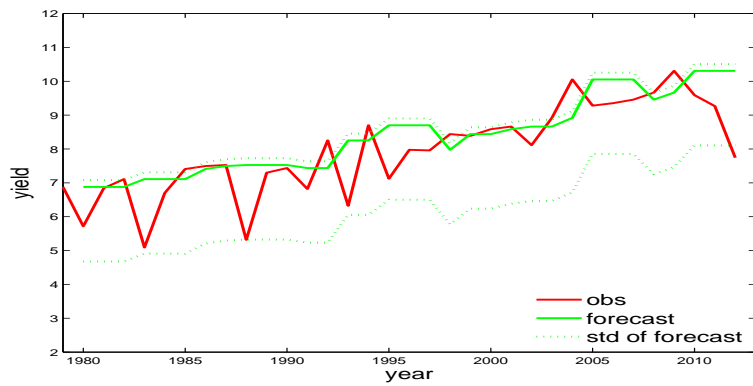
where the forecast distribution from both the errors and the Gaussian mixture are normalised. Most of the forecast distribution from the mixed Gaussian (in green) is centred around the smaller errors, however there is a second smaller Gaussian centred on the larger errors. The probability from Kernel 1 has the highest weight of 0.87. The kernel width of Kernel 2, furthest away from the asymmetric model estimate, is surprisingly narrow at 0.09. The forecast distribution is:

$$p(y|u_1, \sigma_1, u_2, \sigma_2, \omega) = \frac{\omega}{n\sigma_1} \sum_{i=1}^n K\left(\frac{y_i - x_i - \mu_1}{\sigma_1}\right) + \frac{(1 - \omega)}{n\sigma_2} \sum_{i=1}^n K\left(\frac{y_i - x_i - \mu_2}{\sigma_2}\right) \quad (3.14)$$

The asymmetric model estimate and the standard deviation of the forecast distribution are shown in Figure 3.11b. By design the asymmetric kernel causes the standard deviation (green dotted line) of the forecast to be larger below the model estimate than above the model estimate. The asymmetric kernel seems a bit too narrow, quite a few positive outcomes fall outside one standard deviation such as the years 1992, 1994, 2004 and 2009. Bootstrap resampling, however, may provide a better estimate as there are so few points.



(a) Asymmetric kernel



(b) Time series

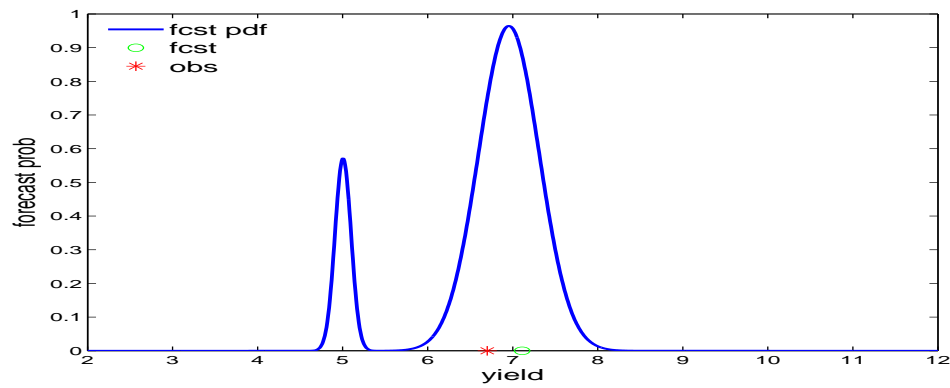
Figure 3.11: Fig 3.11a: The forecast distribution of the asymmetric kernel (green) and sample (blue). Note the second smaller kernel for negative falls. Fig 3.11b. Notice the asymmetric kernel is wider at the bottom than the top.

The shape of the asymmetric kernel is shown in Figure 3.12 for selected years, the two separate Gaussians from the mixed Gaussian can clearly be identified. There is a large area between the two Gaussians which has very low probability, if an outcome fell between the two Gaussians it would have a very low probability mass. In 1988, a drought year, the outcome (red star) falls into the smaller Gaussian kernel and so has a higher probability mass than the other empirical models.

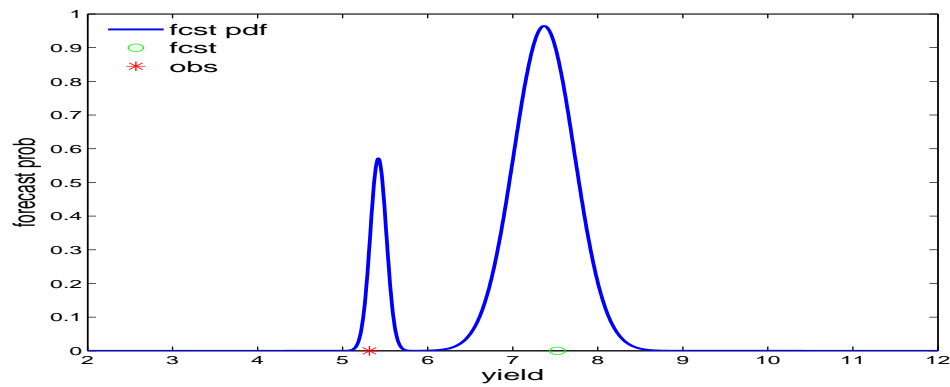
The behaviour of the empirical Ignorance if the kernel width changed slightly was explored in Figure 3.13. Empirical Ignorance is quite robust to small movements in the kernel width.

The skill for the asymmetric model is shown in Figure 3.9c. With a mean relative Ignorance of 0.50 and a standard deviation of 4.51 it is the empirical crop model with the least skill. In 1980 and 1995 the outcome fell in the low probability area between the two mixed Gaussians and in 2004 the outcome fell outside the the two mixed Gaussians.

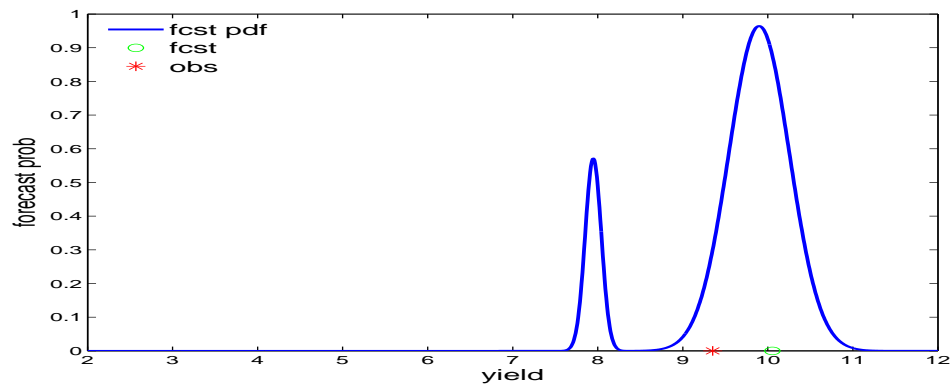
Although the asymmetric model's kernel dressing parameters were chosen using the EM algorithm, the skill of the model is assessed by empirical Ignorance. The effect on skill if the kernel dressing parameters are altered is explored. If σ_1 for Kernel 1 is changed, while all the other parameters are held constant, is in Figure 3.14a. The effect on empirical Ignorance on changing σ_1 was very small, it drifts downwards as the width of the kernel increases. The skill of the asymmetric model if σ_2 (for Kernel 2) was changed is shown in Figure 3.14b. The narrower σ_2 , the more skill the asymmetric model has. Next the weight (ω) was examined in Figure 3.14c. Here the



(a) 1984



(b) 1988



(c) 2006

Figure 3.12: The estimated probabilistic density function by year for the asymmetric model. The two separate Gaussians are clearly identified and there is an area of low probability between them.

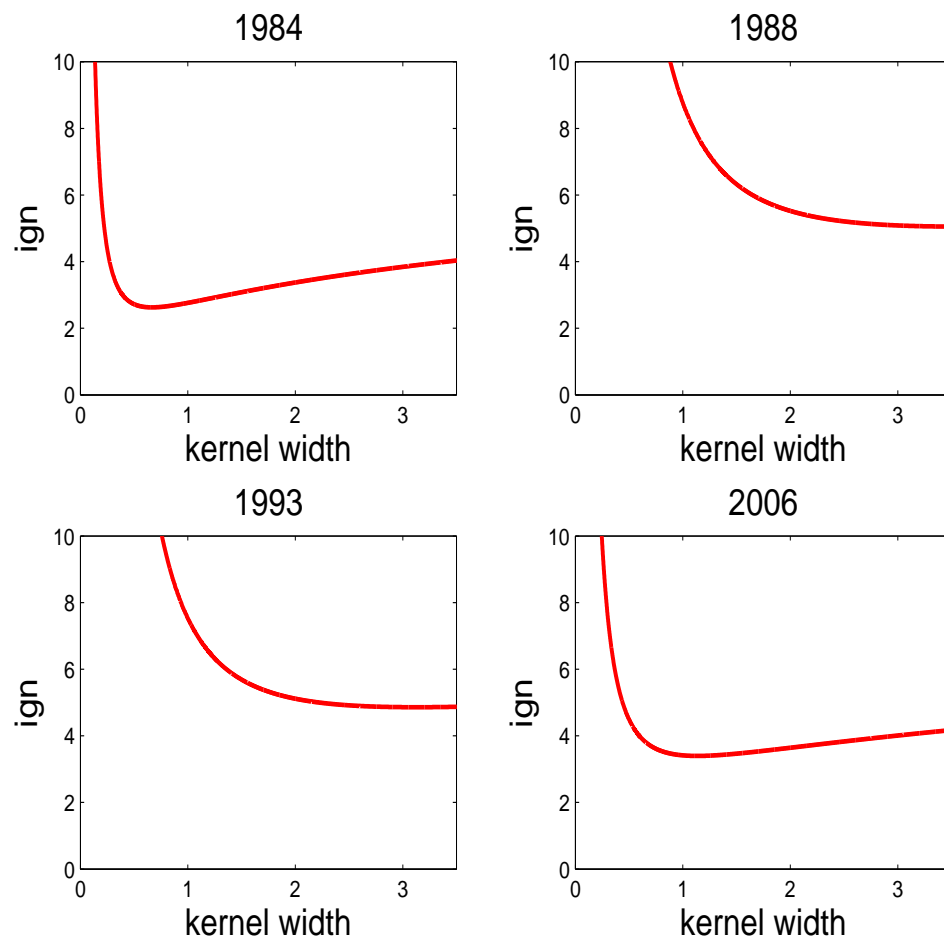


Figure 3.13: The kernel width by year for the asymmetric model. Unless the chosen kernel width is too narrow, the empirical Ignorance is quite robust.

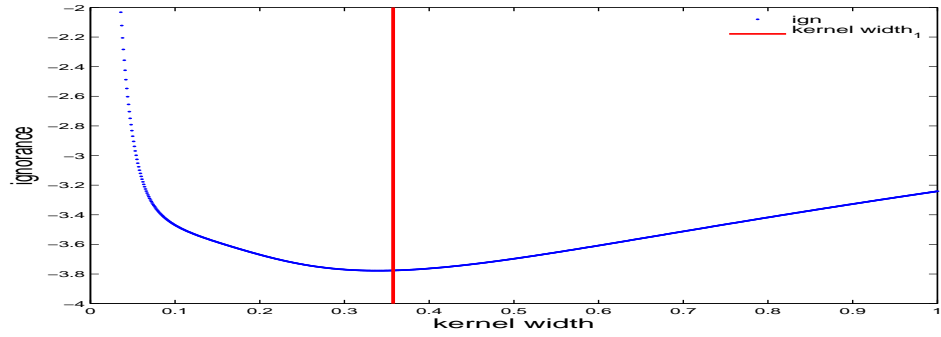
smaller ω became, the more skill the asymmetric model had. It was found that adding an offset to the asymmetric kernel improved the skill, but this is beyond the scope of this thesis.

3.4.4 Gamma Model

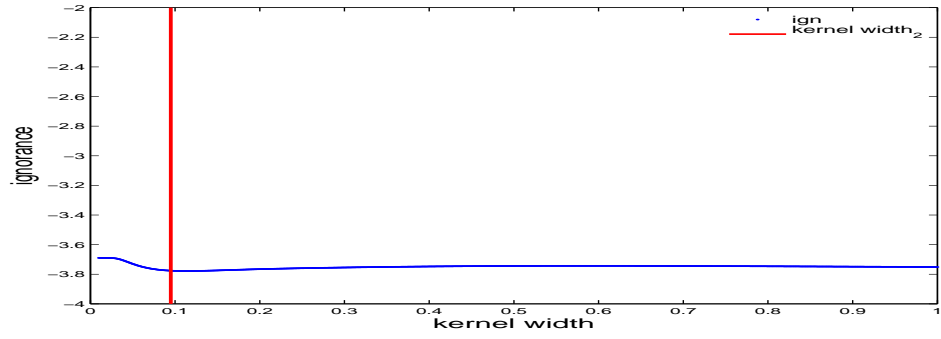
As the mixed Gaussian model has an area of low probability between the two kernels, a different asymmetric kernel model is tested using the Gamma distribution as the kernel. For the Gamma model, the singleton ensemble is from a one step ahead linear regression. This estimate is converted into a forecast distribution by kernel dressing using a negative Gamma distribution. A negative Gamma distribution is used to capture the very low yields in years with adverse weather conditions. The linear regression and Gamma distribution parameters are simultaneously fitted using maximum likelihood estimation. For this a generalized linear model (glm) is implemented in R, with the Gamma family link set as inverse [65]. This means the generalized linear model formula for the mean becomes:

$$E(Y_i) = \frac{1}{a + bu_i} \quad (3.15)$$

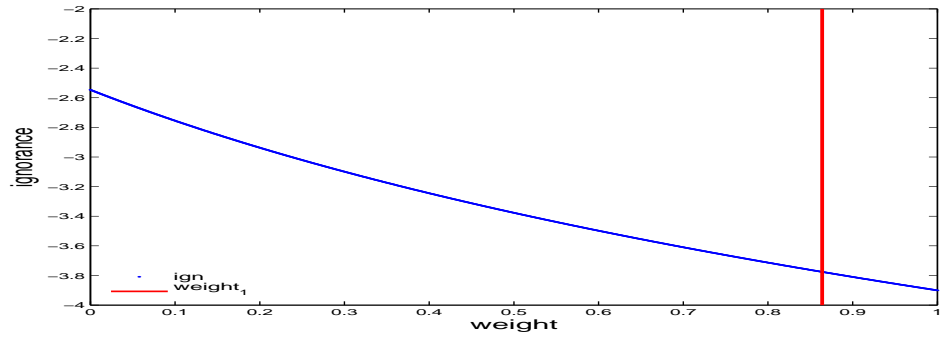
where a is the intercept, b is the parameter, $E(Y_i)$ is expected yield and u_i is the year from $i = 1, \dots, n$. To fit a negative Gamma distribution using the generalized linear model the following ad hoc method is used. Constant C_0 is added and the yield subtracted, so the mean becomes:



(a) σ_1



(b) σ_2



(c) ω

Figure 3.14: The skill of the asymmetric model when 3.14a σ_1 from the 1st mixed Gaussian is altered, 3.14b when σ_2 from the 2nd mixed Gaussian is altered and 3.14c the weight (ω) is altered. There are no sudden jumps in empirical Ignorance for the three kernel dressing parameters.

$$C_0 - E(Y_i) = \frac{1}{a + bu_i} \quad (3.16)$$

To examine what value C_0 should be, the Gamma model used a range of values. How well the model with a given constant C_0 fits the data is determined in an ad hoc way by the Akaike's Information Criterion (AIC) [1]. The results are shown in Figure 3.15, the lowest AIC, and therefore the better fit, is when C_0 is 21 or 22. The AIC increases if C_0 becomes too small or too large. Setting C_0 to approximately twice the maximum yield provides the best fit. From now on C_0 is simply taken as a constant.

The parameters estimated using maximum likelihood estimation for the mean are:

$$21 - E(Y_i) = \frac{1}{-1.1823774 + 0.0006311u_i} \quad (3.17)$$

To remove C_0 from the formula and calculate negatively distributed Gamma, Equation 3.17 is rearranged as:

$$E(Y_i) = - \left(\frac{1}{-1.1823774 + 0.0006311u_i} - 21 \right) \quad (3.18)$$

The $E(Y_i)$ is the expected value of the negative gamma distribution, see Equation 3.19 where α is the shape ($\alpha > 0$) and β_i is the rate ($\beta > 0$) which changes for each year i . While fitting the parameters in Equation 3.17, α is also calculated as 306.4197[†]

[†]The glm function in R only provides an approximation for α . A more accurate measurement of α is found using the glm summary information and the R function

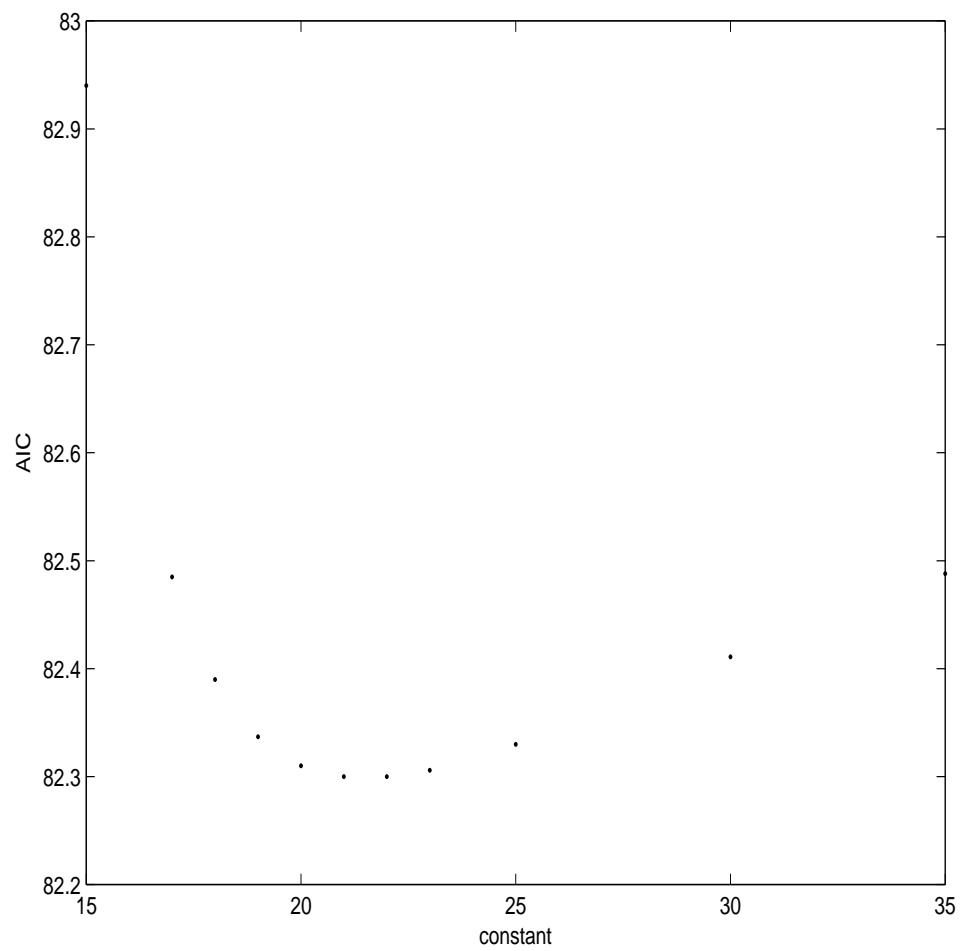


Figure 3.15: Value of AIC by C_0 . Note that when C_0 is 21 or 22 the AIC is lowest.

$$E(Y_i) = \alpha\beta_i \quad (3.19)$$

The β_i parameter for year i is then calculated by rearranging 3.19 and substituting:

$$\beta_i = \frac{E(Y_i)}{\alpha} \quad (3.20)$$

$$\beta_i = -\frac{1}{\alpha} \left(\frac{1}{-1.1823774 + 0.0006311u_i} - 21 \right) \quad (3.21)$$

The variance is calculated as:

$$Var(Y_i) = \alpha\beta_i^2 \quad (3.22)$$

$$std(Y_i) = \sqrt{\alpha}\beta_i \quad (3.23)$$

A comparison between the Gamma model forecasts and observations is presented in Figure 3.16. In years with extreme weather the yields fall outside twice the standard deviation of the probabilistic forecast, for example in the drought years 1988 and 2012.

The fitted distribution function of the yield for selected years are shown in Figure 3.17. As the α parameter for the Gamma distribution is so high the kernel is quite similar to a Gaussian kernel. In years where there is a large fall in yield, the probability mass on the outcome is very low as in Figure 3.17b.

gamma.shape [91].

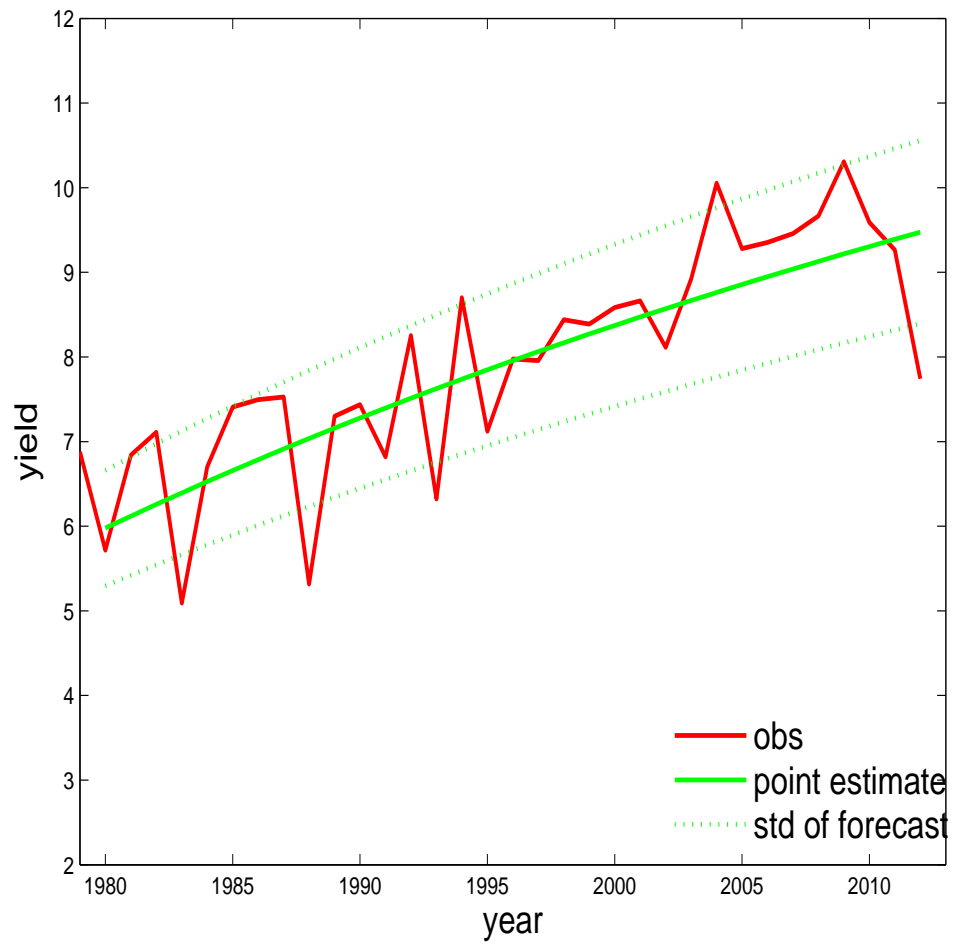
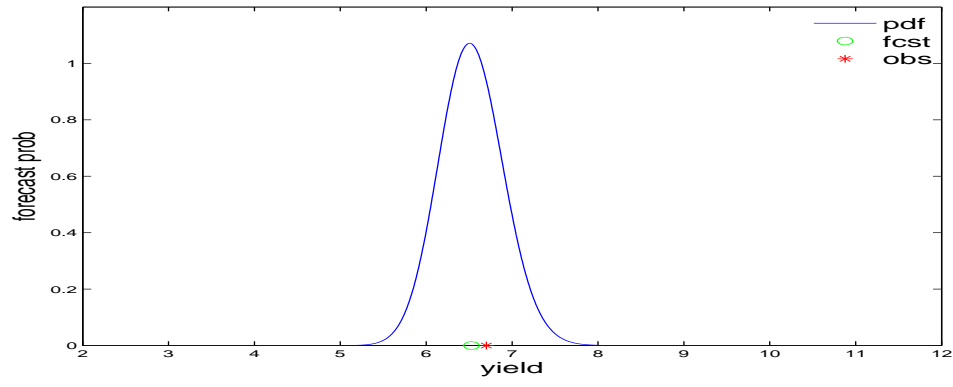
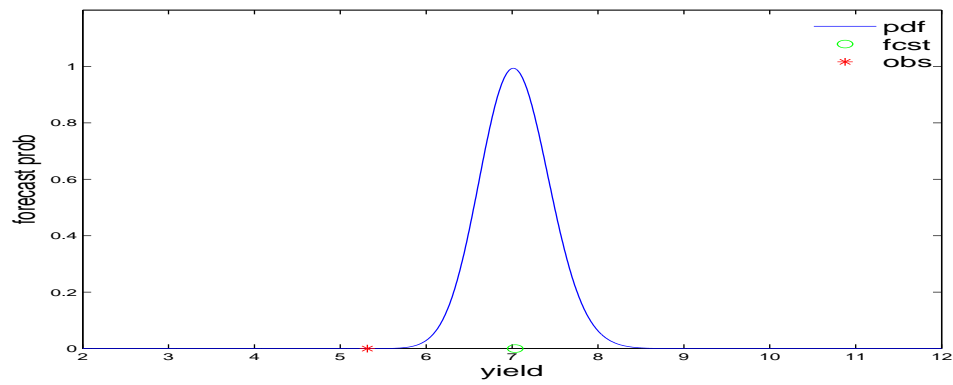


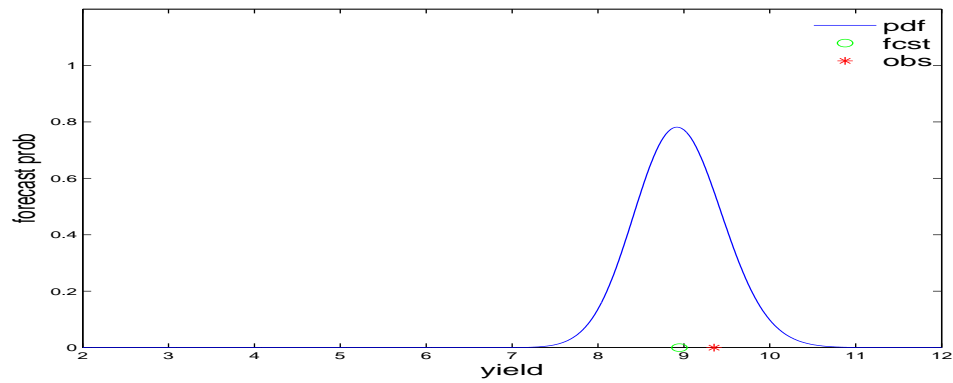
Figure 3.16: The Gamma model forecast (green) and the observations (red). In years with adverse weather such as a drought in 2012 the observation falls outside the standard deviation (dotted green line).



(a) 1984



(b) 1988



(c) 2006

Figure 3.17: The estimated probabilistic density function by year for the gamma model. As the α parameter is high the kernel shape is quite similar to a Gaussian kernel.

Model	Basis	Kernel dressing params
CERES-Maize	errors ($y_i - x_i$)	$\sigma_i = 0.6305 \pm 0.0001$
Persistence	differences ($y_{i+1} - y_i$)	$\sigma=1.13$
Dynamic Clim	differences ($y_{i+1} - y_i$)	$\sigma=1.13$
Ratio	differences ($y_{i+1} - y_i$)	$\sigma=1.13$
Asymmetric	errors ($y_i - x_i$)	$\sigma_1 = 0.09, u_1 = -2.1,$ $\sigma_2 = 0.36, u_2 = -0.16$

Table 3.3: A summary of kernel width (σ) by model.

3.5 Estimating the kernel width

Kernel widths (σ) are chosen as the standard deviation across a set of errors (or differences). Using leave-one-out cross-validation creates a set of σ estimates and the model used the median across this set. A summary of kernel widths by crop model is in Table 3.3. In seasonal weather forecasts kernel dressing parameters were chosen by minimising Ignorance [18], discussed in Section 1.10. In this section, these two different methods to choose the kernel widths are compared for crop models.

Empirical Ignorance by kernel width is shown in Figure 3.18 as blue dots. For all the crop models, as the kernel width becomes narrower, the skill increases until eventually the kernel width becomes too narrow. At this point empirical Ignorance becomes unstable, rapidly moving towards infinity with only a small decrease in kernel width. To compare the two methods the kernel width set by the standard deviation of the errors (or differences) is shown as a red dotted line and the kernel width set by minimising Ignorance

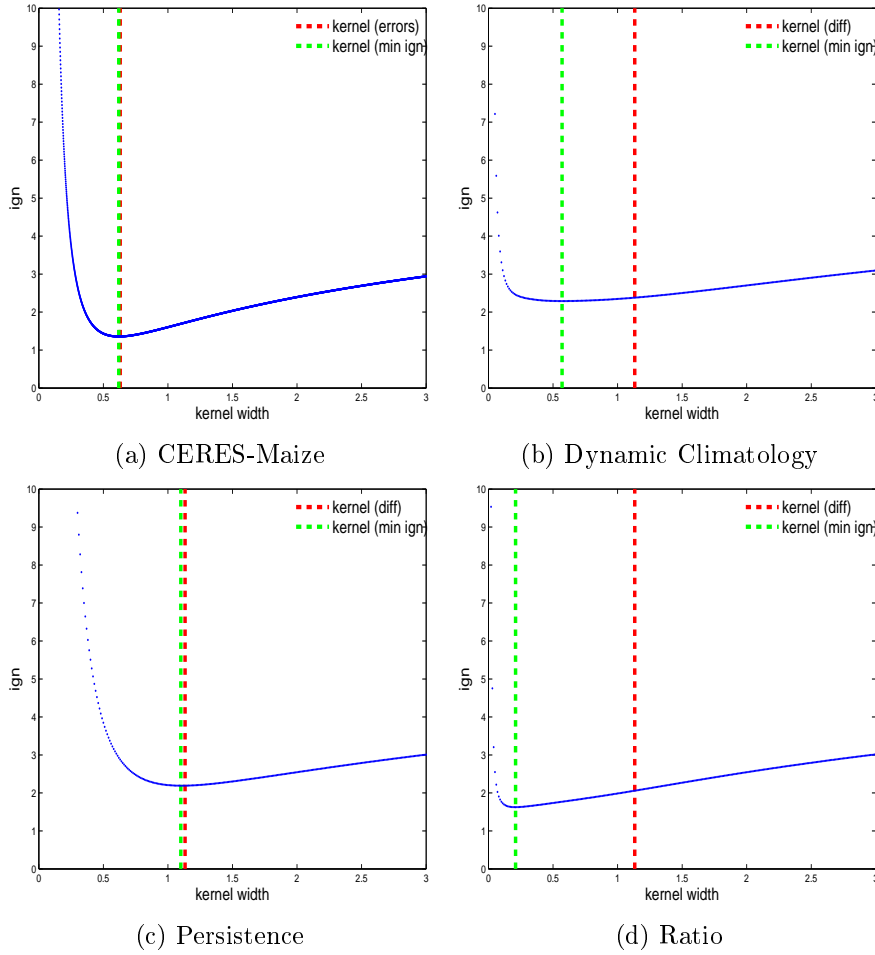


Figure 3.18: The empirical Ignorance by kernel width is shown in blue. The kernel width, from the standard deviation of the differences (or errors), is in red. The kernel width from minimising Ignorance is in green. Minimising Ignorance has the narrowest kernels, and so the most skill, for all the empirical models.

is shown as a green dotted line. For the CERES-Maize model in Figure 3.18a both methods choose very similar sized σ , across all the crop models this is the narrowest σ . With the persistence model, Figure 3.18c, both kernel widths are again close although when σ is chosen by minimising Ignorance it is slightly smaller. For both the dynamic climatology model and the ratio model in Figures 3.18b and 3.18d the kernel from minimising Ignorance is significantly narrower.

3.6 Kernel dressing methods and skill

The skill of probabilistic forecasts, where the kernel width was chosen by different methods, was determined for each model by considering the mean empirical Ignorance and the 5% to 95% bootstrap resampling intervals of the empirical Ignorance. The x -axis shows empirical Ignorance when kernel widths are chosen by minimising Ignorance. The y -axis shows Ignorance when kernel widths are chosen by the standard deviations of the errors (or differences). Both methods use leave-one-out cross-validation with the kernel width used by the model, the median of the n estimates.

The model with the most skill for both kernel width selection methods is the CERES-Maize model (green). As the CERES-Maize falls directly on the diagonal line there is little difference in skill between the two methods. The same is true for the persistence model (red), which also crosses on the diagonal line. All the other models fall above the line, so choosing the kernel width by minimising Ignorance provides more skill. The order of model skill remains consistent across both methods for choosing the kernel width.

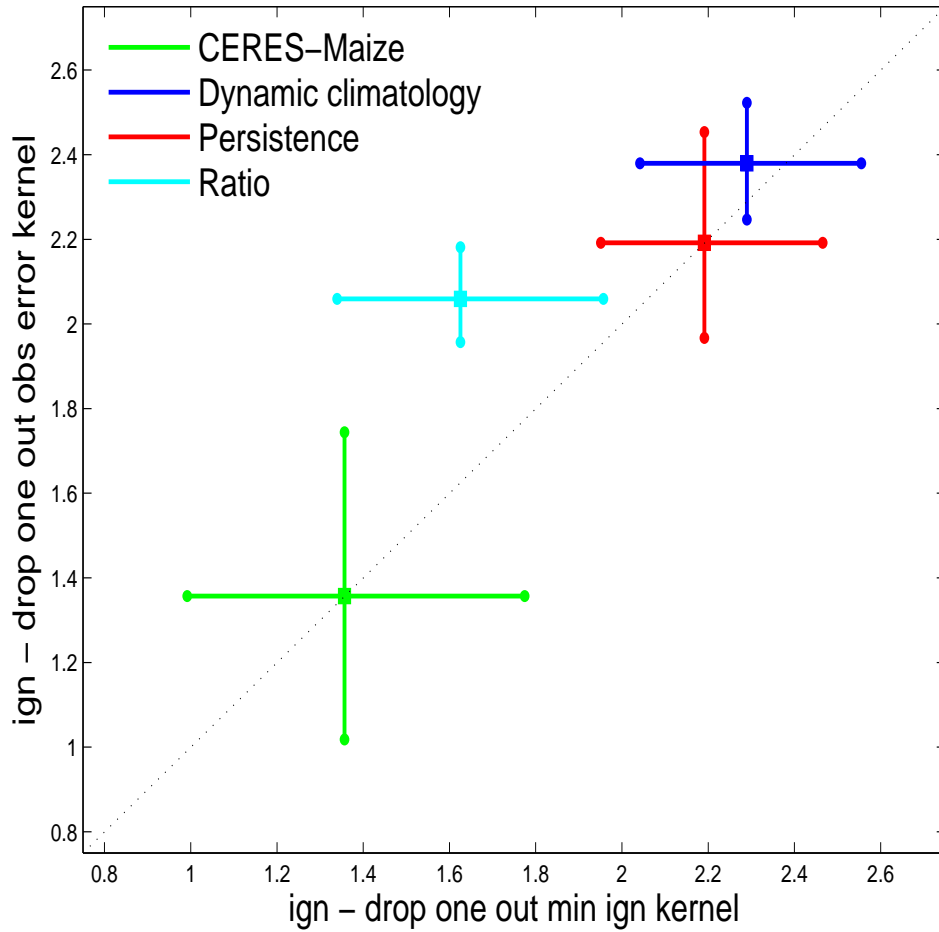


Figure 3.19: Comparison of skill when the kernel widths are chosen by different methods. Ignorance and 5 to 95% bootstrap resampling intervals when kernel width is chosen using the drop one out method by minimising Ignorance (x -axis) and when the kernel width is chosen using drop one out errors (y -axis). Notice that CERES-Maize is the model with the most skill, irrelevant of how the kernel widths are chosen.

The ratio model (cyan) and the dynamic climatology model (blue) cross above the diagonal line. For these empirical models choosing kernel widths by minimising Ignorance produces a more skillful forecast. The bootstrap resampling intervals, however, are much wider when minimising Ignorance is used.

Skill relative to persistence when the kernel widths are chosen by minimising Ignorance, with the exception of the asymmetric and gamma model, are shown in Figure 3.20. The ratio model has significantly more skill than the other empirical models. The bootstrap resampling interval for the asymmetric model is significantly larger than the other models. In the next section it is examined whether the skill of probabilistic forecasts is improved by including an empirical model in a multi-model forecast.

3.7 Multi-model crop forecasts

Does a probabilistic forecast from a multi-model crop model have more skill than a single crop model? Here results are examined when two crop model forecasts are equally weighted together. For all the models, except the asymmetric model and the gamma model, the kernel widths have been selected by minimising Ignorance.

Skill relative to persistence, is shown in Table 3.4. For single models, the “one” column, the CERES-Maize model has the most skill and the asymmetric model the least. However, when two models are equally weighted, the multi-models with more skill than persistence are the CERES-Maize with the asymmetric model, a skill score of -0.90, and the CERES-Maize with

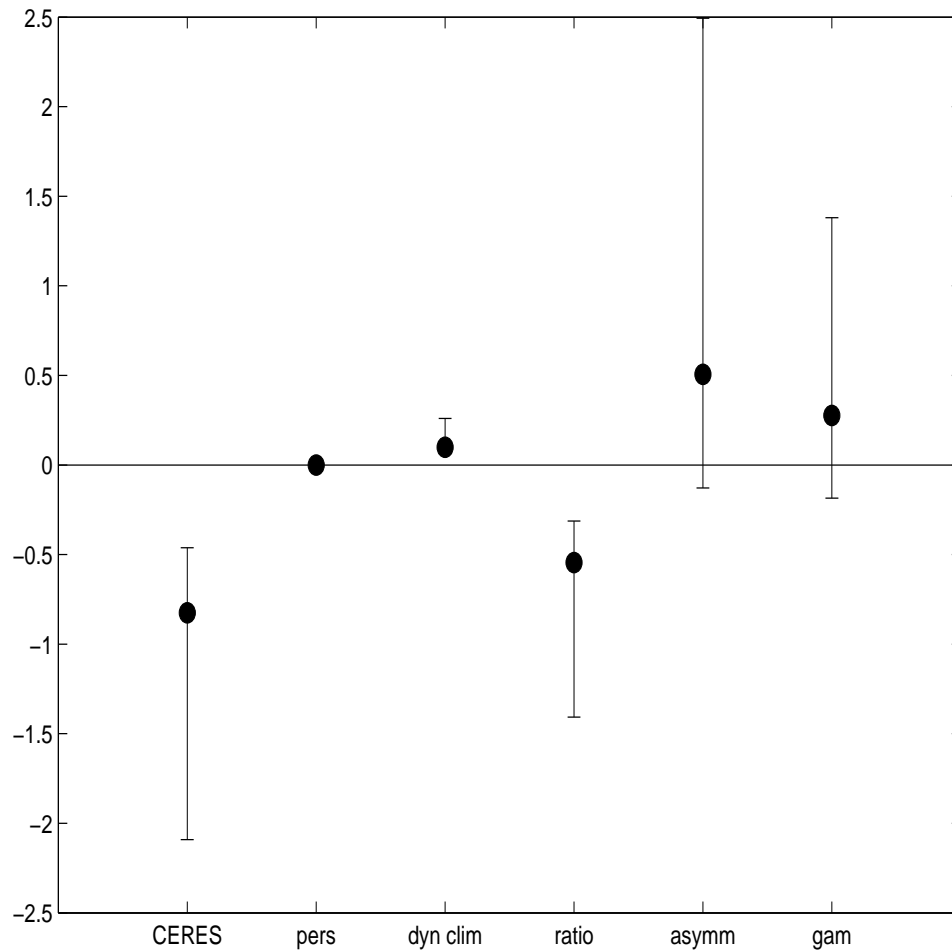


Figure 3.20: Comparison of model skill relative to persistence when the kernel widths are chosen by minimising ignorance. The dot is the mean relative Ignorance and the line is the 5% and 95% bootstrap resampling interval. The ratio model has significantly more skill than the other empirical models as the relative ignorance is lower.

the ratio model, a skill score of -0.84. This is a surprising result for the asymmetric model, as it is the crop model with the least skill. When it is combined with the CERES-Maize model however, the probabilistic forecast improves.

Table 3.4: Skill relative to persistence for single models (1 model) and two equally weighted models.

model	one	CERES	ratio	dyn clim	asym	pers	gam
CERES	-0.82	-0.82	-0.84	-0.61	-0.90	-0.65	-0.79
ratio	-0.54	-0.84	-0.54	-0.33	-0.52	-0.38	-0.64
dyn clim	0.10	-0.61	-0.33	0.10	-0.30	0.04	-0.28
asym	0.51	-0.90	-0.52	-0.30	0.51	-0.37	-0.44
gamma	0.28	-0.79	-0.64	-0.28	-0.44	-0.30	0.28
pers	0	-0.65	-0.38	0.04	-0.37	0	-0.30

Figures 3.21 and 3.22 compare the skill between forecasts from a single models and two equally weighted models. The single model is on the left in green and the two equally weighted models are on the right in blue. The x -label shows the name of the single model on the left and the name of the second model (that this is equally weighted with) on the right. A single + sign means there is no second model. The mean skill, measured relative to persistence, is the dot in the middle and the vertical lines represent the 5% and 95% bootstrap resampling interval. The horizontal line, at zero, is the skill of the bench mark model, persistence. It is important to note that bootstrap resampling is not the same as standard deviation. The bootstraps are asymmetric, not symmetrical and therefore Gaussian distribution cannot be assumed.

In this section the focus is on forecasts from equally weighted models and how they compare with forecasts from the CERES-Maize model. Our conclusions are as follows:

1. Figure 3.21a, compares the CERES-Maize model (in green) against the CERES-Maize model equally weighted with an empirical model, on average the CERES-Maize is not always the model with the most skill.
2. Equally weighting the CERES-Maize model with the asymmetric model slightly improves the skill, but does not reduce the spread of the bootstrap resampling interval.
3. Equally weighting the CERES-Maize model with the ratio model, slightly improves the skill and provides a smaller bootstrap resampling interval.
4. As both the CERES-Maize and asymmetric models have the widest bootstrap resampling intervals, equally weighting these two models does not reduce this interval.
5. Figure 3.23 shows the skill relative to the CERES-Maize model. In this Figure a negative score means the equally weighted model has more skill than the CERES-Maize model.
6. When the additional model is the asymmetric model the skill is negative, indicating more skill than the CERES-Maize model. Under bootstrap resampling only 59.9% of the results are less than zero, indicating this might not be robust.

7. When the additional model is the ratio model the skill is negative. Under bootstrap resampling only 58.6% of the results are less than zero, indicating this might not be robust.
8. For all the other equally weighted models both the skill and more than 50% of the bootstrap resampling are above zero suggesting the CERES-Maize model provides a probabilistic forecast for yield with more skill.

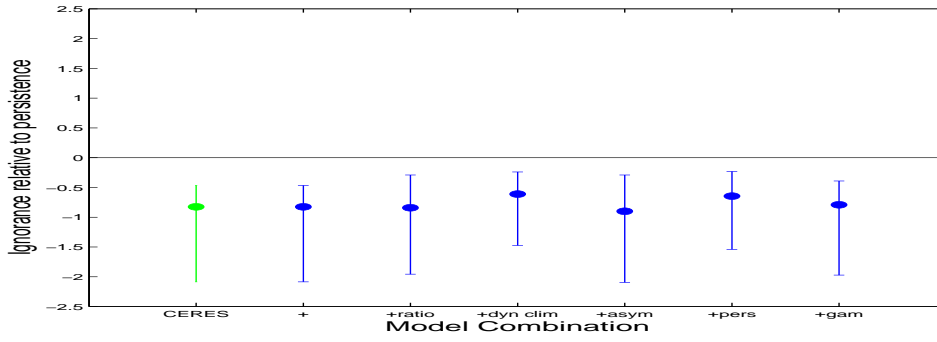
3.8 Multi-model forecasting at US state level

The eight states which produced the most maize (by volume) in 2013 are listed below in Table 3.5 [78]. From these eight states only two states, Nebraska and Kansas, are mainly irrigated [78].

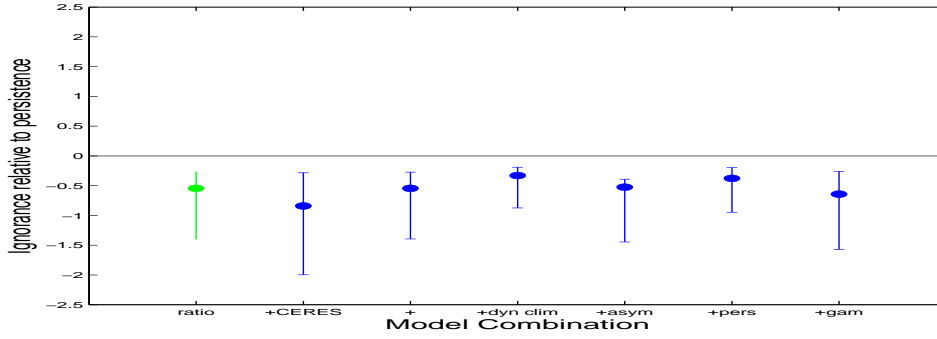
Table 3.5: Volume of maize produced by the top eight maize producing states in 2013

State	volume (bushels)	volume (%)
Iowa	2,161,500,000	18.4
Illinois	2,100,400,000	17.9
Nebraska	1,623,500,000	13.8
Minnesota	1,304,000,000	11.1
Indiana	1,035,450,000	8.8
South Dakota	808,680,000	6.9
Ohio	661,980,000	5.6
Kansas	508,000,000	4.3

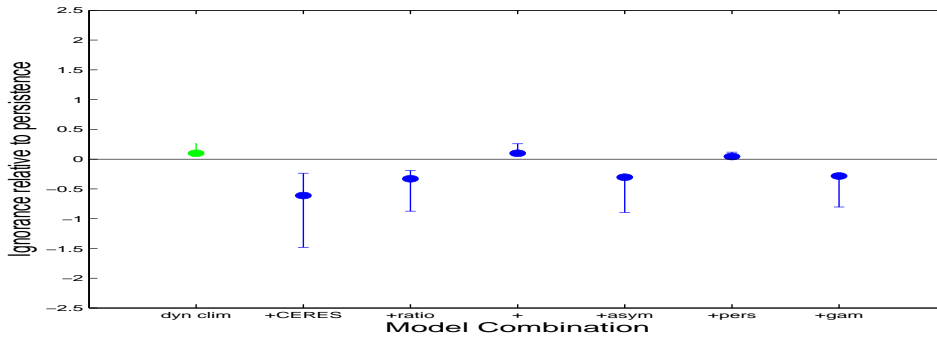
How much skill does the CERES-Maize model have at forecasting US state level maize yield? The kernel width was set individually for each US state



(a) CERES-Maize

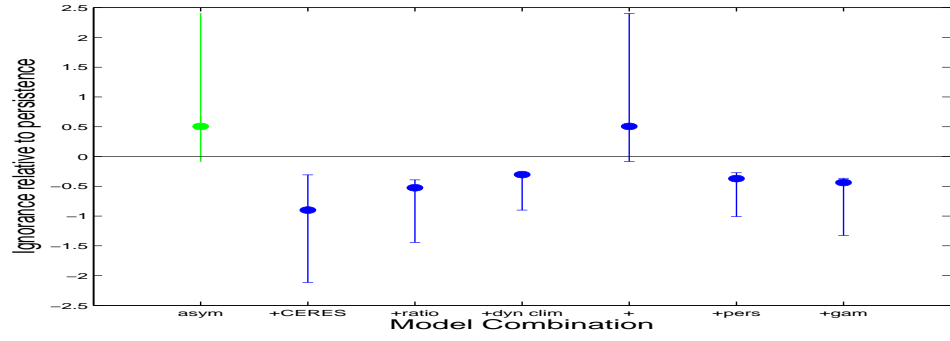


(b) Ratio

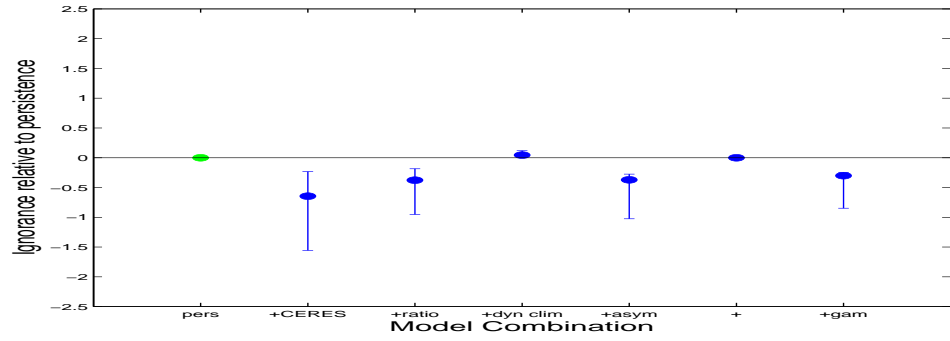


(c) Dynamic Climatology

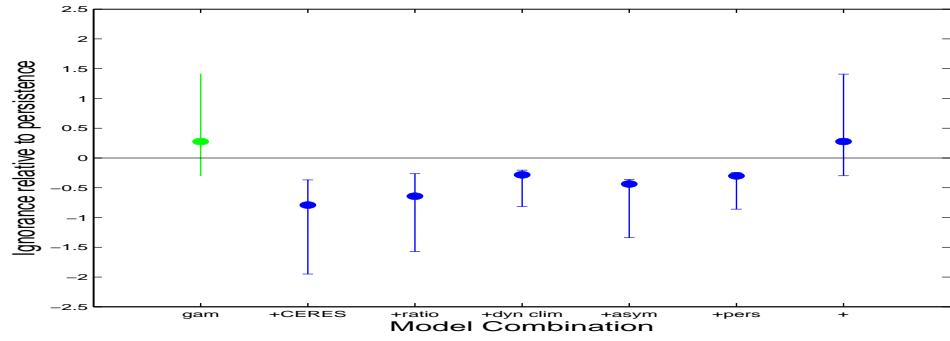
Figure 3.21: Comparison of skill between a single model and two equally weighted models for the CERES-Maize, Ratio and Dynamic Climatology model. The top and bottom of the line represent the 5% and 95% bootstrap resampling interval. Notice how the most skillful model is the equally weighted CERES-Maize and asymmetric models in (a).



(a) Asymmetric



(b) Persistence



(c) Gamma

Figure 3.22: Comparison of skill between a single model and two equally weighted models for the asymmetric, persistence and gamma model.

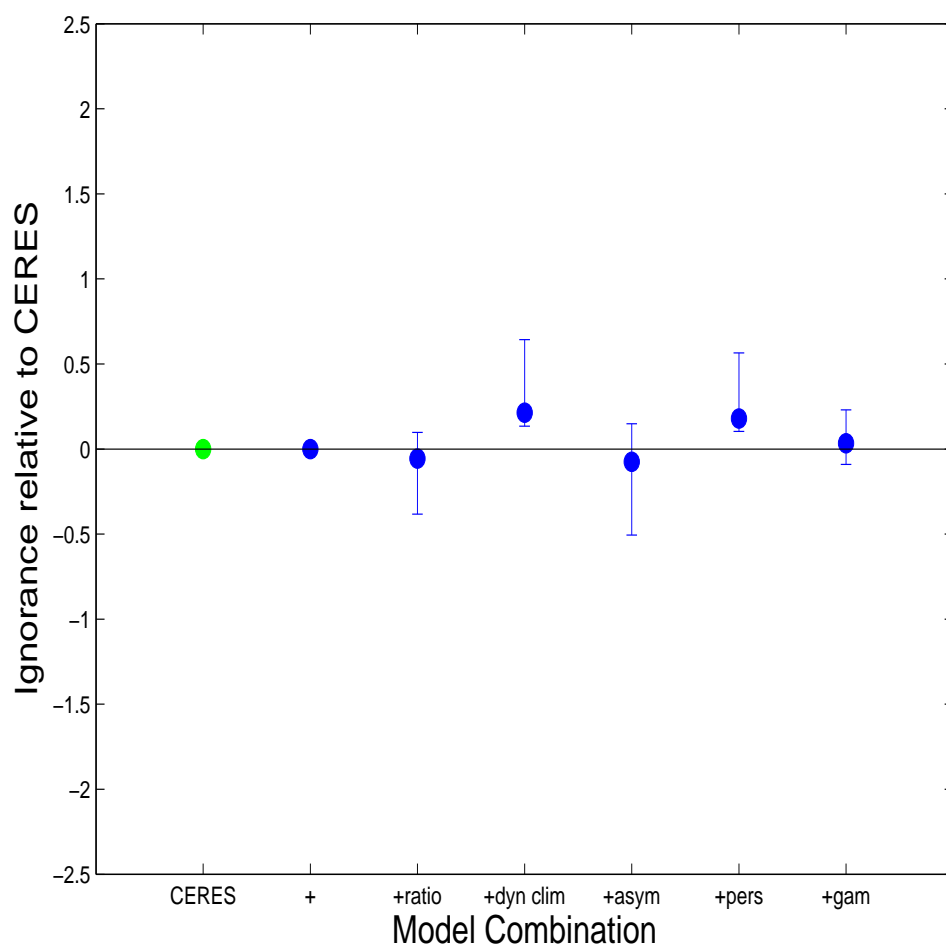


Figure 3.23: CERES-Maize: the skill is relative to the CERES-Maize model. When CERES-Maize is equally weighted with the ratio model and the asymmetric model it has more skill than the CERES-Maize model

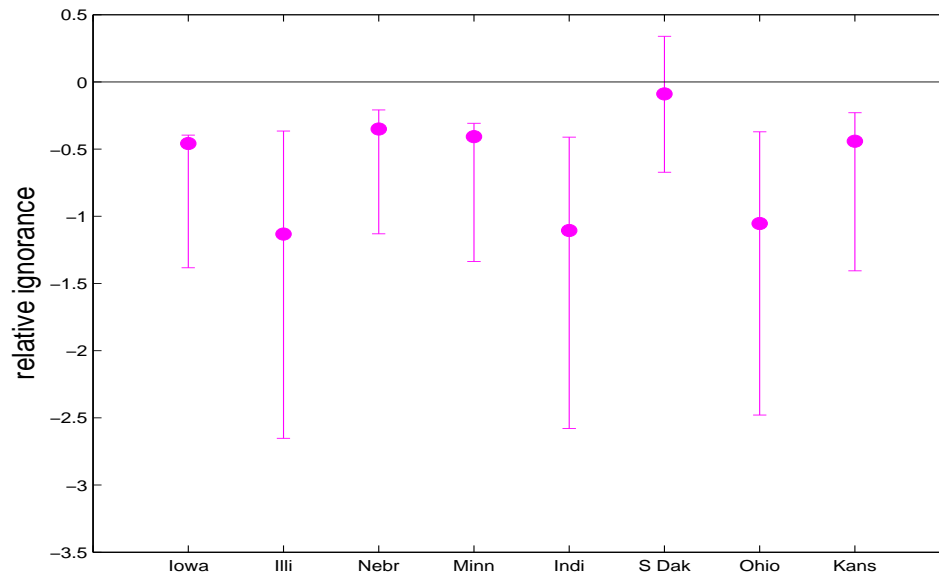
by minimising Ignorance [18]. The skill in Table 3.6 is measured relative to the persistence model. A negative score shows the probabilistic forecasts from the CERES-Maize model have more skill than persistence. Here again just the results of the models with the most (and least) skill are considered. The skill for probabilistic forecasts for the top eight maize producing states is shown in Figure 3.24a. For clarification:

1. The states with more skill are Illinois (-1.13), Indiana (-1.11) and Ohio (-1.05).
2. The state with least skill is South Dakota with a mean relative Ignorance of -0.09, hardly beating persistence. The CERES-Maize model has less skill at forecasting yield for South Dakota than any other state.
3. There is little difference in the skill of probabilistic forecasts from irrigated states (Nebraska and Kansas) and non-irrigated states.

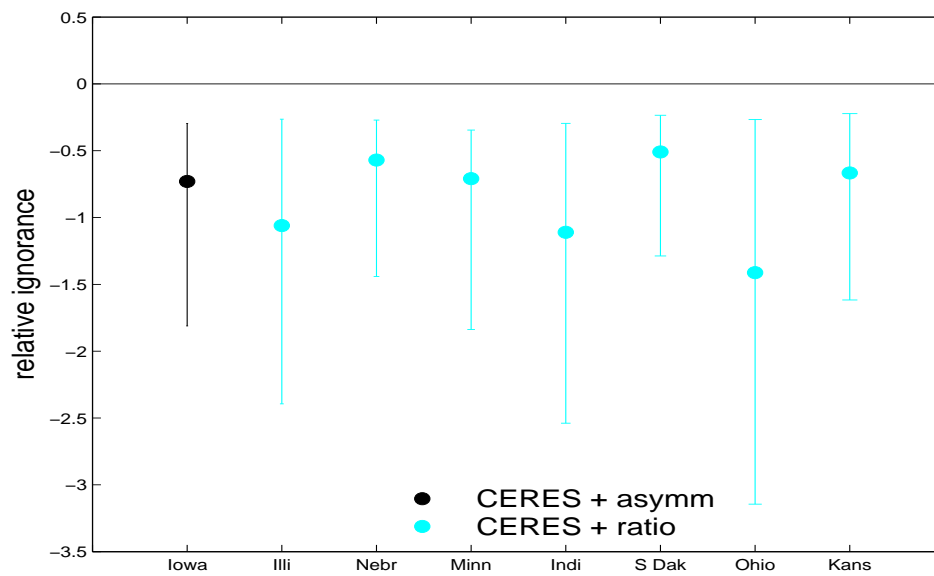
Table 3.6: Skill of the CERES-Maize model relative to persistence for 8 states.

State	rel Ign	5%	95%
Iowa	-0.46	-0.96	-0.05
Illinois	-1.13	-1.50	-0.74
Nebraska	-0.35	-0.80	0.15
Minnesota	-0.41	-0.96	0.08
Indiana	-1.11	-1.50	-0.74
South Dakota	-0.09	-0.58	0.43
Ohio	-1.05	-1.44	-0.68
Kansas	-0.44	-0.95	0.32

If two crop models are equally weighted together, which multi-model provides



(a) CERES-Maize model



(b) Multi-models

Figure 3.24: Fig 3.24a: Comparison of skill relative to persistence for the CERES-Maize model for the top eight maize producing states. The more skillful CERES-Maize models are the states of Illinois, Indiana and Ohio. Fig 3.24b: Comparison of the most skillful equally weighted crop models by state. For 7 out of 8 states the most skillful multi-model is CERES-Maize with the ratio model.

forecasts of yield with the most skill, relative to persistence, by US state?

The results are in Figure 3.24b. Our conclusions are as follows:

1. The multi-model with the most skill is consistently the CERES-Maize with the ratio model for each state apart from Iowa.
2. The bootstrap resampling intervals have not significantly decreased by using a multi-model.
3. For Iowa the multi-model with the most skill is the CERES-Maize with the asymmetric model, here again the bootstrap resampling intervals remain approximately the same size.

Although these results show us that there is an improvement in the average skill if the CERES-Maize model is equally weighted with an empirical model, it does not show whether or not this improvement is significant. To consider this the skill of the CERES-Maize model is compared directly against the models which have the most skill by state. There are three models which have the most skill by state, the ratio model, the CERES-Maize and ratio model and finally the CERES-Maize and asymmetric model. The skill relative to the CERES-Maize model is shown in Figure 3.25. Different US states have different models with the most skill. In summary:

1. For Iowa the multi-model CERES-Maize with the asymmetric model has the most skill, though all three models have negative relative Ignorance so more skill than the CERES-Maize model.
2. For Illinois and Indiana the CERES-Maize model has the most skill.

3. The multi-model CERES-Maize and ratio has the most skill in Nebraska, Minnesota, Ohio and Kansas.
4. For South Dakota the ratio model has the most skill, for this state the CERES-Maize model had low skill. All three models have more skill than the CERES-Maize model.
5. The asymmetric model only has more skill than the CERES-Maize model for Iowa and South Dakota.
6. The CERES-Maize and ratio model tend to have the most skill, or almost the most skill across all the states.

The results are shown in Table 3.7 where South Dakota, Iowa and Nebraska have noticeable improvements in skill from the CERES-Maize model. Skill for Illinois obviously remains the same but the skill for Indiana and Ohio have little difference between the CERES-Maize model and the multi-model.

Table 3.7: Model with the most skill relative to persistence by state.

State	model	rel Ign	5%	95%	> 0(%)
Iowa	CERES+asym	-0.73	-1.08	-0.43	97.1
Illinois	CERES	-1.13	-1.53	-0.76	0
Nebraska	CERES+ratio	-0.57	-0.87	-0.30	92.4
Minnesota	CERES+ratio	-0.71	-1.12	-0.36	94.9
Indiana	CERES+ratio	-1.11	-1.43	-0.82	30.2
South Dakota	ratio	-0.56	-0.85	-0.26	96.8
Ohio	CERES+ratio	-1.41	-1.73	-1.15	99.1
Kansas	CERES+ratio	-0.66	-0.95	-0.44	64.3

Equally weighting the CERES-Maize model with an empirical model can increase the skill relative to persistence, though which model to use depends

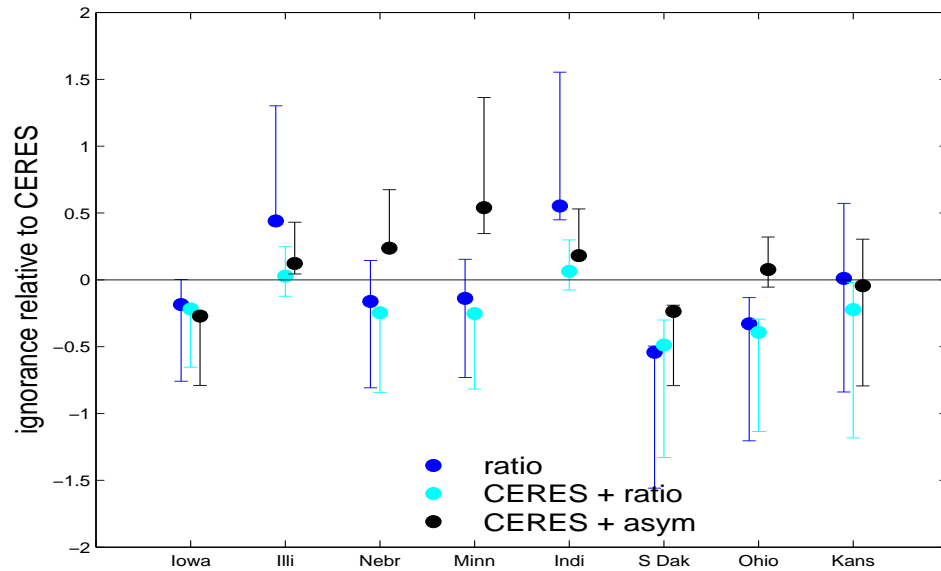


Figure 3.25: Comparison of the top three models relative to the CERES-Maize model. For Illinois and Indiana the CERES-Maize model provides the best forecast.

on the state. If a CERES-Maize forecast for a particular state has low skill, an empirical model may provide more skill.

The fact that data is scarce, that it will take years to get considerable more data and that available observations, where known, were training the CERES-Maize model makes it difficult for any near term future research to establish significance with confidence.

3.9 Conclusions

1. A new methodology and illustration for the construction of a probabilistic forecast from a crop model is in Section 3.1.
2. Establishing a baseline using as a bench mark model an empirical model

from maize yield observations, persistence, is explained in Section 3.3

3. The development of a new empirical crop model which creates an ensemble around a rolling maximum using ratios is in Section 3.4.2.
4. The development of a new empirical crop models using asymmetric kernels is in Section 3.4.3 and Section 3.4.4. To our knowledge this the first time an asymmetric kernel has be used in crop modelling.
5. Why minimising Ignorance to select kernel widths provides probabilistic forecasts with the most skill is shown in Figure 3.19
6. Comparison of the skill of single crop models relative to persistence for US maize when the kernel widths have been set by minimising Ignorance is in Figure 3.20. The CERES-Maize model has the most skill and the ratio model is the empirical model with significantly more skill than the others.
7. Which two crop models equal weighted together have the most skill is examined in Section 3.7. Figure 3.23 shows equally weighted the CERES-Maize and asymmetric models has the most skill.
8. The skill of the CERES-Maize model at state level is in Figure 3.24a. Forecasts for South Dakota have significantly less skill than other states.
9. The most skillful model or multi-model by state is in Section 3.8. In Figure 3.25 it is demonstrated that including an empirical model can improve the skill of the CERES-Maize probabilistic forecasts.

Chapter 4

Analysis of weather and yield

Accurate estimates of weather across crop growing areas are a useful tool for predicting the variability in crops. Weather has a large non-linear impact on yield, particularly when there are adverse weather conditions such as drought [44, 64, 75, 77, 97]. This chapter considers the impact of daily meteorological observations on maize yield. Daily observations from the USHCN [52] are examined in Section 4.1, along with the quality of these observations for Iowa, the US State which produces the largest amount of maize. The timing of any adverse weather events in the crop cycle of maize is important. For example in Iowa maize is usually planted during April and May and delays to this can impact yield [58]. Another key date is the harvesting of maize during October and November [58]. The main causes of observational uncertainty in the meteorological observations time series from the USHCN are reviewed in Section 4.2.

Drought and extreme wet weather causes maize yield levels to fall [57]. Additionally, high temperature alone can cause crop failure [30,45], particularly if it coincides with the flowering stage of the crop cycle. In Section 4.3 the impacts from the highest daily t_{max} and precipitation on maize yield are considered. The number of days greater than 29°C and yield is also examined.

Maize yield at county level for the state of Iowa is explored in Section 4.4. The CERES-Maize modellers calibrate for technical advancements by fitting a simple linear regression through both the outcomes and the forecasts [19,37], as discussed in Section 1.5. In Section 4.4 the variability in slope parameters from fitting a linear regression to the maize yield time series is examined. Adverse weather events on maize were examined by considering the year with the largest percentage fall in maize yield for each county in Iowa. The impact on the surrounding counties between years was compared. This information, presented graphically, is new as far as we are aware.

Disentangling variation in maize yield time series caused by weather from variation caused by technical advancements is difficult. Improvements from technical advancements cause maize yield to increase in a non-linear manner, as do favourable weather conditions. In Section 4.5 a new method is put forward to better identify the technical advancements in the yield curve.

4.1 Meteorological observations from Iowa

The meteorological observations considered in this section are from weather stations operated by the USHCN [52]. The stations are generally located in rural areas or small towns and are spread throughout the US. The weather stations record five daily variables: t_{min} , t_{max} , precipitation, snowfall and snowfall depth. Alongside the daily readings are flags which highlight any suspected problems with the quality of the daily readings. The analysis in this chapter uses forty one years of meteorological observations from Iowa. The years are 1970 to 2010, the most recent available when the data was downloaded from <http://cdiac.esd.ornl.gov/epubs/ndp/ushcn/ushcn.html> on 1st December 2013.

4.1.1 Location of weather stations in Iowa

In Iowa there are 23 USHCN weather stations spread evenly across the state, each located in a different county, as shown in Figure 4.1. Details of the location and symbol representing each weather station are shown in Table 4.1. Each weather station is identified by a unique code where the first two digits identify the state (see Appendix D). Although not every county has a weather station, in this chapter the meteorological observations from the weather stations are used as a proxy for the weather in Iowa.

Meter Code	County Code	County Name	Symbol
130112	135	Monroe	•
130133	109	Kossuth	•
130600	11	Benton	•
131402	67	Floyd	•
131533	145	Page	•
131635	45	Clinton	•
132724	63	Emmet	•
132789	101	Jefferson	○
132864	65	Fayette	○
132977	189	Hancock and Winnebago	○
132999	187	Webster	○
134063	181	Warren	○
134142	83	Hardin	○
134735	149	Plymouth	○
134894	85	Harrison	*
135769	159	Ringgold	*
135976	87	Henry	*
135952	37	Chickasaw	*
137147	119	Lyon	*
137161	25	Calhoun	*
137979	21	Buena Vista	*
138296	171	Tama	△
138688	183	Washington	△

Table 4.1: The USHCN weather stations in Iowa, their county and the symbol used in this thesis to represent them.

4.1. Meteorological observations from Iowa

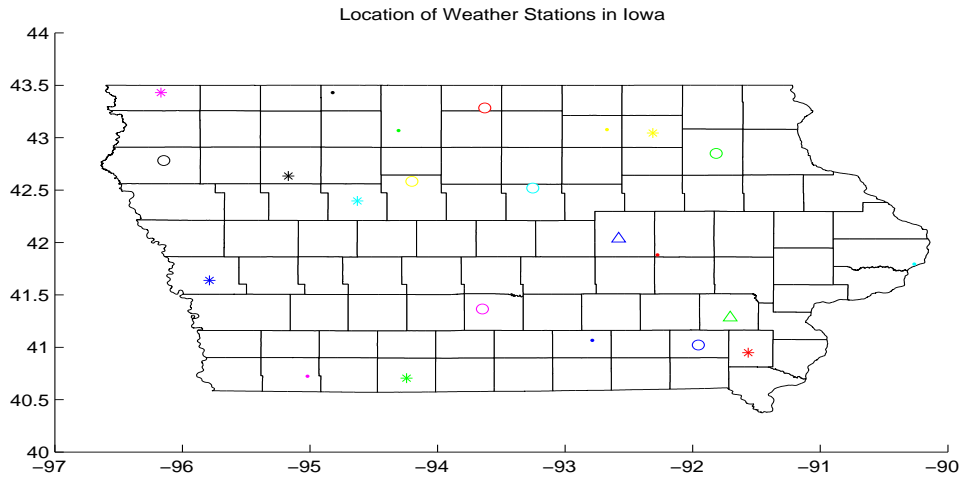


Figure 4.1: Location of USHCN weather stations in Iowa. Note that although the weather stations are evenly spread throughout Iowa, there is not a weather station in every county.

symbol	meaning of symbol
t_{max}	daily maximum temperature
$\bar{t}_{max,m}$	monthly mean of t_{max}
$\bar{t}_{max,y}$	yearly mean of t_{max}
$\bar{t}_{max,ay}$	mean across all years t_{max}
$t1_{max}$	highest daily t_{max} in the year
$\bar{t}1_{max}$	mean highest $t1_{max}$ across all the years
t_{min}	daily minimum temperature
$\bar{t}_{min,m}$	monthly mean of t_{min}

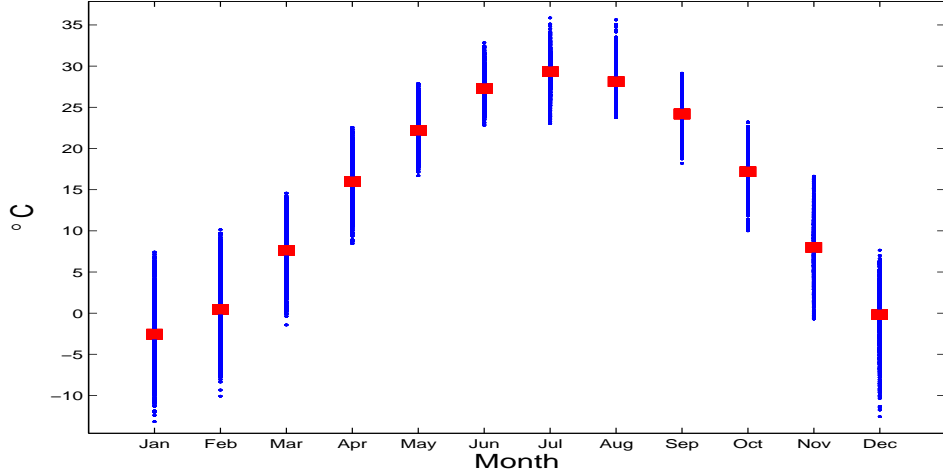


Figure 4.2: Monthly $\bar{t}_{max,m}$ in Iowa. On average higher $\bar{t}_{max,m}$ are during the months June, July and August.

4.1.2 Maximum temperature

A summary of the mean monthly maximum temperatures ($\bar{t}_{max,m}$) is shown in Figure 4.2. Each blue point represents the monthly $\bar{t}_{max,m}$ for a year at a weather station. The red squares show the mean of $\bar{t}_{max,m}$ for the relevant month. The seasonality of $\bar{t}_{max,m}$ is clearly seen in the figure.

To define whether the year's mean maximum temperature ($\bar{t}_{max,y}$) is warmer or cooler than average, the anomaly is calculated. The anomaly measures the individual year's deviation from the overall average maximum temperature ($\bar{t}_{max,ay}$), as shown in Figure 4.3. For each year a blue dot represents the mean anomaly across all the weather stations in Iowa. An average year would have an anomaly that is approximately zero, a warmer than average

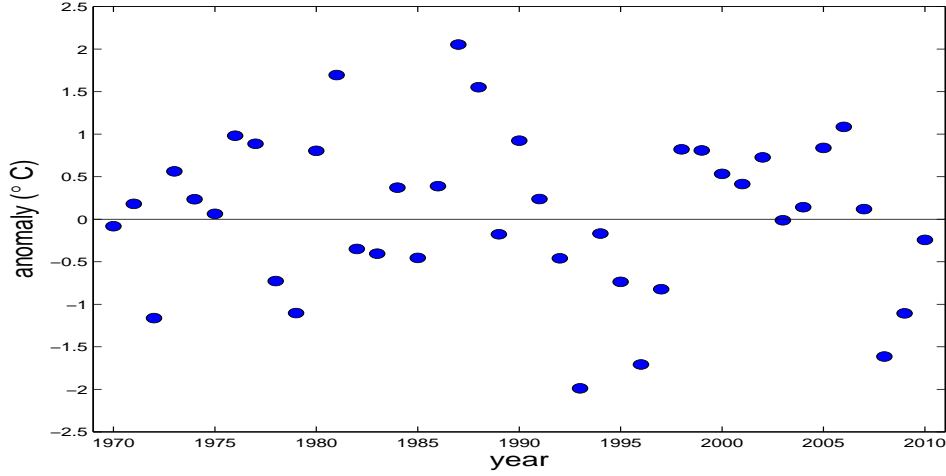


Figure 4.3: Annual anomalies measured from the $\bar{t}_{max,ay}$ in Iowa. Note that 1981, 1987 and 1988 all have large positive anomalies, so are warmer than an average year.

year would have a positive anomaly and a cooler than average year a negative anomaly. The warmest year, shown by the largest positive anomaly, is 1987.

Maximum temperature (t_{max}), is an important variable for predicting crop yield, as just a few days, or even hours, with a very high t_{max} can negatively impact yield [30]. The highest t_{max} in a year, $t1_{max}$, is shown by weather station in Figure 4.4. The symbols represent the individual weather stations and the black line is the average across all weather stations. In Iowa 1988 had the highest mean $t1_{max}$ ($\bar{t1}_{max}$) although some individual weather stations recorded higher maximums in 1974 and 1975. In 1998 maize yield in Iowa was significantly lower than average because of a severe drought [90]. The lowest $\bar{t1}_{max}$ year was 1993, another year when maize yield was low. Weather stations located in the North East of Iowa such as * (for weather station 135952 in Chickasaw) tend to have higher t_{max} than the other counties.

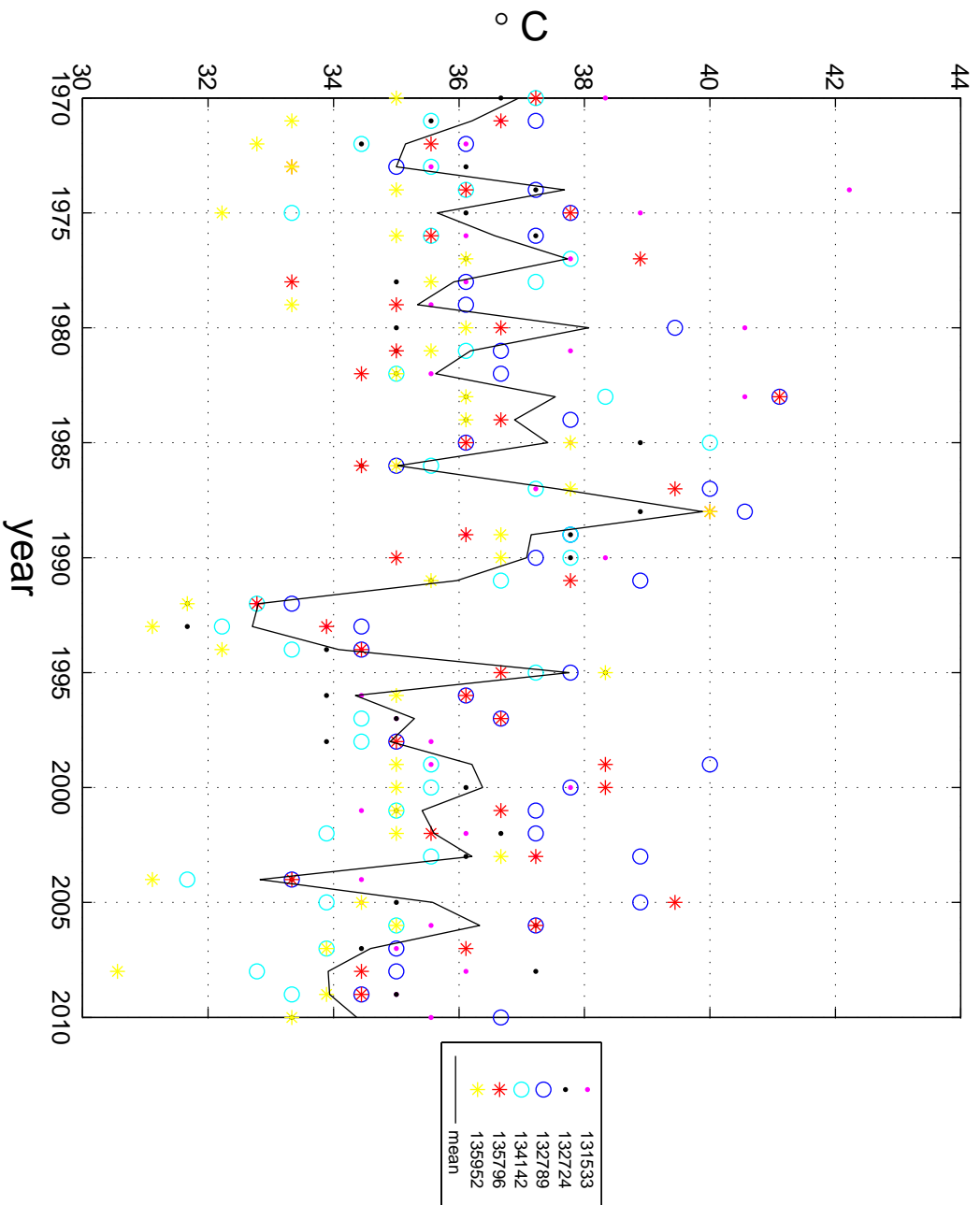


Figure 4.4: Highest t_{max} in each year ($t1_{max}$) by weather station in Iowa. The black line is the mean ($\bar{t1}_{max}$) across the weather stations. The highest $\bar{t1}_{max}$ in Iowa was in 1988.

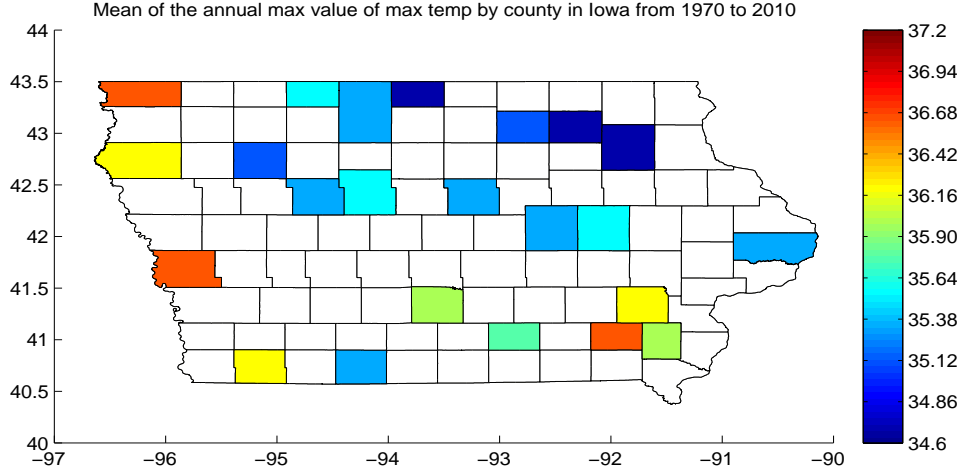


Figure 4.5: The mean of the highest t_{max} in the years 1970 to 2010 ($\bar{t}_{max,ay}$) by county in Iowa. Counties located in the West and South have higher temperatures shown by the orange and yellow colours.

For each county, the mean highest t_{max} (\bar{t}_{max}) across 40 years is shown in Figure 4.5. The hotter counties are shown as red and the cooler counties are shown as blue. White indicates there are no weather stations in that county. Counties located in North East Iowa tend to report cooler temperatures than counties located in West or South Iowa. The differences between counties can be more than 2°C .

4.1.3 Quality Control for Meteorological Observations

USHCN check the quality of the meteorological observations and flag any suspected erroneous readings [52]. The flags are coded to explain why a reading failed the quality check. Readings that pass the quality check have no flags. The flag codes are shown in Tables 4.3, D.2 and D.3. Any missing

readings are usually marked with -9999. In addition, if there are less than 31 days in a month, the missing days are marked with -9999. In practice, however, missing readings are not always marked this way, instead there are several large gaps in individual weather station's time series.

The codes from flags used for t_{max} readings in Iowa are summarised in Table 4.4. In all 99.06 % **qflag** were not set, as the readings had passed the USHCN quality checks. The code with the next highest count was the **qflag** I, where the t_{min} was greater than the t_{max} in 0.92% of the readings.

The analysis in this chapter only includes meteorological observations where the **qflag** is not set. Other exclusions are for monthly (or annual) estimates when there are more than 9 problems with daily readings, these could either be from missing readings or readings with a **qflag**. For example if a weather station was missing daily records for June, July or August the highest annual daily t_{max} was not calculated for that weather station for that year.

4.1.4 Precipitation

Monthly precipitation for Iowa is shown in Figure 4.6. There is seasonality in precipitation with on average higher precipitation in June than the rest of the year. Even after the data is screened for error flags and missing data there are a few precipitation values that seem exceptionally high in January and

Table 4.3: Summary of Q Flags for Temperature Data

Flag Type	Code	Explanation
q	not set	did not fail any quality assurance check
q	A	checks for t_{max} that are below the t_{min} across a three-day window
q	D	checks for duplication of the data across years or months
q	G	a maximum t_{min} that is at least 10°C warmer (or cooler) than all other t_{max}/t_{min} for a given station and calendar month
q	I	the t_{min} is greater than t_{max}
q	K	checks for streaks of 15 or more identical values
q	M	looks for daily t_{max} that are less than the lowest t_{min} (and vice versa) by month for each station
q	N	both t_{max} and t_{min} are equal to 0°C
q	O	checks that the daily t_{max}/t_{min} does not exceed the 15 day climatological mean by more than size standard deviations
q	S	the temperature differs by more than 10°C from the neighbouring stations on the preceding, current and following days
q	T	checks that the daily t_{max} (or t_{min}) does not exceed the t_{max} (or t_{min}) on the preceding and following days by more than 25 °C
q	X	t_{max} (or t_{min}) that fall outside the world extremes for the highest (or lowest) temperature ever seen

Table 4.4: Flags from t_{max} observations in Iowa

flag	code	number of points	percentage
m	blank	258,721	98.7 %
m	L	3353	1.3 %
q	blank	259,610	99.06 %
q	G	5	0.02 %
q	I	2401	0.92 %
q	N	1	0.00 %
q	S	57	0.02 %
s	blank	7244	2.76 %
s	0	254657	97.17 %
s	H	173	0.07 %

December. [†] The annual precipitation by year and weather station is shown in Figure 4.7. A weather station that regularly records higher precipitation than the other weather stations is 135976 (coded by ^{*}) in Henry county and a weather station that often records lower precipitation is 132724 (coded by .) in Emmet county. Figure 4.8 shows precipitation during the growing season by month and year. The drought years 1976, 1988 and 2012 have significantly less precipitation than the other years. In July 1993 (marked as light blue) a significantly larger volume of rain fell than in the other years, this caused wide spread wet rot to maize.

How precipitation varies across counties in Iowa is shown in Figure 4.9. Counties with high precipitation are coloured dark orange and counties with low precipitation are coloured dark blue. Higher precipitation is seen in South East Iowa and lower precipitation in North West Iowa.

[†]Determining the origin of these high points is beyond the scope of the Thesis.

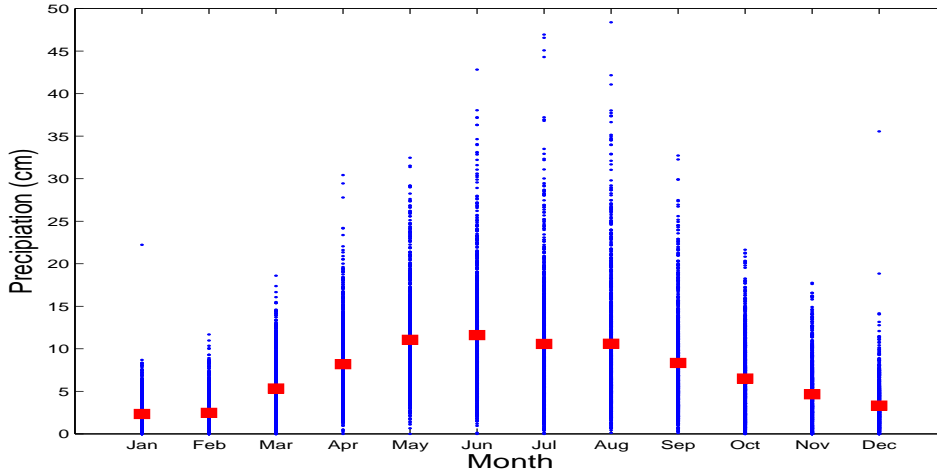


Figure 4.6: Monthly precipitation in Iowa. Note that on average more of the precipitation falls between May and August.

4.2 Causes of observational uncertainties

Some inaccuracies in the meteorological observations are caused by changes in how the observations are measured, rather than changes in the weather. A systematic bias in temperature readings is caused by a shift in the time of day the observations were recorded [92]. From the 1940s individual weather stations switched their time of observations from the afternoon to the morning without recording the date of this change, this caused a cooling bias in the t_{min} and t_{max} time series as discussed in Section 1.3.

Other changes introduced additional inconsistencies into the archive. The weather stations were gradually updated from a manually read liquid-in-glass thermometer contained in a white wooden shelter to an electronically

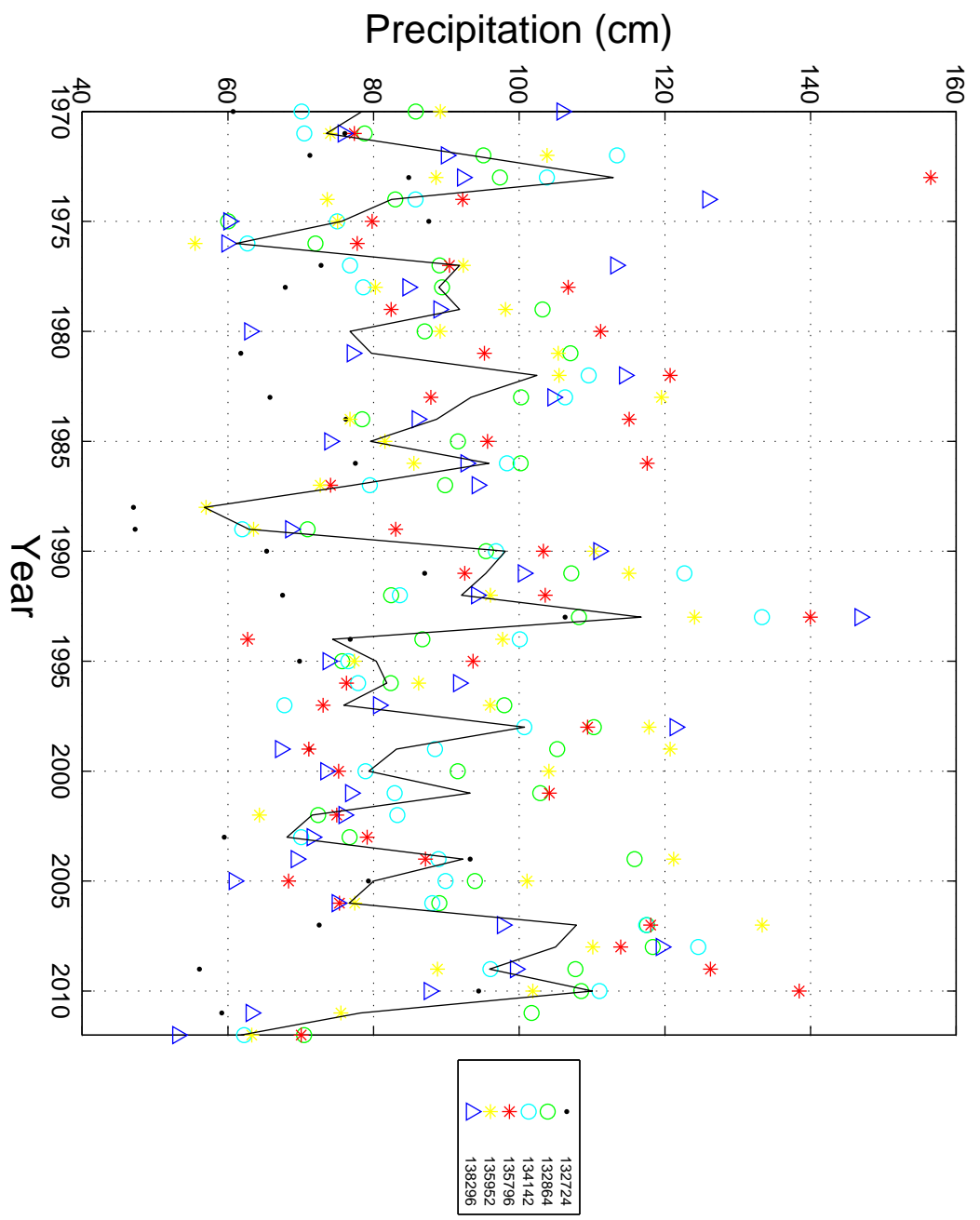


Figure 4.7: Annual precipitation by year and selected weather stations in Iowa. The black line is the mean precipitation across all the weather stations. The years 1973 and 1993 had high precipitation.

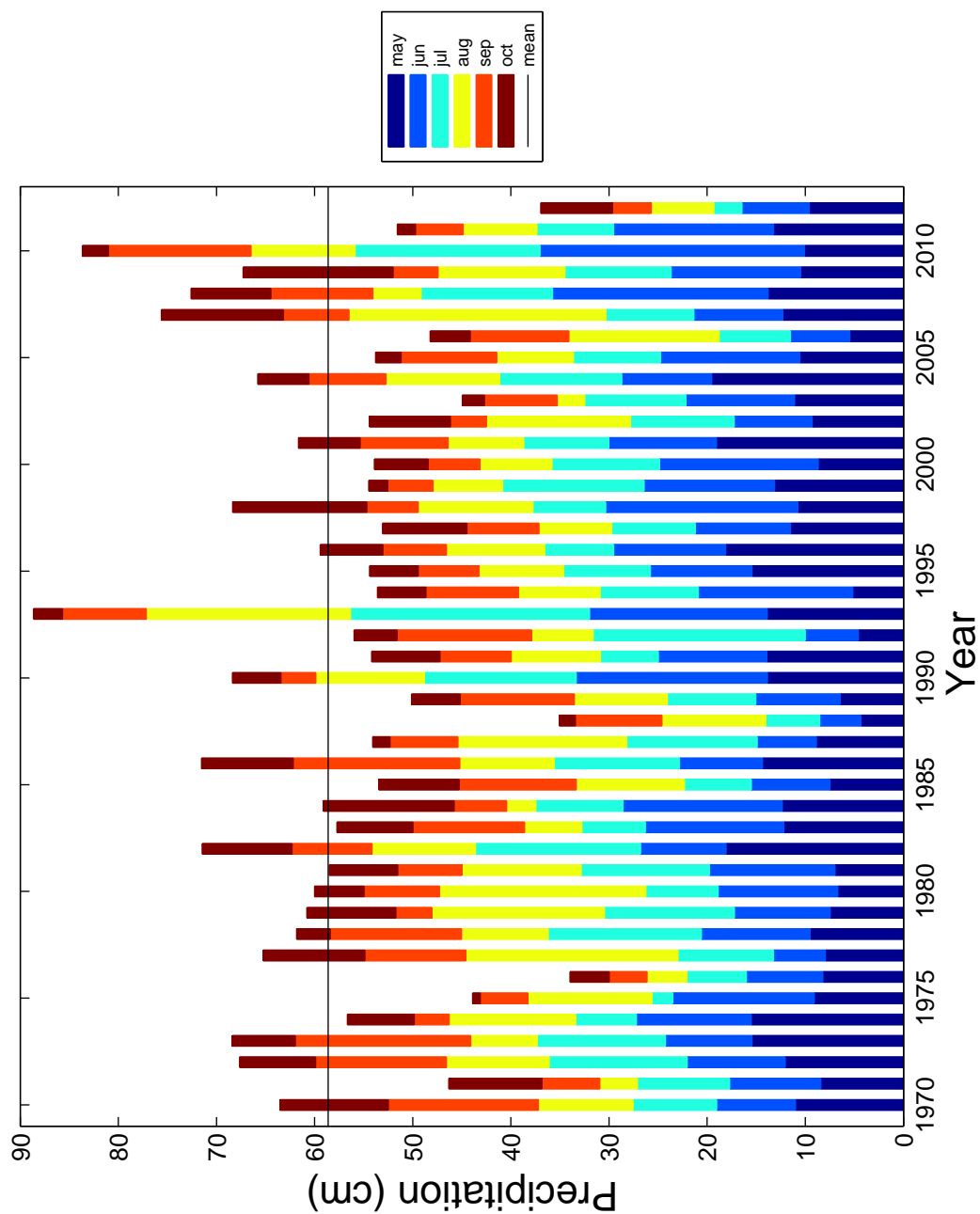


Figure 4.8: Growing season precipitation by year and month in Iowa. The precipitation in July 1993 (light blue) was much more than July in the other years.

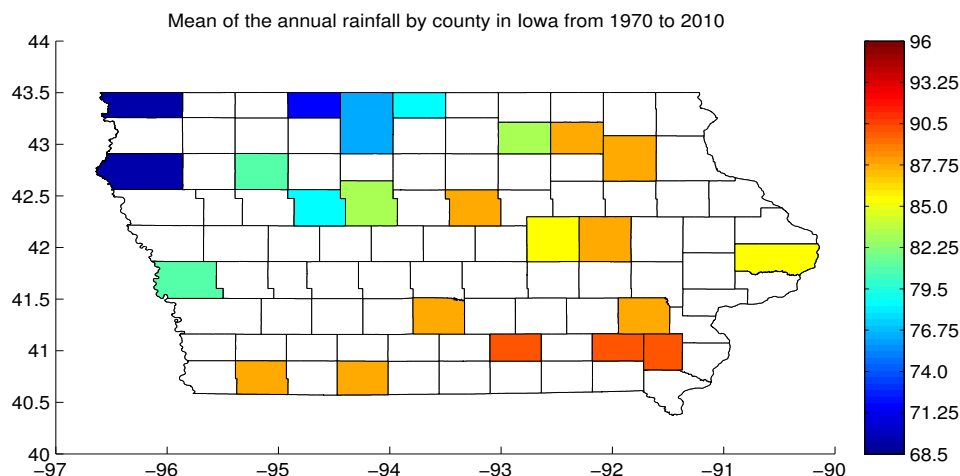


Figure 4.9: Map of the mean annual precipitation by county in Iowa. The counties in the North West of Iowa have lower annual precipitation than the other counties.

read maximum-minimum temperature system (MMTS) [92]. The MMTS introduced a negative bias to t_{max} and a positive bias to t_{min} . The exact date a particular station switched to MMTS was not always recorded; the update change was made piecemeal across the network of the weather stations from the 1980s onwards. The physical location of the weather station may also have been changed; for example it could have been moved from the roof of a building to the ground, or moved to a different location close by, again there is often no record of this move.

Photographs of most of the weather stations in the USHCN network have been taken by surfacestation.org [21]. Figures 4.10 and 4.11 show that some of the weather stations are positioned in less than ideal locations where the surroundings could impact on the meteorological observations. For example, some weather stations are on top of concrete, which acts as a heat source,



Figure 4.10: Weather Station 135796 Mount Pleasant Observation Station, Henry, Iowa * [21]. Meteorological observations will be hindered by the overgrown vegetation which blocks precipitation and sunlight.

or in shade during some of the day. When Menne et al [51] examined the reliability of the data by comparing the mean monthly anomaly from 1971 to 2001 between poorly sited and well sited stations they found no evidence that temperature trends were inflated due to poor station siting .

4.3 Interaction of meteorological observations and crop yield

The impact of weather on maize yield is considered in this section. If, during the growing season, t_{max} rises above a certain threshold (29°C) the maize yield falls sharply [45]. As modelled in the crop model (see Section 1.5) if this occurs during flowering the negative impact on yield is greater than at other times during the crop cycle. Both high t_{max} and low precipitation have



Figure 4.11: Weather Station 137147 Rock Rapids Observation Station, Lyon, Iowa * [21]. Meteorological observations will be affected by the surroundings.

negative impacts on maize yield [30].

In this thesis the interaction between maize yield, the highest annual daily t_{max} and precipitation using USHCN meteorological observations from Iowa are examined. To consider the impact from just weather, the effects from technical advancements in the maize yield time series need to be separated out. For this section, a linear regression is fitted using linear least squares to set the fixed parameters (α_1 and α_2) [35]. The fitted linear regression is shown as the blue line in Figure 1.2. The residuals are then compared against the weather variables, the residuals are calculated as

$$r_i = y_i - \alpha_1 + \alpha_2 i, \text{ where } i = 1, \dots, 34 \quad (4.1)$$

where the i^{th} outcome is y_i , r_i is the i^{th} residual for year $i = 1, \dots, n$.

The residuals were sorted by size and divided into three equally likely bins labelled high, medium and low. High residuals are marked by red stars, medium residuals by green circles and low residuals by blue squares. A comparison of the highest annual daily t_{max} , the precipitation (over the growing season) and the size of the residual is in Figure 4.12. The high residuals (red stars) are mainly clustered towards the middle of the figure, so in general maize yield is higher when there is average weather. The low residuals (blue squares) are separated out in two areas of the scatter plot; the top left when there was higher than average precipitation and lower than average maximum t_{max} and the bottom right when there was higher than average maximum t_{max} and lower than average precipitation, drought conditions.

What is the impact of high temperatures on yield? This is considered by comparing the number of days in a year where t_{max} was greater than a threshold temperature (29°C) against the yield residuals. The results are shown in Figure 4.13. There are low residuals without a high number of t_{max} days, but never high residuals when there is a high number of t_{max} days. A linear regression for the data is shown as the red dashed line. In 1993 extreme wet weather in July severely damaged crops. For no other year did any such phenomenon occur. As only the impact of high temperatures are considered, and given this is an extreme and identifiable cause, this point has been omitted from the regression. The linear regression shows that as

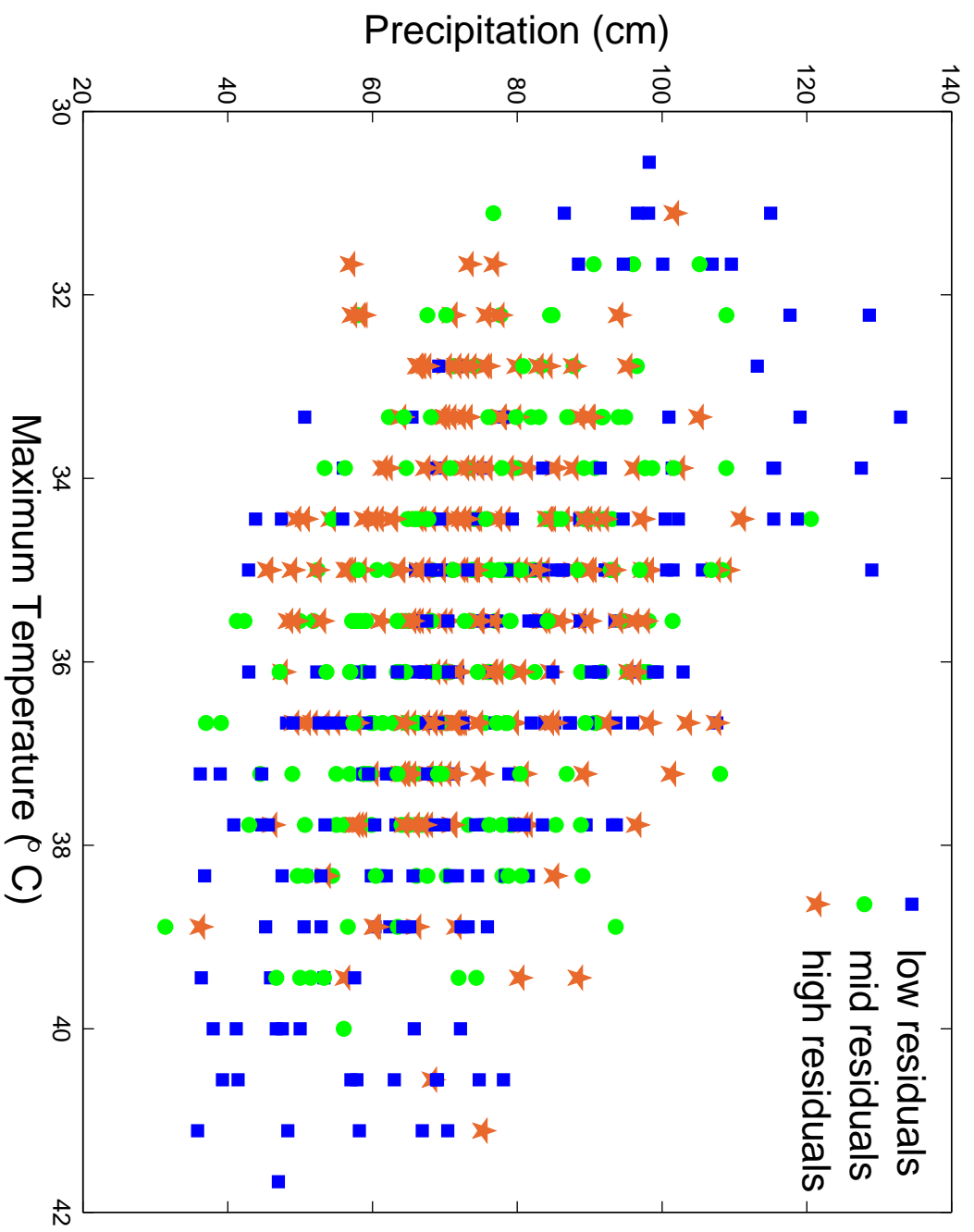


Figure 4.12: Comparison of the size of the residuals, the highest annual daily t_{max} and precipitation for Iowa. The highest values for residuals (red stars) are mainly found in years where the maximum temperatures and precipitation are average.

the number of days above 29°C increase, the residuals decrease as the yield has been impacted by the high temperatures. The difference, measured as yield residual, between the linear regression estimate and actual 1993 value is 66.5.

In the next section maize yield by county in Iowa is examined.

4.4 Maize yield by county in Iowa

Iowa has 99 counties, this section examines the maize yield from these for the time frame 1970 to 2011 [52]. A selection of six counties' maize yields by year are shown in Figure 4.14, these show large discrepancies in yield between counties. For example the yield for county 7 is lower and more volatile than yield from the other five counties. Additionally, the increase in yield over time is less in county 7. The mean yield by county is shown as a histogram in Figure 4.15a, most counties have means towards the upper end of the range which is from 100 to 140 bushels/acre. The location of the mean yield by county is shown in Figure 4.16 where the higher mean yields are orange, and the lower mean yields are blue. The southern counties tend to have lower yields and the central northern counties tend to have higher yields. The standard deviation of the yields by county is shown in Figure 4.15b, most of the counties have an average standard deviation of about 32. A few counties have a significantly higher standard deviation (of between 38 and 40) coloured orange in in Figure 4.17 and located in the North West of Iowa. Over the years as the technology has improved, the yield has

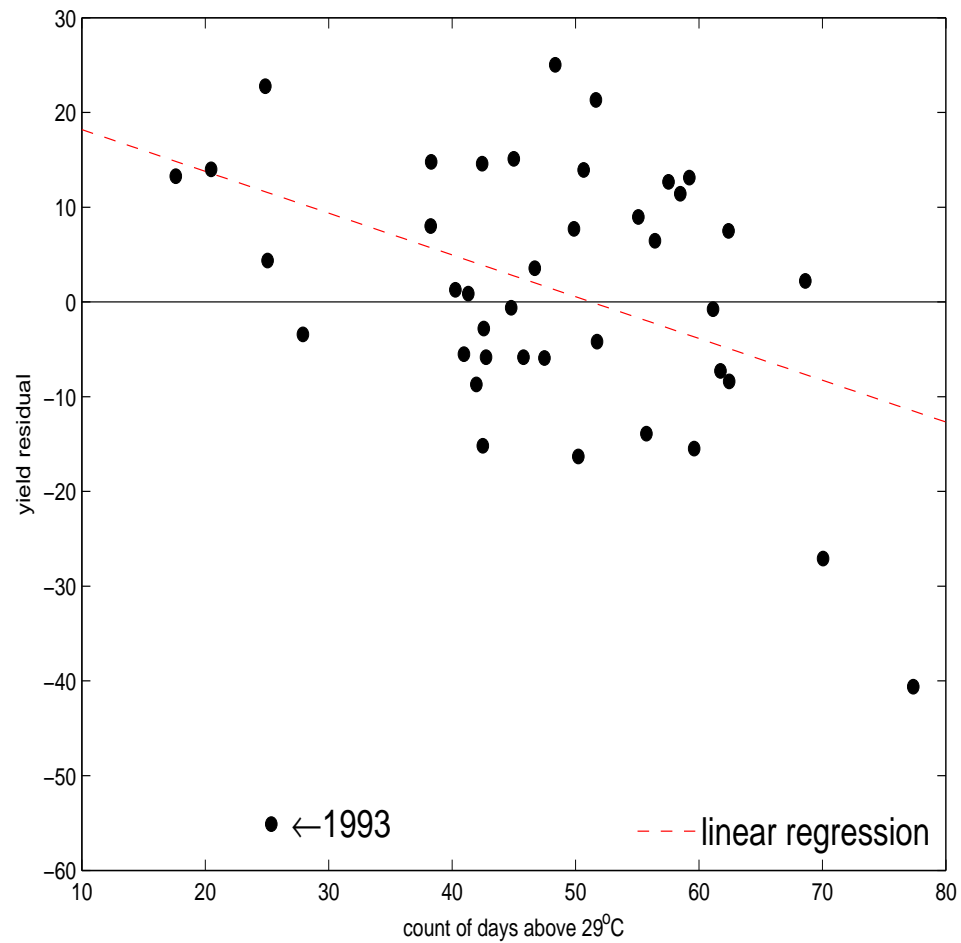


Figure 4.13: Comparison of the number of days over 29°C and the yield residual for Iowa. Note the low residual yield for 1993 when the wet weather damaged crops. A linear regression, red dashed line, was fitted to the data (excluding 1993) as the number of days above 29°C increases the yield falls. See text for discussion of 1993.

4.4. Maize yield by county in Iowa

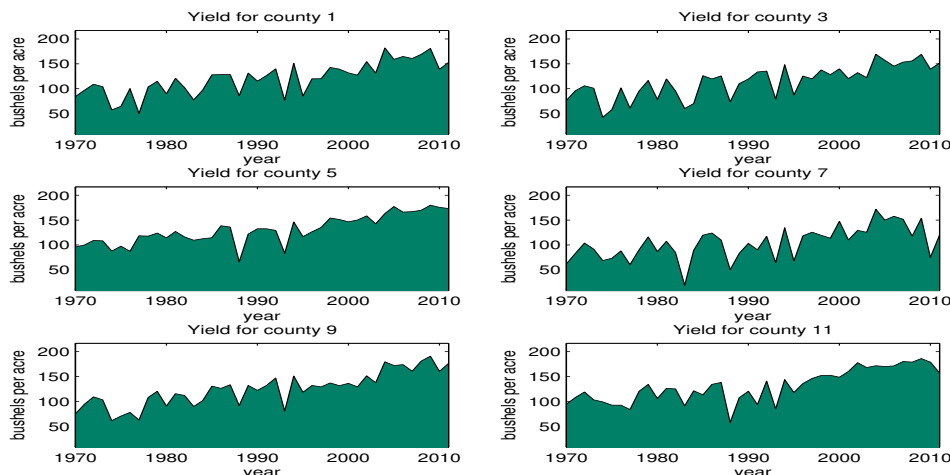
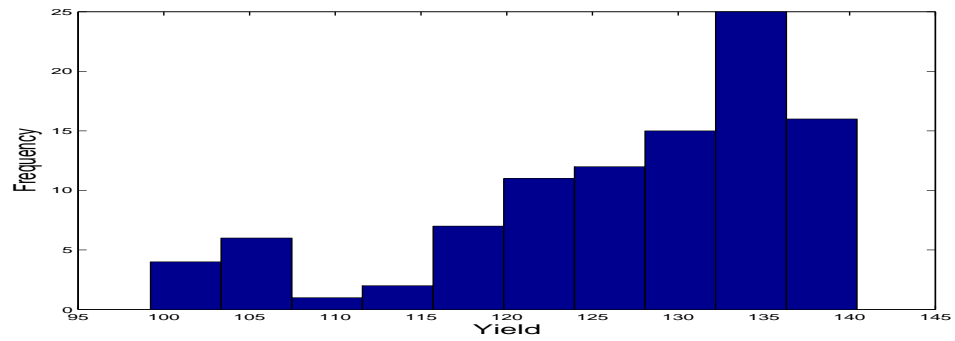


Figure 4.14: Yearly maize yield for six counties in Iowa. For most counties there is a large rise in yield between 1970 and 2012.

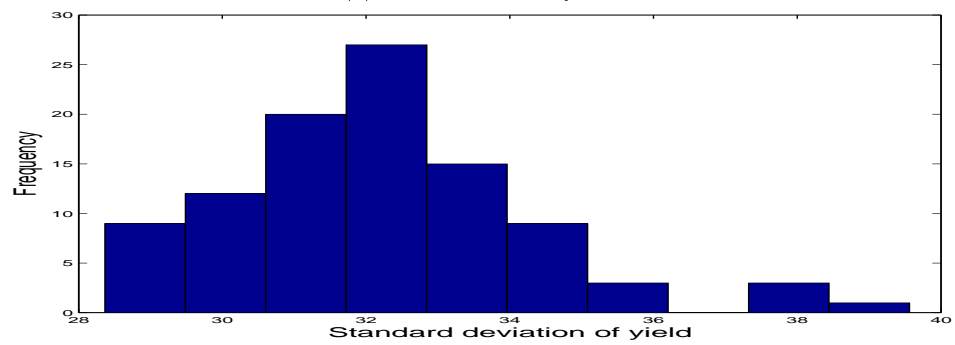
increased. To calibrate the CERES-Maize model a linear regression is fitted to the outcomes by the CERES-Maize modellers [19,37] (see Section 1.5). The slope parameters from fitting a linear regressions to county level maize yield are shown in the histogram in Figure 4.15c. This histogram shows how across the counties the slope level varies between 1.4 and 2.8. The highest values of the slope parameter are for counties coloured orange in the North West of Iowa and the lowest values are for the counties coloured blue in the South of Iowa as shown in Figure 4.18.

4.4.1 How widespread is crop failure?

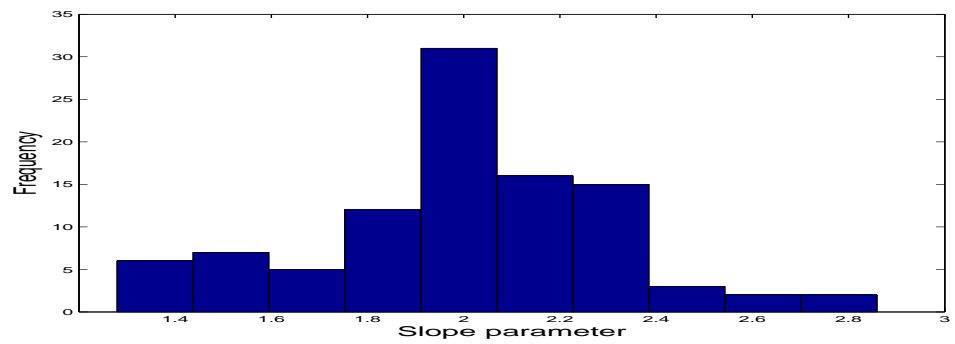
An investigation into how widespread crop failure is across counties in Iowa is made by considering the largest fall in maize yield for each county between 1970 and 2012. The years with the biggest percentage fall in yield by county



(a) Mean of maize yield



(b) Standard deviation of maize yield



(c) Slope parameter of maize

Figure 4.15: Histogram of maize yield in 4.15a shows for most counties yield is towards the upper end of the range. The average standard deviation of maize yield in 4.15b is approximately 32. The slope histogram 4.15c shows most slope parameters are approximately 2.

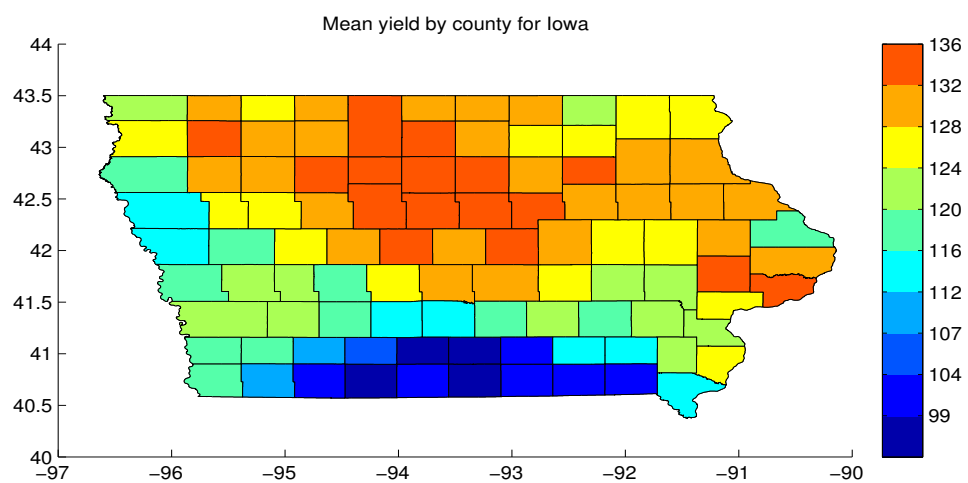


Figure 4.16: Mean yield by county in Iowa. Southern counties have lower mean yields, coloured blue.

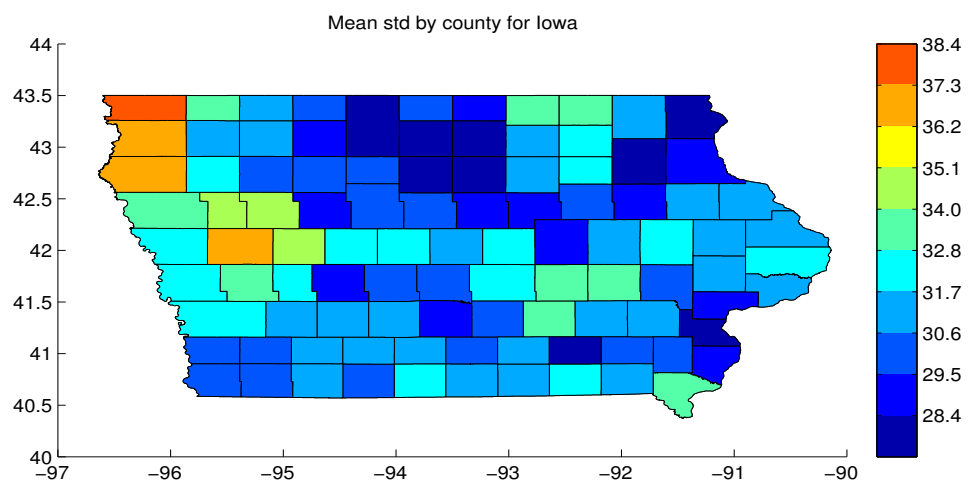


Figure 4.17: Mean standard deviation of maize yield by county in Iowa. The counties in the South West of Iowa have the highest standard deviations coloured orange.

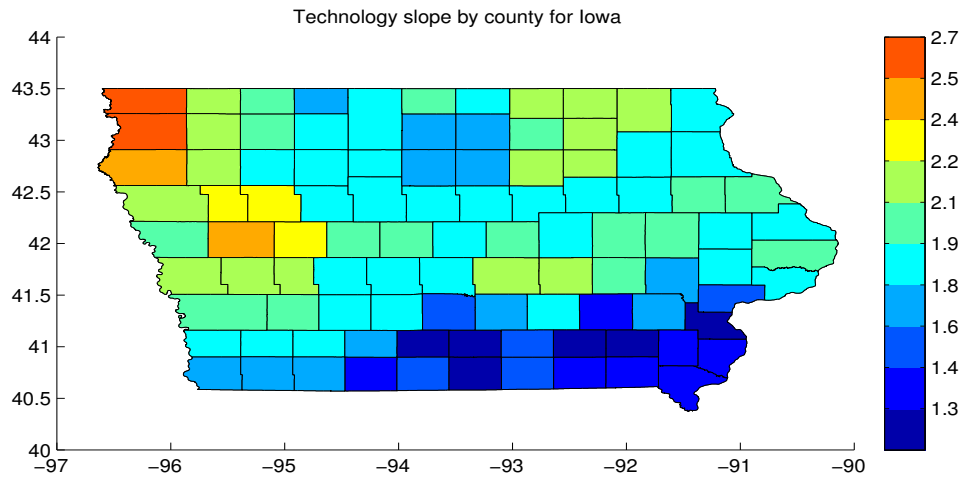


Figure 4.18: Technology slope by county in Iowa. The counties in the North West have the higher slopes (orange) and the counties in the South have lower slopes (blue).

are shown in Figure 4.19. The biggest falls all occur in one of only five years, the years 1974, 1977, 1983, 1988 and 1993. An interesting point from this figure is that in the last 18 years there has not been a record breaking fall in yield for any county in Iowa. The largest yearly fall in yield for any county was a fall of 81.4% in 1977. The percentage fall in maize yield by county was plotted separately for these five years. For these figures, the larger the percentage fall in yield, the darker the blue. A county was coloured white if the yield remained constant or increased. Figure 4.20 shows 1974 when there was a fall in yield across the whole of Iowa but particularly in the Western counties. In 1977, as shown in Figure 4.21, yield fell in approximately half of Iowa, this is the only year where the large fall in yield is so localised. For 1977 a couple of counties near the Southern edge were severely impacted, as shown by the very dark blue. In 1983, Figure 4.22, the worst affected

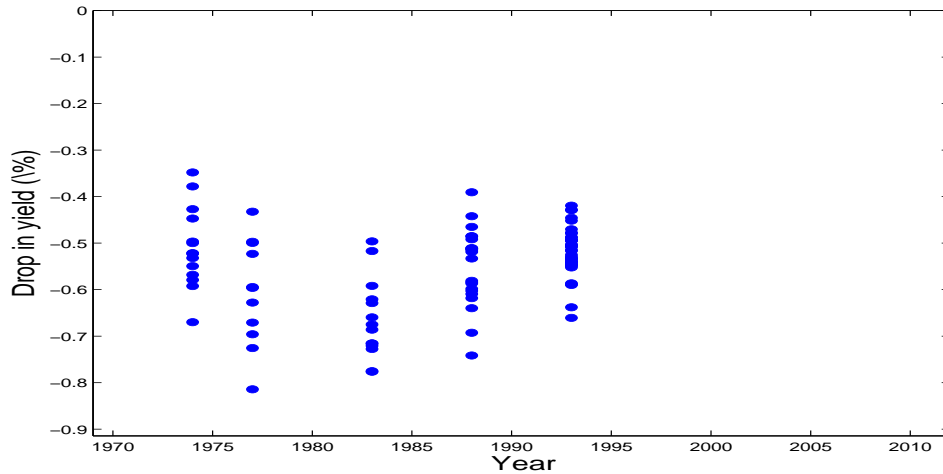


Figure 4.19: For each county in Iowa a blue dot marks the year with largest percentage fall in yield. The years with the largest percentage falls are all found in five years.

counties were in the South East of Iowa. In 1988, Figure 4.23, a larger area of Iowa was severely affected with dark blue colours showing yield falls of 50% or greater across large swathes of West Iowa. However the largest and most widespread falls in yield are in 1993, Figure 4.24, when nearly every county is coloured in dark blue indicating a large fall in yield across the whole of Iowa.

4.4.2 Maize yield in adverse weather years

The annual crop summary from the NASS lists the yield and growing conditions by crop. Listed below are the different adverse weather conditions which were recorded as negatively impacting yield for maize in Iowa.

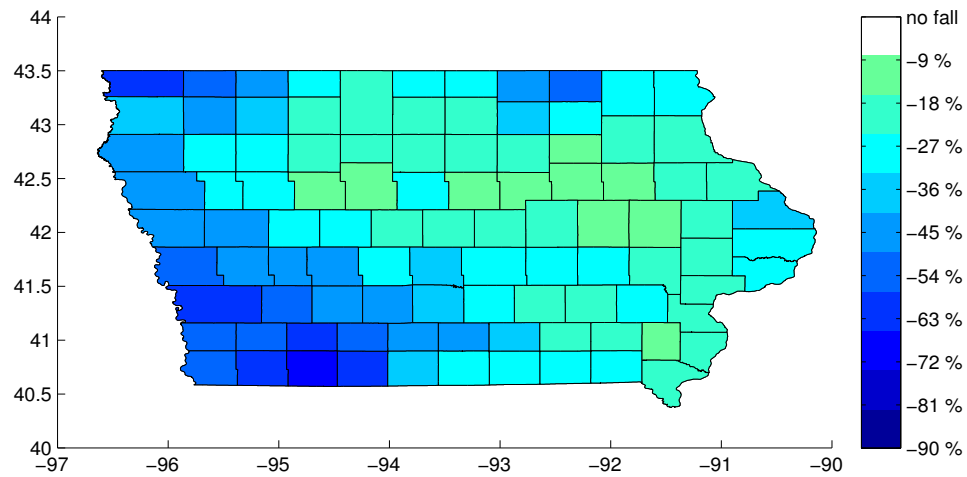


Figure 4.20: Percentage fall in yield by county for Iowa in 1974. The darker the blue, the larger the fall in yield here the largest falls are in the west of Iowa.

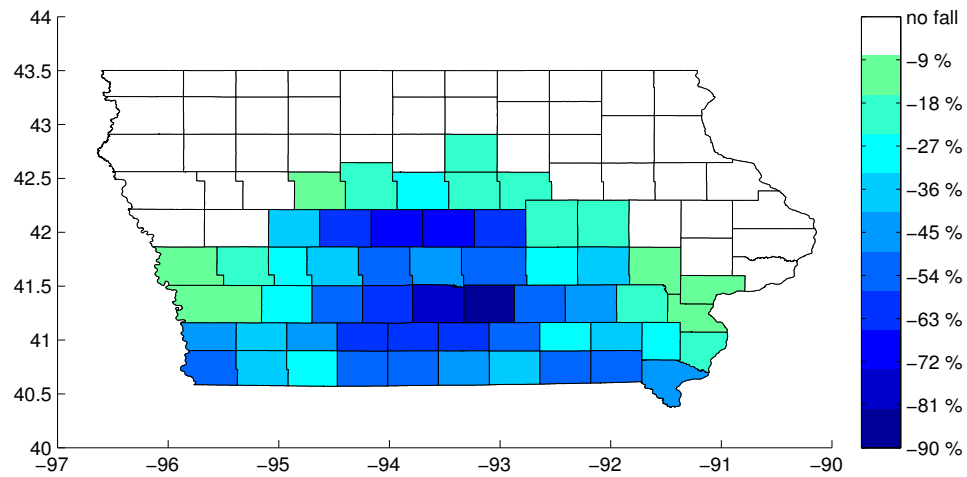


Figure 4.21: Percentage fall in yield by county for Iowa in 1977. Although this year had the largest fall in yield for any county it is more localised than the other years.

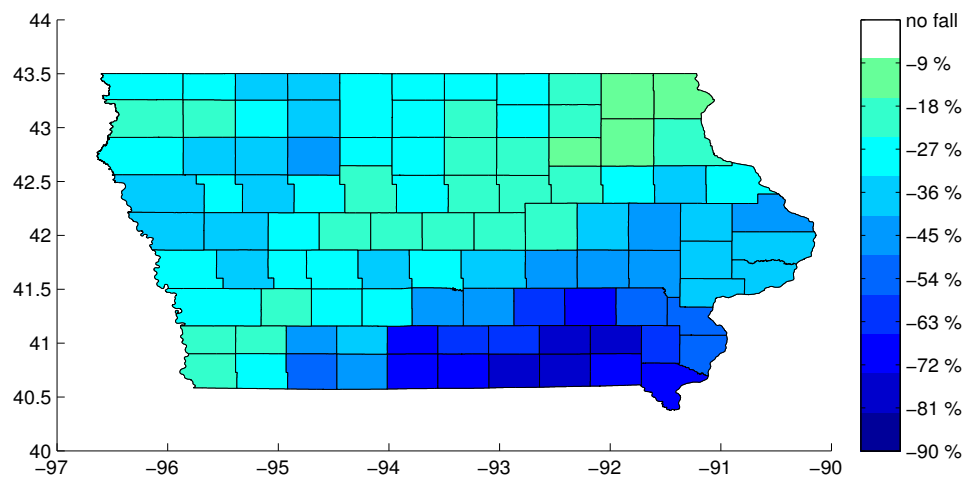


Figure 4.22: Percentage fall in yield by county for Iowa in 1983. The largest falls in yield were for counties in the South East of Iowa.

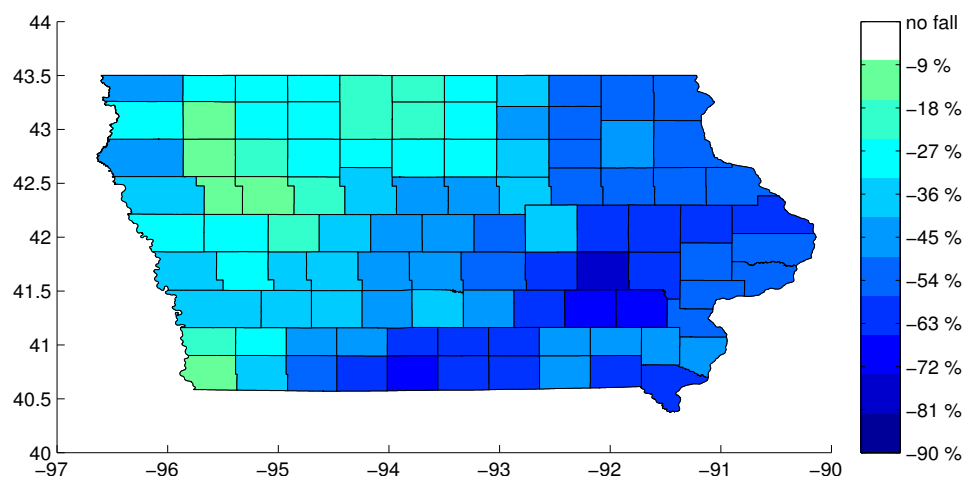


Figure 4.23: Percentage fall in yield by county for Iowa in 1988. The largest falls in yield were for counties in the East of Iowa.

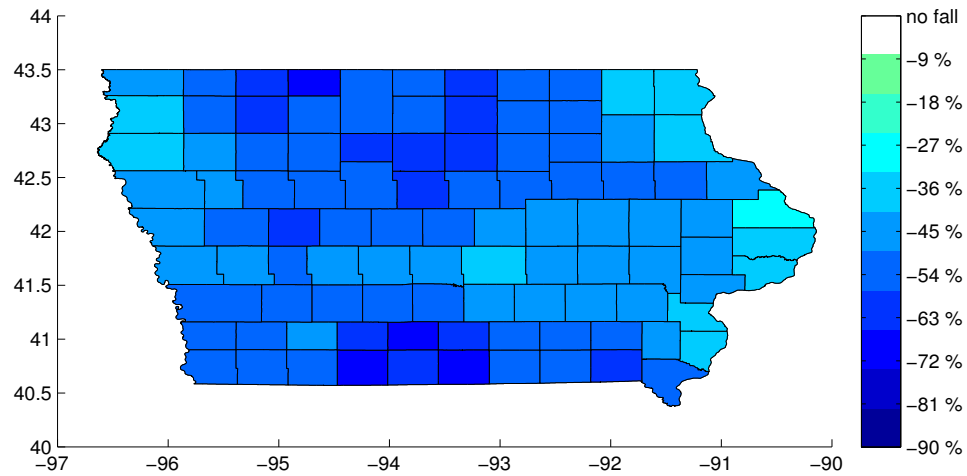


Figure 4.24: Percentage fall in yield b county for Iowa in 1993. All the counties in Iowa have been significantly impacted.

Year	Summary of Extreme Events
1936	Dust bowl year
1945	"Serious obstacles" for maize - late planting due to cold weather and flooding, early/on time frosts
1947	Unfavourable weather, flooding
1951	Smaller harvested area due to abandonment from floods. Crops late due to cool and wet weather and then got frost damaged in September
1955	Drought in late July
1974	Hot dry weather in July caused heat stress
1977	High temperatures in a localised area in Iowa during June.
1983	Hot and dry weather conditions in July and August
1988	Drought in July

Year	Summary of Extreme Events
1993	Excessive moisture in July damaged maize
1995	Hot weather in August, record yield the year before
2012	Widespread drought and extreme temperatures in June and July

Adverse weather causes crop yield levels to fall. In the next section it is considered whether irrigation provides protection from adverse weather events.

4.4.3 Yield from irrigated fields

In the US only four maize producing states have a large percentage of their farm land irrigated, for these the NASS records the irrigated and non-irrigated yields separately. The irrigated states are Colorado, Kansas, Nebraska and Texas. As Nebraska is the 3rd biggest producer of maize in 2012 [78], the data from Nebraska from 1947 onwards has been examined in this section. In Figure 4.25, the irrigated harvested areas are shown in blue and the non-irrigated harvested areas are shown in red. Between 1947 and 2012 Nebraska has changed from mainly non-irrigated to mainly irrigated. The total harvested area is shown as the green line. The percentage of irrigated and non-irrigated land is compared in Figure 4.25 which shows that between 60 to 70% of the harvested area was irrigated between the 1980s

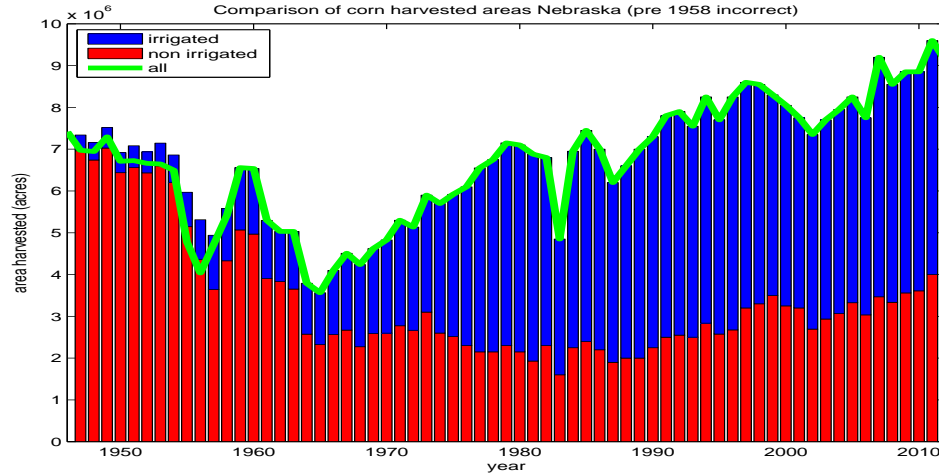


Figure 4.25: Nebraska harvested area (acres) by year for irrigated (blue) and non irrigated (red). The irrigated harvested area has increased from 1946. The green line is the total area. Note the total area is wrong in the early years, this is an error in the NASS figures.

and the present day. In Figure 4.27 the yield weighted between irrigated and non-irrigated yield is green, non-irrigated yield is red and irrigated yield is blue. In the years 1974, 1980, 1993 and 2012 there are noticeable falls in the weighted yield (green line). For the years 1974, 2002 and 2012 however falls in yield only occur in non-irrigated maize yield (red) so for these years irrigation does appear to provide some protection against adverse weather conditions.

In the next section a method to estimate the technical advancements is proposed which takes account of the non-linear increases in the maize yield time series.

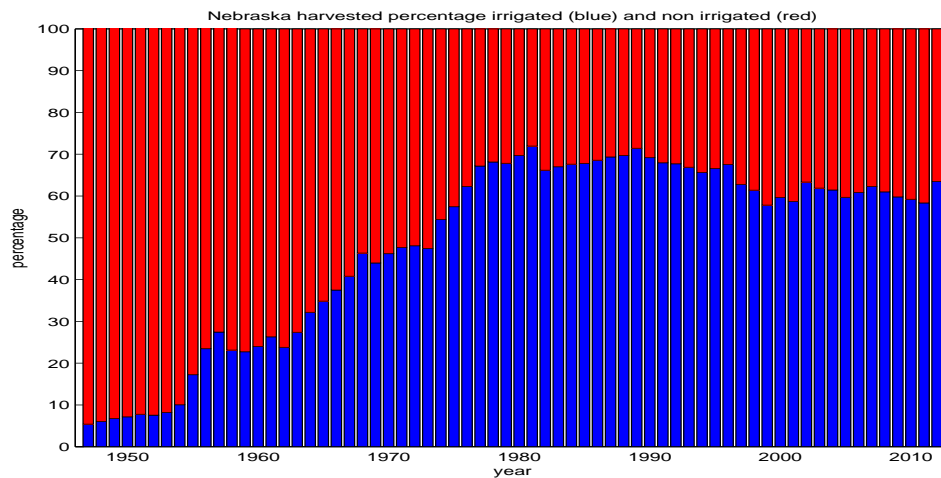


Figure 4.26: Nebraska percentage of harvested area irrigated (blue) and non irrigated (red). The percentage of irrigated area has slightly decreased in the last twenty years.

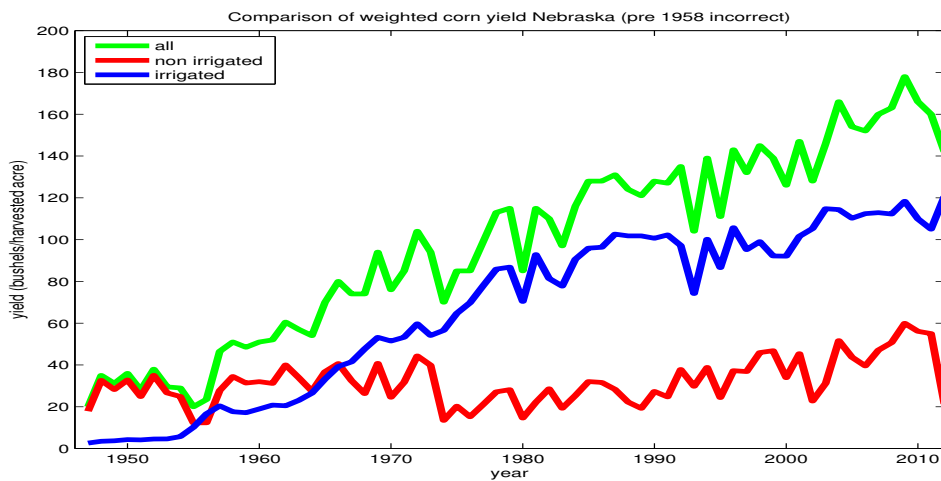


Figure 4.27: Weighted yield curve for Nebraska, total yield (green), irrigated yield (blue), non-irrigated yield (red). Most of the yield increase is from the irrigated lands.

4.5 Introducing the prior return to model technology advancement in maize yield

Technology increase in maize yield is said to have started with the introduction of fertilisers [40]. Is it possible to identify the year where the technology increase first starts? γ is defined as the year where the technology increase starts. One method to identify γ is to find the best fit using a fitted trend[†]. Where the fitted trend is a straight line from 1860 until γ and then a linear regression from γ to 2012. The best fit is found by computing the minimum root mean square error between the fitted trend and the yield.

The root mean square errors for the fitted trend line by γ are compared in Figure 4.28. The best fit is when γ is 1943, however the bottom of the curve is quite flat so similar values of RMSE to 1943 are found for near by years.

$$y_i = \begin{cases} 38.18 & \text{if } i \leq \gamma, \\ 38.18 + 1.92k & \text{where } k = i - \gamma, k > 0 \end{cases} \quad (4.2)$$

This best fitted trend is shown in Figure 4.29. There is a gap between the fitted line and the observed yield at the “elbow”.

Is it possible to differentiate between increases caused by favourable growing

[†]thanks to Piotr Fryzlewicz for this suggestion.

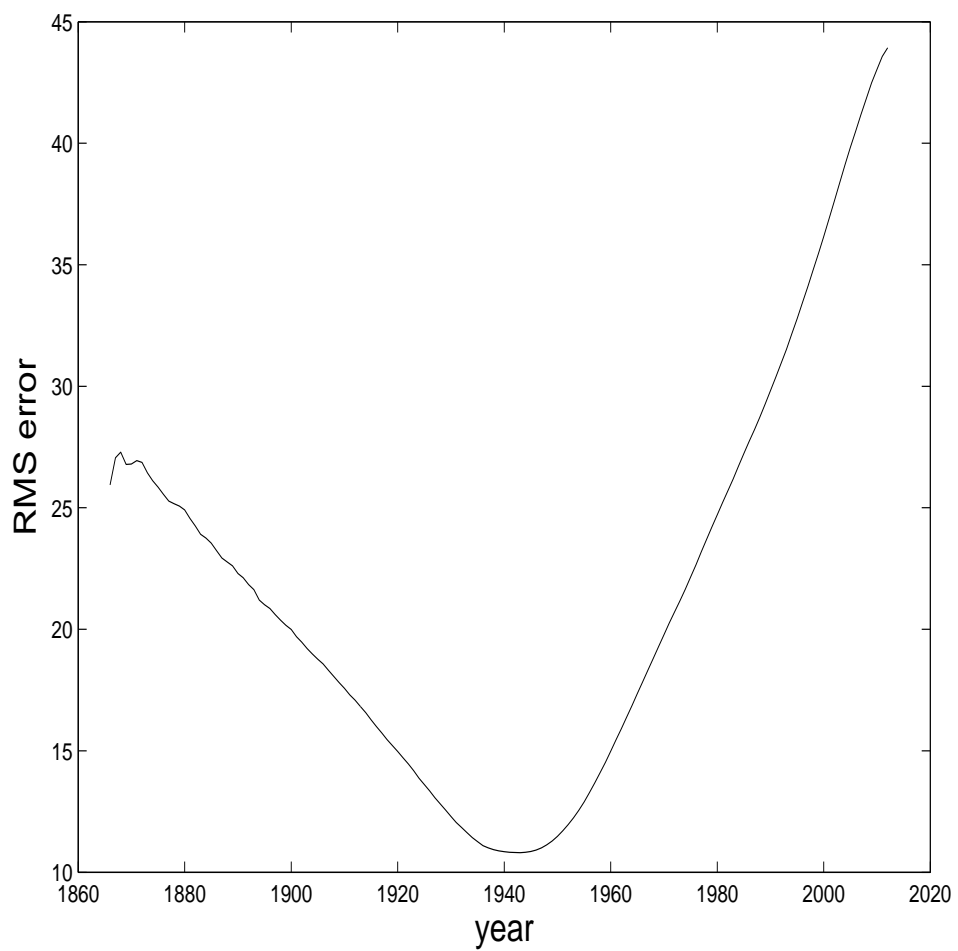


Figure 4.28: The root mean square error between the fitted trend line with the “elbow” at each year and the observations by year for US maize yield time series. Notice how the bottom of the curve is quite flat.

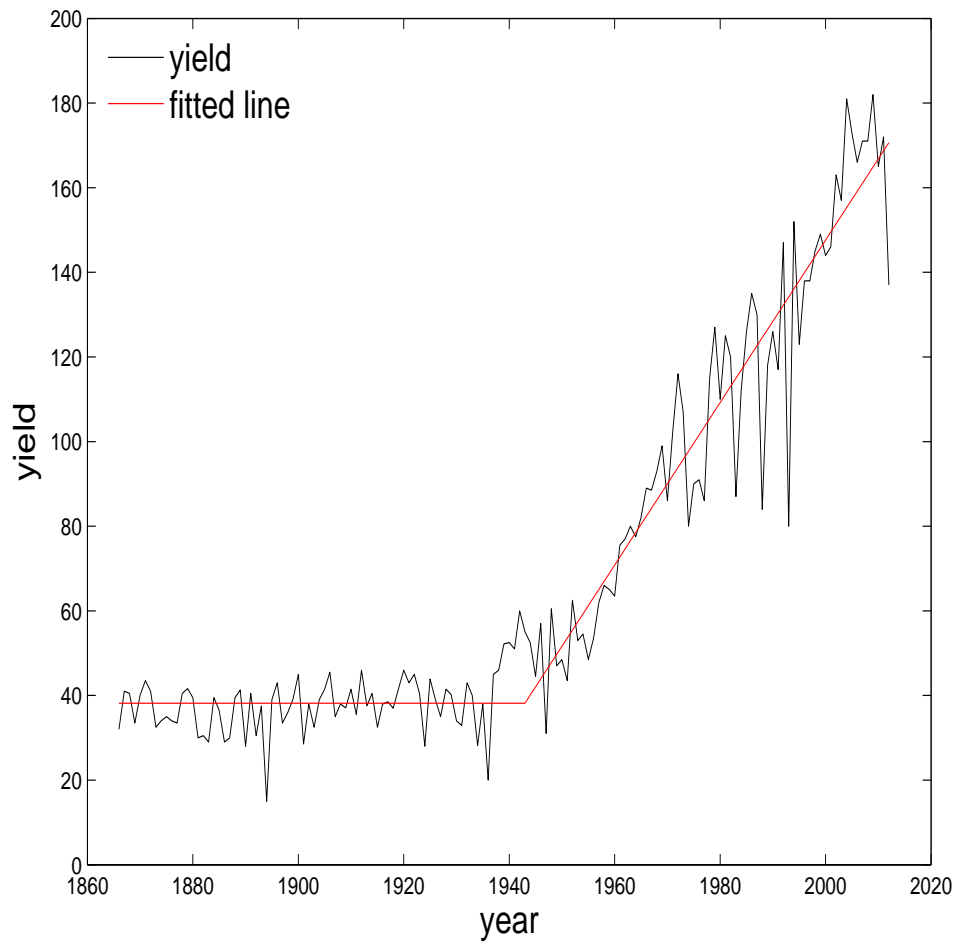


Figure 4.29: Identifying the year (γ) when technical advances start. The RMSE between the best fitted trend line and data is at a minimum when $\gamma = 1943$. Note there is quite a large gap between the “elbow” of the fitted trend and the yield.

conditions from increases caused by technical advancements? A new method to approximate the non-linear technical advancements using prior ratio is proposed where the prior ratio is:

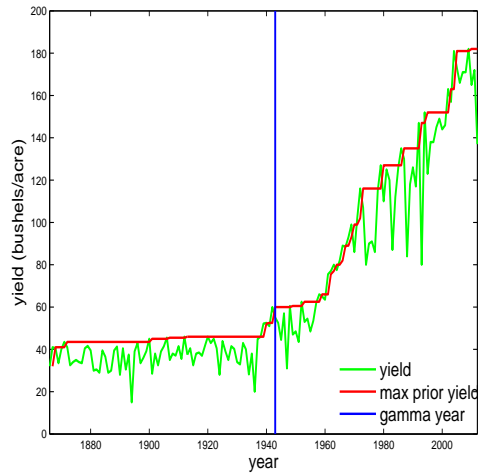
$$r_i = \frac{y_i}{z_j}, \quad j < i \quad (4.3)$$

where y_i is the outcome for the i^{th} year, z_j is the maximum prior yield, $\max(y_1, \dots, y_{i-1})$, for the years $j = 1, \dots, i - 1$ and r_i is the maximum prior ratio for the i^{th} year.

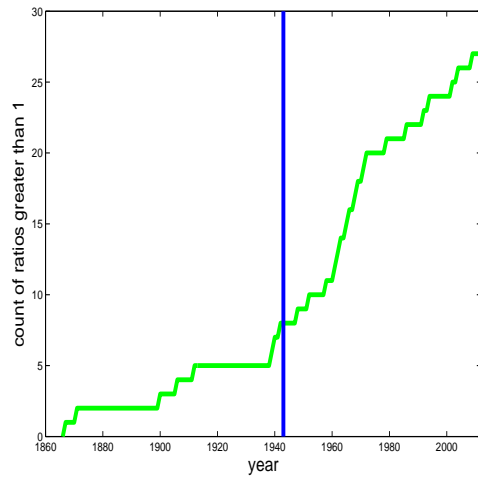
What does the prior ratio (r_i) mean? A prior ratio (r_i) greater than 1 means the yield has reached a new high. How much of this new high has been caused by technical advancements and how much has been caused by favourable weather conditions on top of pre-existing technical advancements, however, is difficult to determine. A prior ratio of less than 1 shows that the maximum yield has not been exceeded. A prior ratio significantly below 1 shows that there were adverse crop growing conditions, such as a drought, in that year. Prior ratios are examined at state level and county level for Iowa below.

4.5.1 State level

For Iowa the maize yield time series (green) and the maximum prior yield (red) are illustrated in Figure 4.30a. The year where technical advancements



(a) Maximum prior yield



(b) Step function of maximum prior yield greater than one

Figure 4.30: Figure 4.30a: Comparison of the state yield (green line) with the maximum prior yield (red line) for Iowa. The blue line marks γ the start of the trend. Notice there is a steep rise in yield between 1961 and 1973. Figure 4.30b: A step function of the years where the prior yield (r_i) was greater than one for all the states in Iowa. Note that between 1961 and 1972 nearly every year had an increase in yield, which is unlikely to be caused just by favourable growing conditions.

is estimated to begin is the vertical blue line. The maximum prior year (red line), shows two time periods where there were significant increases in yield. The first of these was between 1961 and 1973 when yield increased from 66 bushels per acre to 116 bushels per acre, an increase of 50 bushels per acre over 12 years, in fairly regular increments. The second large change in yield was between 2002 and 2005 when yield increased from 152 to 181 bushels per acre. From 2005 the maximum prior yield has hardly moved from 181 bushels per acre implying there have been no significant technical advancements in the last 8 years.

The step function increased by one each year the prior ratio (r_i) was greater than one, as shown in Figure 4.30b in green. The start of the trend, γ , is the vertical blue line. This figure shows slightly different information from Figure 4.30a as the green line increases uniformly every year the prior ratio (r_i) is greater than one, there is no indication about the size of the prior ratio increase. From the Figure 4.30b the two steepest sections of the step function are between 1937 and 1942, and between 1961 and 1972. The stagnation from 2005 onwards is not so obvious as a small increase in the maximum prior yield from 181 to 182 bushels per acres in 2010 makes the step function increase.

4.5.2 County Level

At county level yield is not available before 1926, so the time series is shorter.

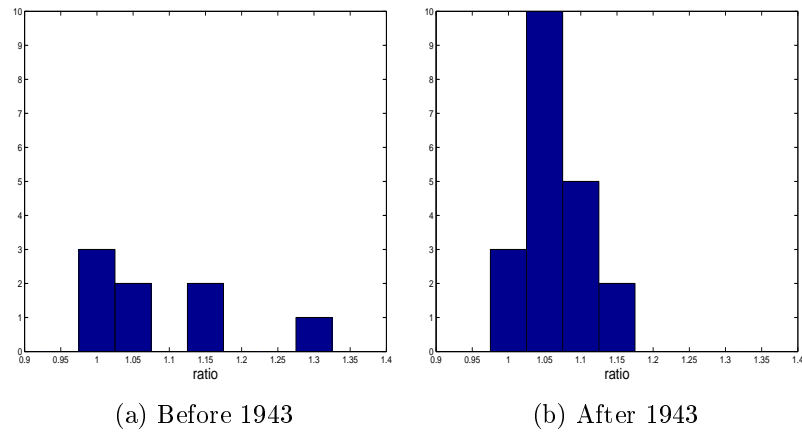


Figure 4.31: Histogram for prior ratios greater than 1 for counties in Iowa by 4.31a early years and 4.31b later years. Notice how the later years have more prior ratios greater than 1 than the early years indicating that there was more technological advancements in the later years.

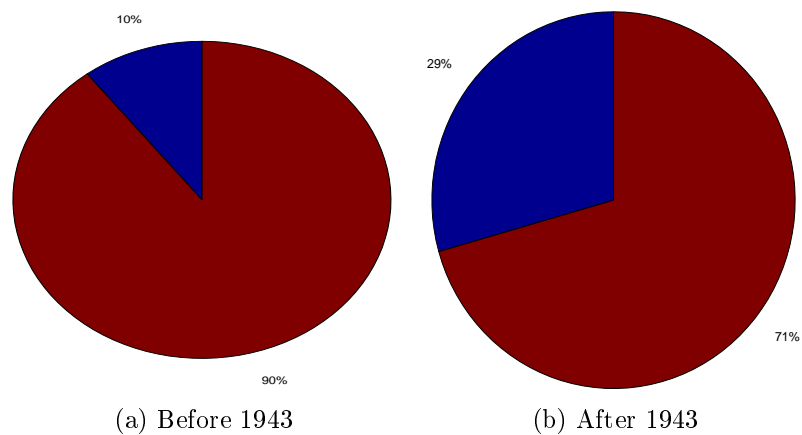


Figure 4.32: Iowa percentage of years where the ratio is greater than 1 (blue) and less than 1 (red) for the early years (left) and the later years (right). Note that there are significantly more ratios greater than 1 after 1943.

In Figure 4.33 the maize yield by individual county (green dot) is plotted with the state level maximum prior yield (red line). The county level time series contains significantly more variability than the state level time series.

The prior ratio (r_i) from Equation 4.3 is considered for each county and for each year in Figure 4.34. There are a few years where all the ratios are less than one, such as 1936 (the dust bowl year), 1946, 1974, 1983, 1993 and 2012. These are all years with well known and widespread adverse weather conditions. Of more interest are the years where the majority of the individual county prior ratios are greater than one; there are only a few of these years: 1972, 1992 and 1994. It would be interesting to determine if these years had particularly good growing weather or if these years saw the introduction of significant improvements in technology[†].

The distribution of prior ratios greater than one is shown in Figure 4.35: the distribution before 1943 has a similar shape to after 1943. The percentage of both groups that have prior ratios greater than one are shown in Figure 4.36; note there is little difference in the fraction pre-1943 (27%) and post-1943 (28%). This is unexpected as a prior ratio greater than one is either from an improvement in technology or good weather that has boosted the yield for that year and the number of prior ratios greater than one would have been expected to be significantly larger in post-1943 group. The γ year estimated from the shorter county level time series is 1947. The fraction of prior ratios

[†]This is beyond the scope of this thesis.

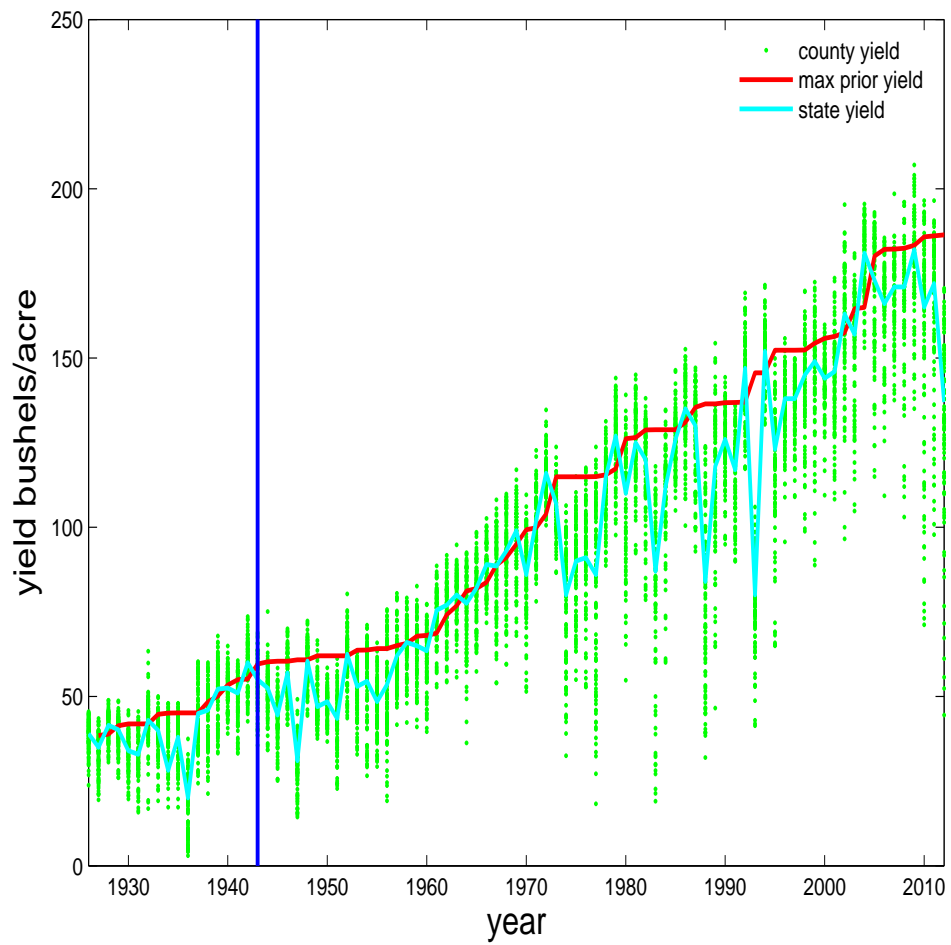


Figure 4.33: Comparison of the state level maximum prior yield (red line), the state level yield (green line) and the county level yield (green dots). There is significantly more variability in county level time series than the state level time series.

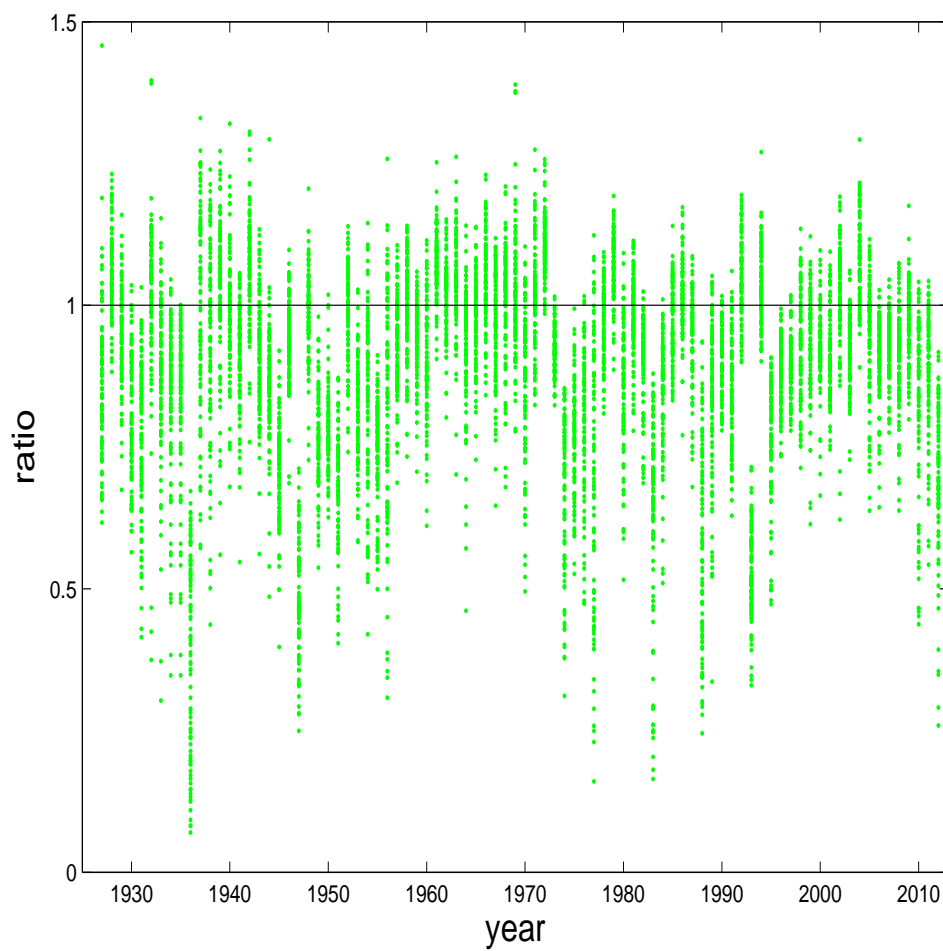


Figure 4.34: Prior ratios for all the counties in Iowa by year. The prior ratios less than one have a bigger downward movement than the ratios greater than one have upward movement.

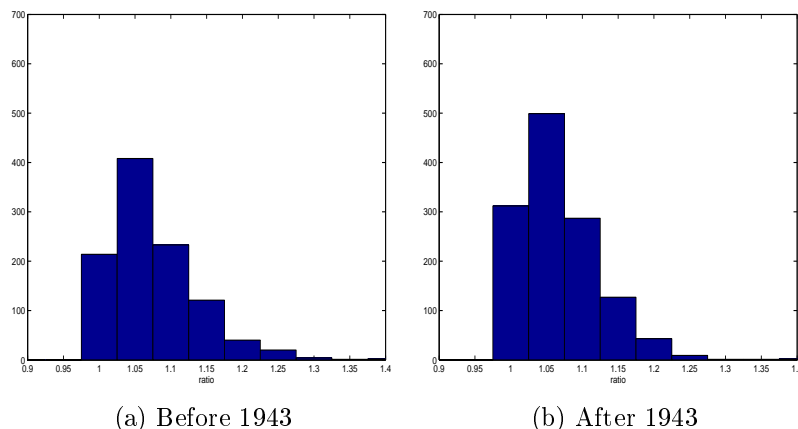


Figure 4.35: Histogram of the prior ratios greater than one for all the counties in Iowa. Notice how most prior ratios are just greater than one.

greater than one in the pre-1947 and post-1947 remained almost identical to Figure 4.36. Dividing the time series by prior ratios greater than one has not been a successful method in separating years with technology increases from years with good weather after technology increases.

4.6 Conclusions

1. Comparisons of the variability of meteorological observations across different counties in Iowa is presented graphically in Section 4.1.2 and 4.1.4.
2. An examination of how the highest annual daily t_{max} and precipitation impact maize yield is in Section 4.3. Effects of technical advancements were removed from the maize yield time series by subtracting the linear regression and considering just the residuals. The residuals were divided by size into three equally likely bins. Comparisons of t_{max} , pre-

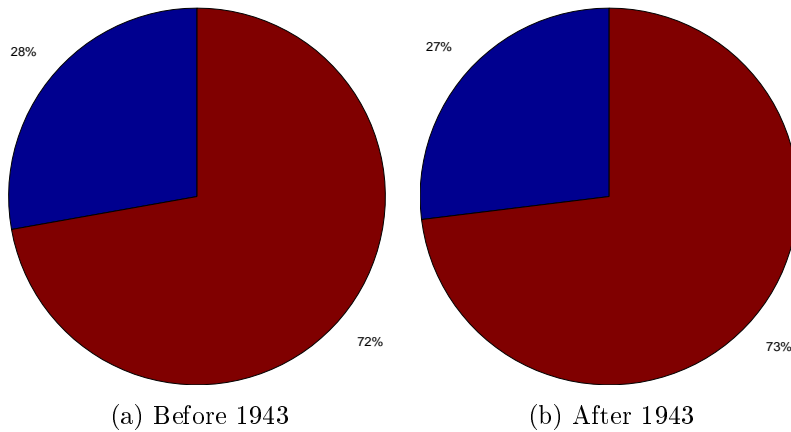


Figure 4.36: Percentage of years where the prior ratio is greater than one (blue) and less than one (red) for all the counties in Iowa. Notice how this is almost identical for years both sides of the technology increase year.

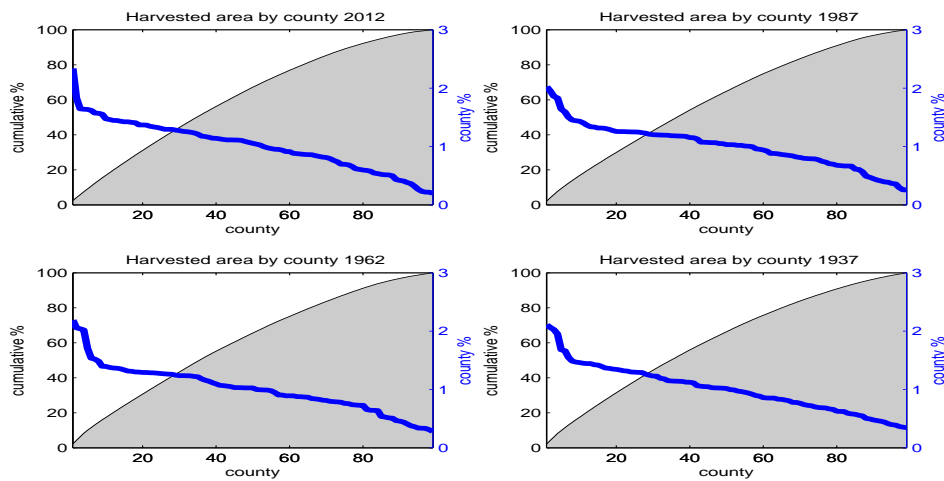


Figure 4.37: These plots each show two things for the years 2012, 1987, 1962 and 1937 (1) What percentage of the total harvested area each county is to the state of Iowa which is recorded on the right y-axis in blue. (2) The cumulative of harvested area percentage by county is recorded on the left y-axis in grey. Across all the years the amount harvested in each county is fairly consistent

precipitation and the bin size are made in Figure 4.12, which shows both high t_{max} with low precipitation and low t_{max} with high precipitation negatively impact the yield.

3. The variation of maize yield mean and standard deviation is presented graphically across counties in Iowa in Section 4.4.
4. A study of how technology advancements differ between counties in Iowa by comparing the linear regression slope parameter is in Section 4.4.
5. A comparison of county level patterns of crop failure by considering the largest percentage falls in maize yield in each county is in Section 4.4.1.
6. The prior ratio method is a novel simple statistic used here to detect technical advancements and is discussed in Section 4.5.

Chapter 5

Creating initial conditions for gridded weather

The CERES-Maize model uses one approximation of weather across the maize growing area to estimate yield [19, 37] as discussed in Section 1.4. This weather approximation contains uncertainty, explained in Section 1.6, which impacts the crop model's estimate [77, 95]. To see the impact from this observational uncertainty on the yield estimate, an ensemble of initial conditions can be input into the crop model. The use of ensembles to create a probabilistic forecast is demonstrated for chaotic dynamical systems in Chapter 2. In this chapter an ensemble is made which captures sampling uncertainty from one weather approximation. This ensemble is designed to be read directly by the CERES-Maize model run by Dr Joshua Elliott, University of Chicago, who will use this in future work to better understand the uncertainty in the crop model estimates. The ensemble of crop

1	2	3
8	0	4
7	6	5

Figure 5.1: One grid and the lay out of the cells surrounding cell number 0

yield estimates will be converted into a probabilistic forecast using standard kernel dressing [5]. The skill of this ensemble forecast can then be compared against the skill of a one member ensemble forecast.

The gridded data analysed is from Iowa and contains twenty years of daily data from 1st January 1980 to 31st December 2010 [20, 33, 53]. The observational uncertainty in this data is estimated by considering differences in meteorological data between individual cells. One grid contains nine cells in a 3 by 3 pattern. The centre is defined as cell 0 and the border as cells 1 to 8, as shown in Figure 5.1:

5.1 Uncertainty in gridded minimum temperature

For one grid in the centre of Iowa the differences in minimum temperature (t_{min}) between cell 0 and the border cells are calculated. The results are shown in Figure 5.2 as histograms. Cell 3 (Figure 5.2c) has mostly positive differences, so t_{min} , is slightly warmer in cell 3 than cell 0. In cell 7 (Figure

Table 5.1: Mean differences in t_{min} from border cells and cell 0 for one grid in Iowa.

grid	mean diff (K)	mean std
1	0.21	0.72
2	0.36	0.57
3	0.74	0.74
4	-0.52	0.68
5	0.37	0.47
6	-0.59	0.81
7	-0.10	0.37
8	0.26	0.57

5.2g) most of the differences lie close to zero with little spread, so t_{min} is very similar to the central cell. The mean and standard deviation of these differences are in Table 5.1. Cell 3 with the highest mean of 0.74 is the cell that least resembles cell 0. Cell 7 with a mean difference of -0.10 most resembles cell 0. A chi squared test on differences from the eight cells show that none have a Gaussian distribution, however for simplicity when estimating t_{min} later it is assumed they all have a Gaussian distribution.

To estimate the uncertainty in t_{min} gridded data for Iowa, the data is divided into 16 equally sized areas, all squares. In each area the differences between cell 0 and the border cells are calculated. As there is seasonality in t_{min} the data is further divided up by month before a Gaussian distribution is fitted.

Histograms of the 16 monthly Gaussian parameters for t_{min} by month are shown in Figures 5.3 and 5.4. The monthly μ parameters are all centred around 0. The monthly σ parameters appear fairly flat between 0.5 and 2.5.

To create an ensemble of initial conditions N_e , where N_e is 9, random perturbations are added to the initial condition, t_{min} at cell 0. The perturbations

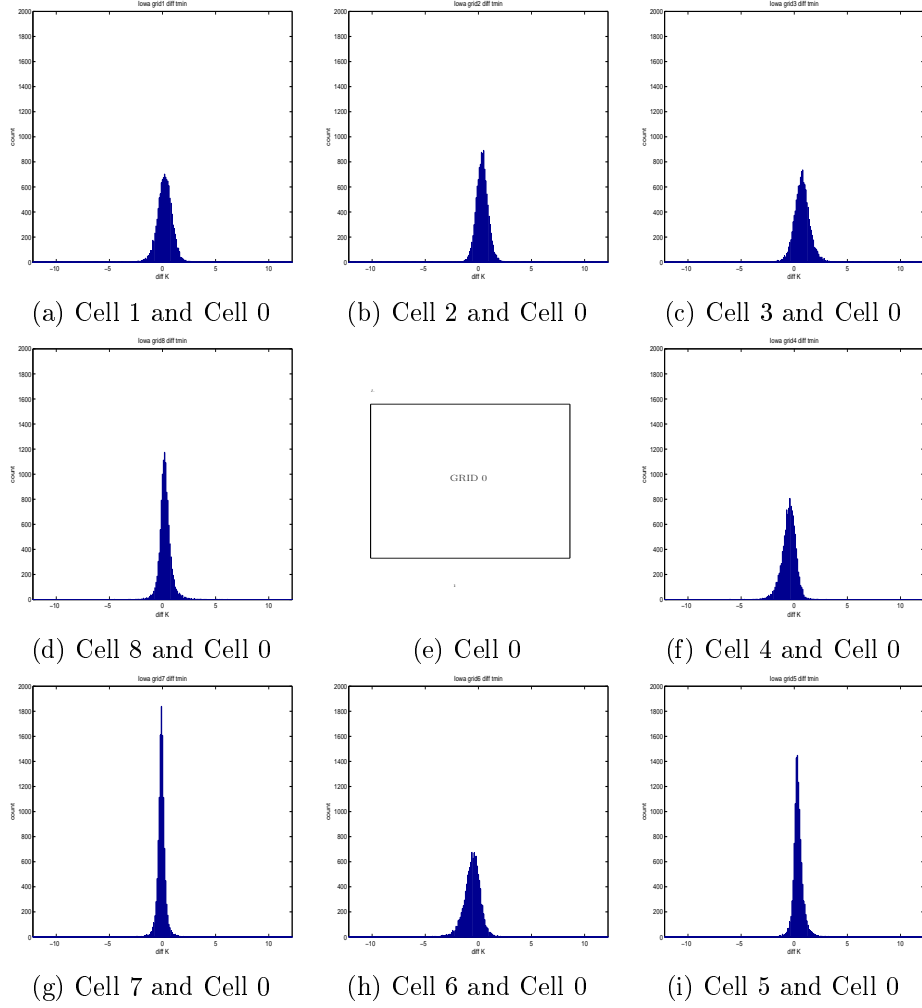
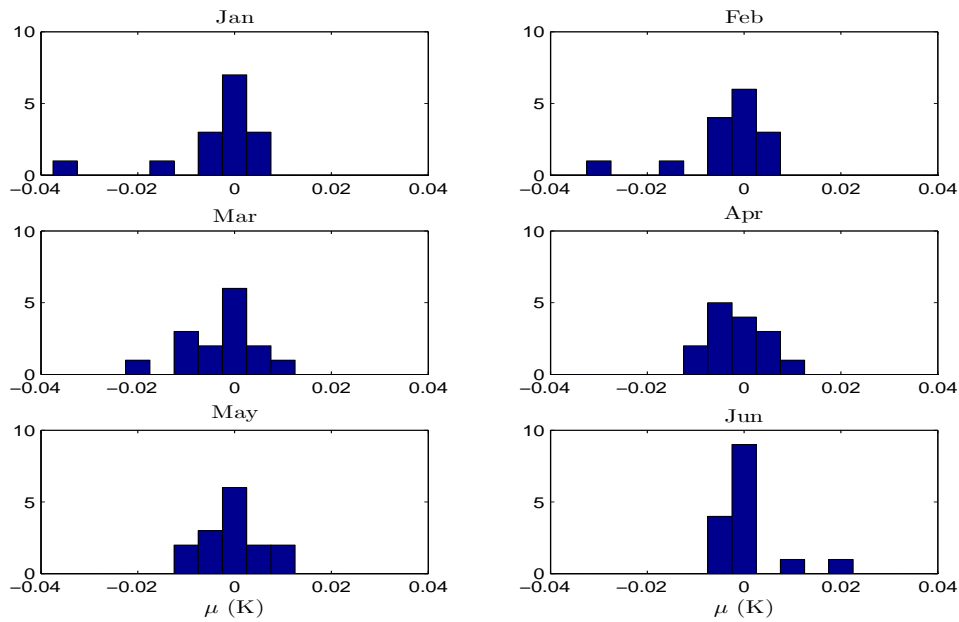
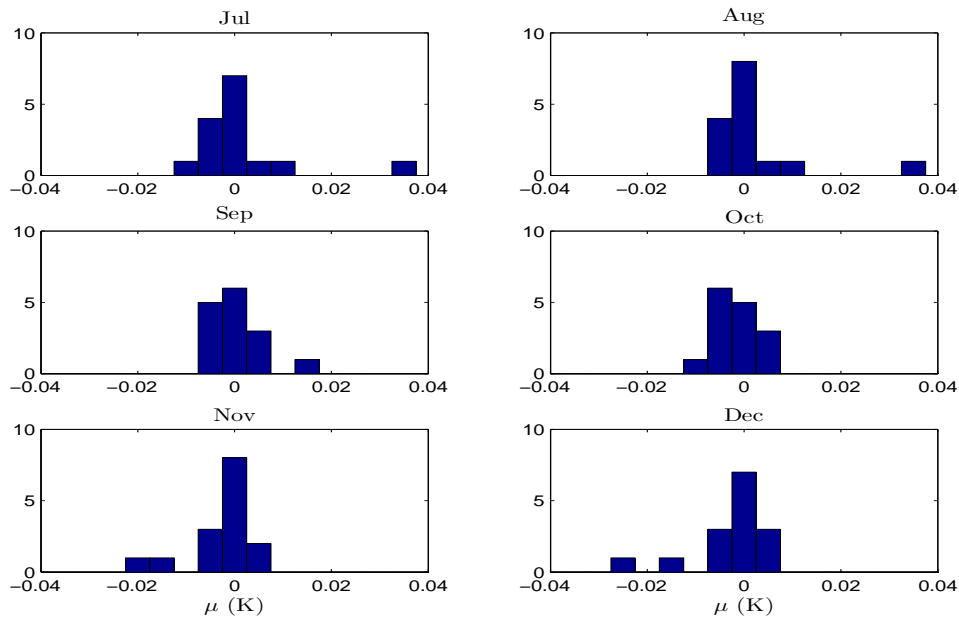


Figure 5.2: Differences in gridded t_{min} between border cells and cell 0 for one grid in the centre of Iowa. The expected difference is zero.

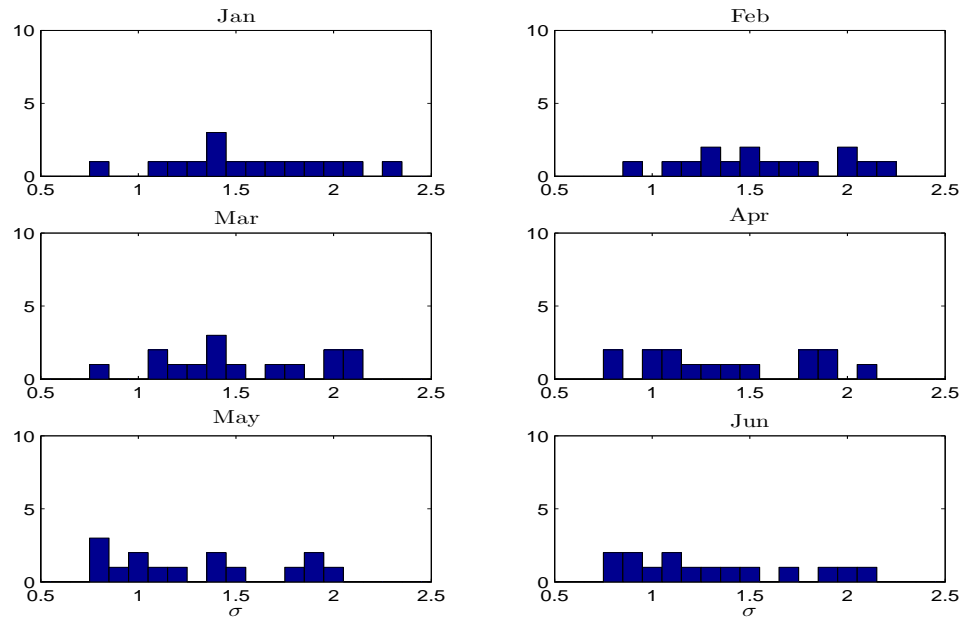


(a) January to June

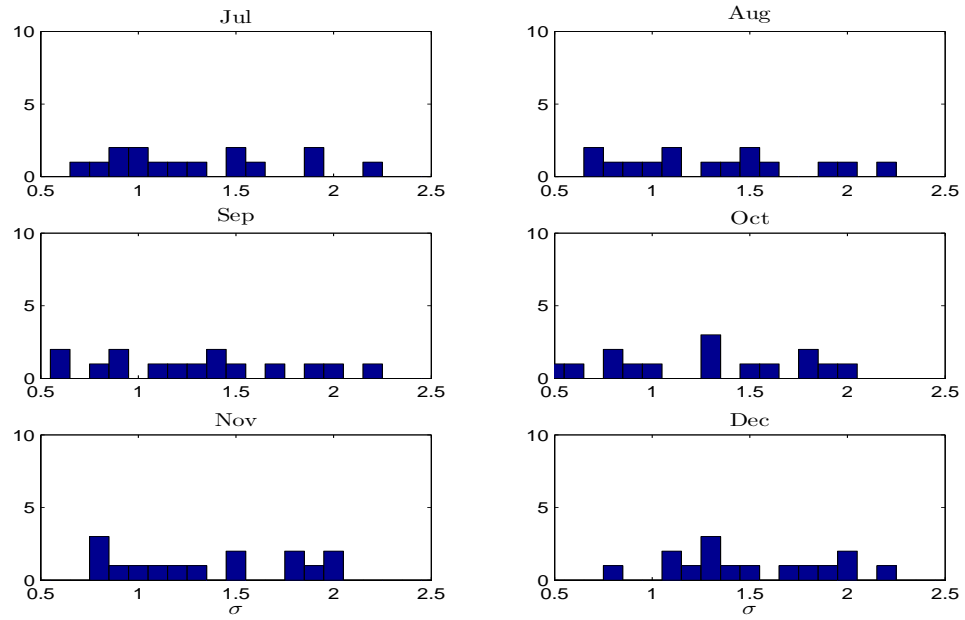


(b) July to December

Figure 5.3: μ for Iowa gridded t_{min} by month for each of the 16 areas. The μ are mainly clustered around 0.



(a) January to June



(b) July to December

Figure 5.4: σ for Iowa gridded t_{min} by month for each of the 16 areas. The monthly σ is spread out fairly evenly across all the months.

are drawn from an assumed Gaussian distribution, with μ and σ parameters determined uniquely for each month and area, see Equation 5.1.

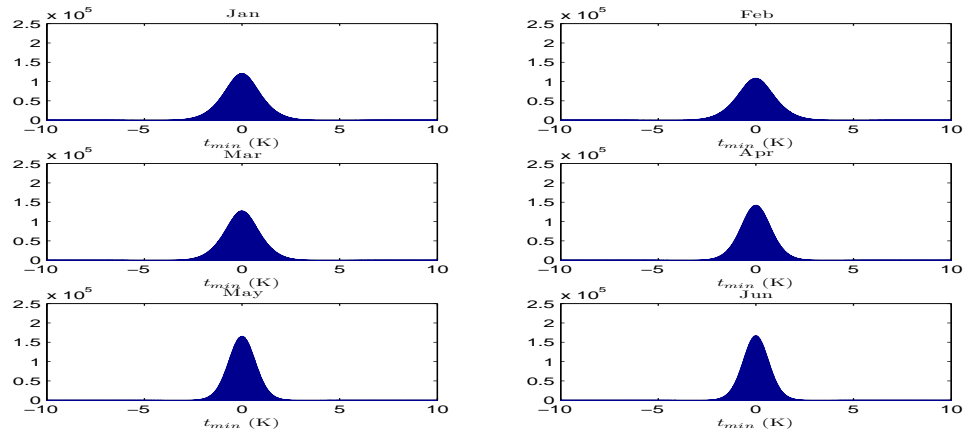
$$x_i^j = g_i + \zeta_i^j \quad (5.1)$$

where t_{min} value for the i^{th} cell point is defined as g_i , the i^{th} cell point perturbations are ζ_i^j for $j = 1, \dots, N_e$ and N_e is the number of ensemble members, 9. Here ζ_i^j is randomly drawn from the Gaussian distribution $N(\mu_{u,v}, \sigma_{u,v})$, where u is the month $u = 1, \dots, 12$ and v is the area for $v = 1, \dots, 16$.

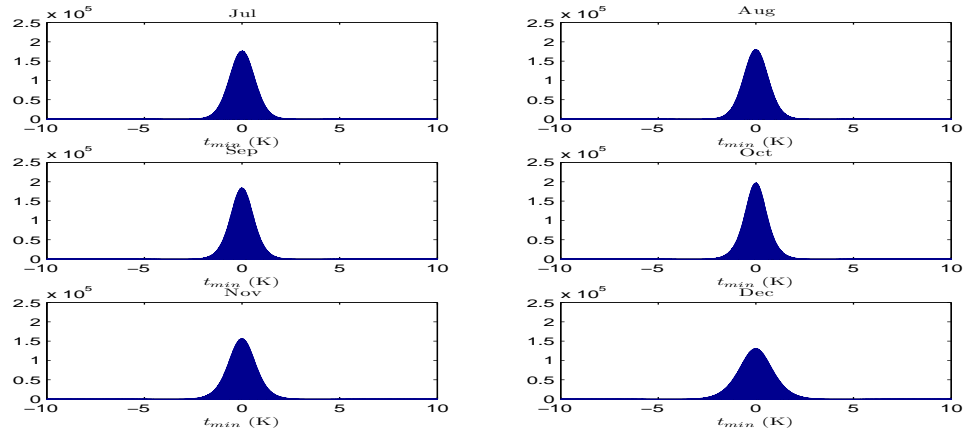
A comparison between gridded t_{min} and the 9 member ensemble is in Figure 5.5, this is from twenty years of daily gridded data. The smallest differences are in September and October where the histogram is at its narrowest and the largest differences are in January and December.

5.2 Uncertainty in gridded maximum temperature

Gaussian parameters used to estimate the uncertainty in gridded t_{max} are calculated in the same way as the gridded t_{min} uncertainty. Histograms of the sixteen Gaussian parameters by month are in the Appendix E in Figures E.1a and E.1b. For each area and for each month the value of the μ is consistently close to zero. The value of σ is spread between a minimum of 0.5 and a maximum of 2.5 for all the months. When the Gaussian parameters from estimating uncertainty in gridded t_{min} and t_{max} are compared they appear very similar as all the means are approximately zero and the σ are



(a) January to June



(b) July to December

Figure 5.5: Difference between the ensemble and t_{min} gridded data for the first six months (5.5a) and the last six months (5.5b). The differences centre around 0.

evenly spread.

5.3 Uncertainty in gridded precipitation

Creating a realistic ensemble for gridded precipitation is not as straightforward as for temperature. With precipitation there are dry days or wet days. Precipitation for wet days is normally approximated by a stochastic weather generator using either a two parameter gamma distribution or a semi-empirical distribution [69, 76, 99, 100]. To generate a gridded precipitation ensemble it is not the precipitation that needs to be estimated, but the uncertainty in the gridded precipitation. This uncertainty is estimated by drawing ensemble members using the distribution from the border cells.

In this thesis uncertainty estimates are calculated separately for the precipitation state of cell 0. The precipitation states are either 0 if it is a dry day, or 1 if it is a wet day. Gridded precipitation data for Iowa are divided into four equally sized areas by month and precipitation state. To estimate an ensemble member for precipitation the following is needed.

1. The probability of a wet day and the probability of a dry day for the border cells of cell 0
2. Precipitation from border cells sorted by volume into nine equally likely bins
3. For the bin containing the highest volumes of precipitation, parameters from fitting an exponential distribution

The flow chart in Figure 5.6 shows the steps for selecting a precipitation ensemble member when cell 0 is a dry day.

5.3.1 Estimating the probability for a dry or wet day

To estimate probabilities for a particular area v , month u consider the case where cell 0 is dry. Figure 5.7a shows the probability that the border cells are dry given that cell 0 is dry, the probability is high ranging from 87% to 95%. If cell 0 is dry, it is less likely that the border cells will be wet, with probability ranging from 8% to 14% in Figure 5.7b. The probabilities display seasonality, for example in areas 3 (red) and 4 (light blue) between June and August the probability of being dry is lower than the rest of the year. When cell 0 is wet a different set of probabilities is calculated.

The ensemble member is chosen by randomly drawing 0 or 1 from the calculated probabilities. Drawing a 0 means the ensemble member becomes 0, and drawing a 1 means the precipitation amount needed to be estimated. Which set of probabilities to use is determined by the state of cell 0.

5.3.2 Estimating nine equally likely bins

The gridded data is divided into two sets; one set when cell 0 is dry and another set when cell 0 is wet. Any precipitation from the border cells is collected and divided into nine equally likely bins where each bin has equal counts of precipitation.

The boundaries of the nine equally likely bins are shown in Figure 5.8a for precipitation bins 1 to 8 with the much wider 9th bin shown separately in

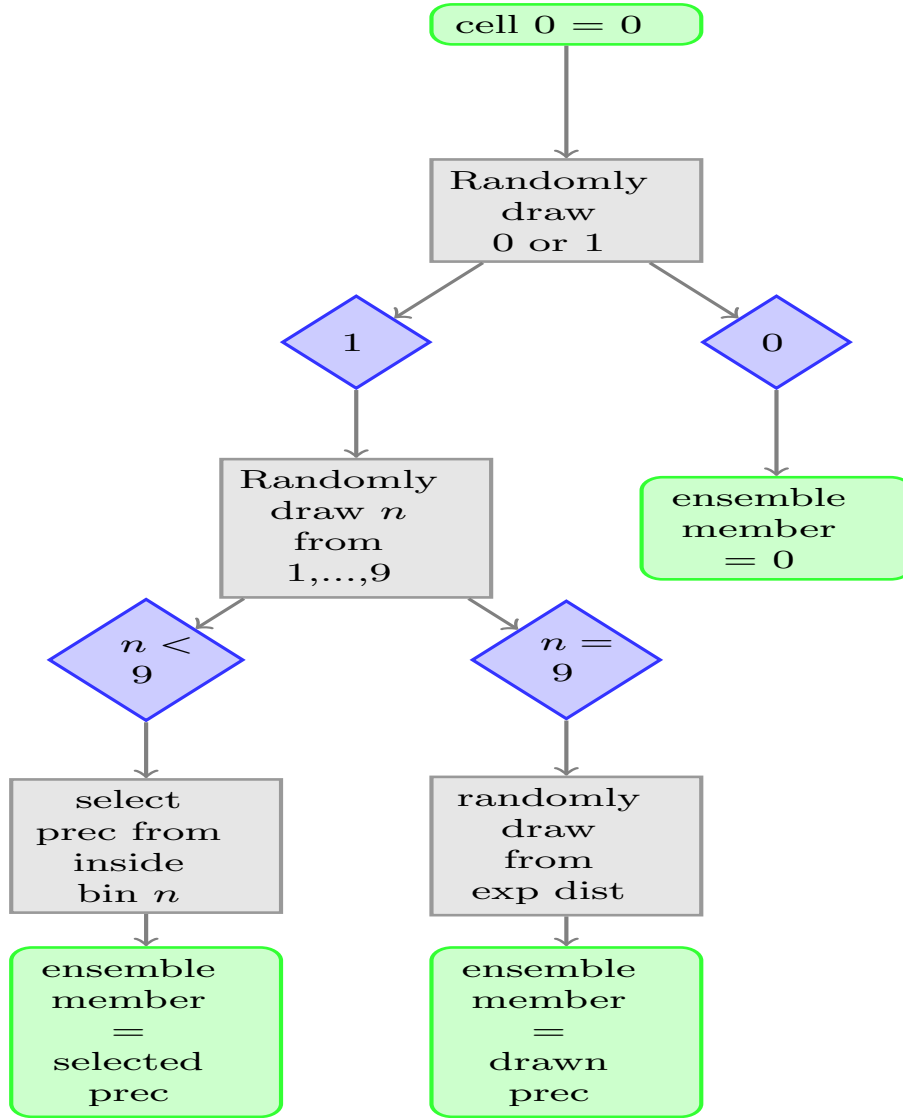
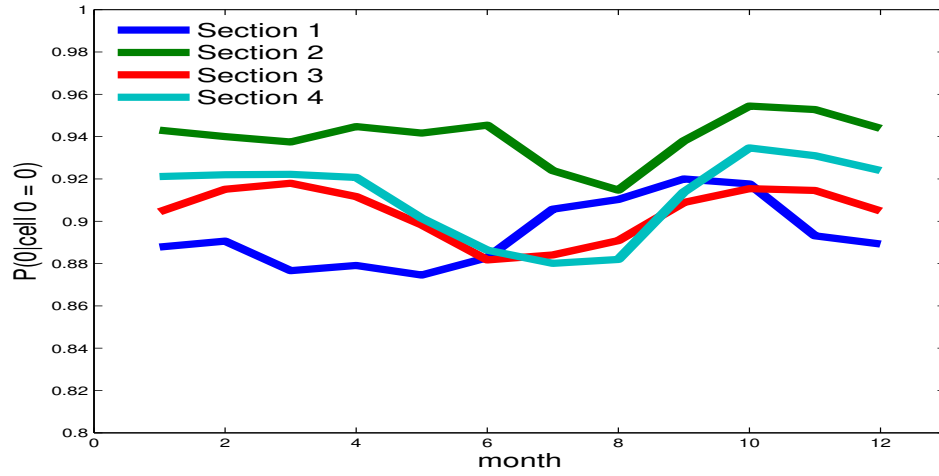
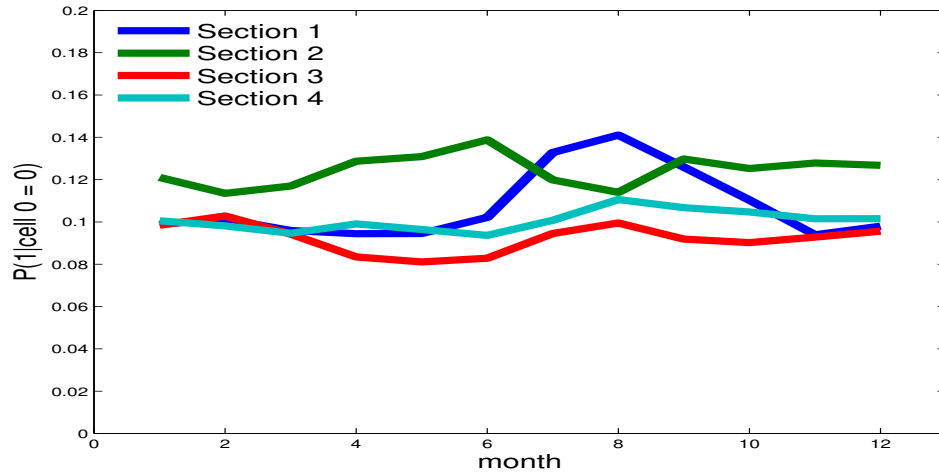


Figure 5.6: Setting precipitation for an ensemble member when cell 0 is dry. Either 0 or 1 is randomly drawn with probabilities conditional on cell 0 being dry. If 0 is selected the ensemble member has no precipitation. If 1 is selected precipitation needs to be calculated for the ensemble member. Historical precipitation is divided into 9 equally likely bins and a number between 1 and 9 is drawn. If the number is less than 9, precipitation is drawn from within that bin number. If the number is 9 precipitation is randomly drawn from the exponential distribution. If cell 0 is wet different probabilities need to be calculated.



(a) $P(\text{border cells} = \text{dry} | \text{cell } 0 = \text{dry})$



(b) $P(\text{border cells} = \text{wet} | \text{cell } 0 = \text{dry})$

Figure 5.7: The probability of a) a dry day and b) a wet day, given that cell 0 is dry, by area and month. Note the probability for a) and b) are on different scales. The probability of a dry day, given that cell 0 is dry, is high across all the months and areas.

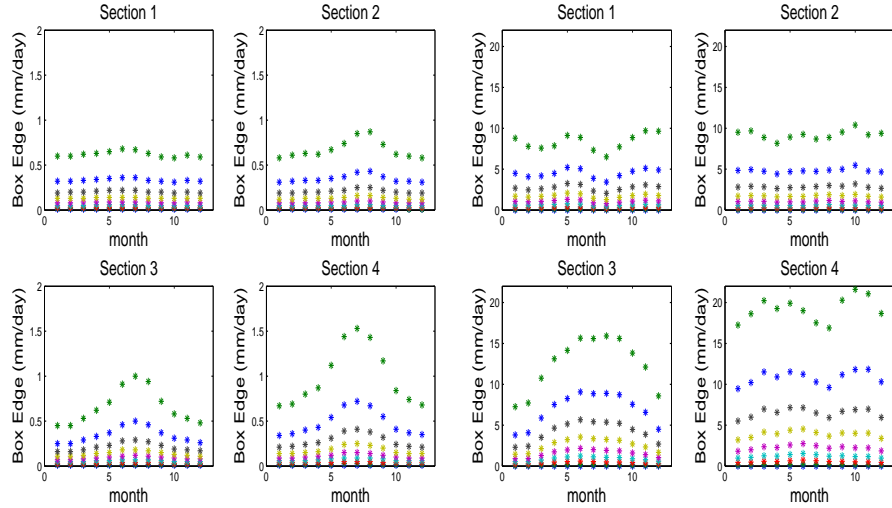
Figure 5.8c. When cell 0 is dry not much precipitation falls in the border bins, so all eight bins are approximately zero across all areas. This is not the case for the 9th bin, which is much wider with a bigger range of daily volumes. The precipitation bins, from the set when cell 0 is wet, are much wider as seen in Figure 5.8b. Area 3 shows seasonality in the precipitation volumes with the largest volumes of daily precipitation occurring between June and September. Area 4 has more daily precipitation than the other areas.

A precipitation value for each ensemble member is chosen by randomly drawing a number between 1 and 9, each number has an equally likely chance of being selected. The number drawn determines the bin the precipitation is taken from. The randomly selected bin is divided into 20 subsections and a number between 1 and 20 (with equal probability of being selected) is drawn. The precipitation for the ensemble member is set as the value in the middle of the selected subsection.[†] If the number drawn is 9, however, the precipitation is a random draw from the exponential distribution.

5.3.3 Estimating the exponential distribution

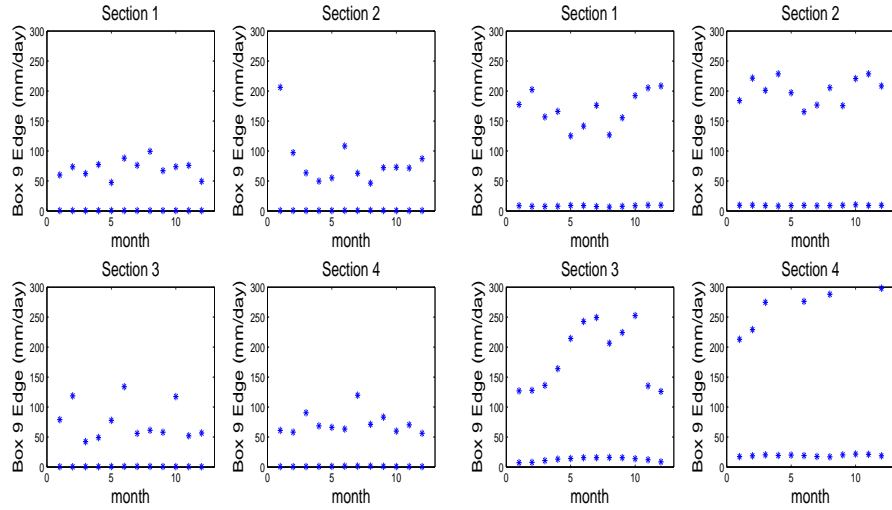
Data in the 9th precipitation bin, when cell 0 is wet or dry, is used to separately fit the exponential distribution. Figure 5.9a shows precipitation observations from bin 9 when cell 0 is wet (for area 1, for the month of January). The majority of the precipitation observations are in the left hand edge of bin 9 with very few data points on the right hand side of the bin. If bin 9

[†]An exception to this is if one of the subsection contains only identical numbers, in this case the precipitation is set to be this number.



(a) Equal bins when cell 0 is dry

(b) Equal bins when cell 0 is wet



(c) 9th bin when cell 0 is 0

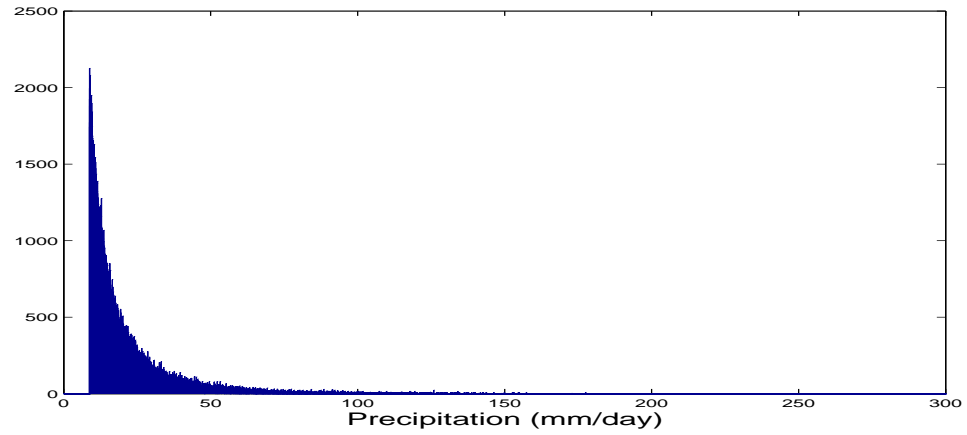
(d) 9th bin when cell 0 is wet

Figure 5.8: Equally likely bins for when cell 0 is dry (left) and when cell 0 is wet (right). Figures are (a) and (b) location of the edges of the first 8 bins and (c) and (d) the width of the last bin. Note that the box edge axis in (a) and (b) is different. When cell 0 is dry the first eight bins are very small as there is less rain in the surrounding cells.

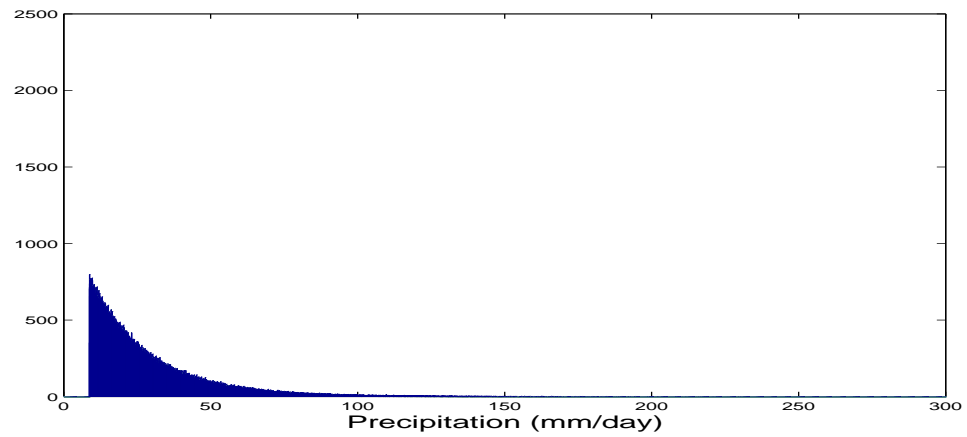
is divided into 20 equally spaced subsections to match the other 8 bins, the random precipitation selected would be equally spaced across the entire box so too many large values would be selected and too few small values. Instead, an exponential distribution is fitted to all the precipitation observations that fell into bin 9 (by area and month). The exponential distribution is truncated at the edge of box 8. Figure 5.9 compares the observations with randomly drawn numbers from the exponential distribution for area 1 in January. The exponential distribution tends to overestimate the precipitation as there are fewer lower precipitation values than in the observations.

The value assigned to the ensemble member when a 9 is drawn is selected by randomly drawing from the exponential distribution. The ensemble of initial conditions are compared with the gridded precipitation, the differences are shown in Figures 5.10 and E.3. The differences are centred around 0, although a few of the ensemble members differ by as much as 40 mm/day from the original cell 0.

Future work will use these ensembles of gridded data in the CERES-Maize model. It will be interesting to measure the uncertainty in the ensemble forecast and also to compare the skill between a nine member ensemble forecast and the current singleton ensemble forecast. Once the results from this have been considered it would be interesting to apply this to more areas than just Iowa.



(a) Precipitation Observations



(b) Precipitation ensemble members

Figure 5.9: Comparisons of observations in bin 9 for area 1 in January (5.9a) with randomly drawn numbers from the exponential distribution (5.9b). The lowest precipitation values randomly drawn from the truncated exponential distribution are less than the observations.

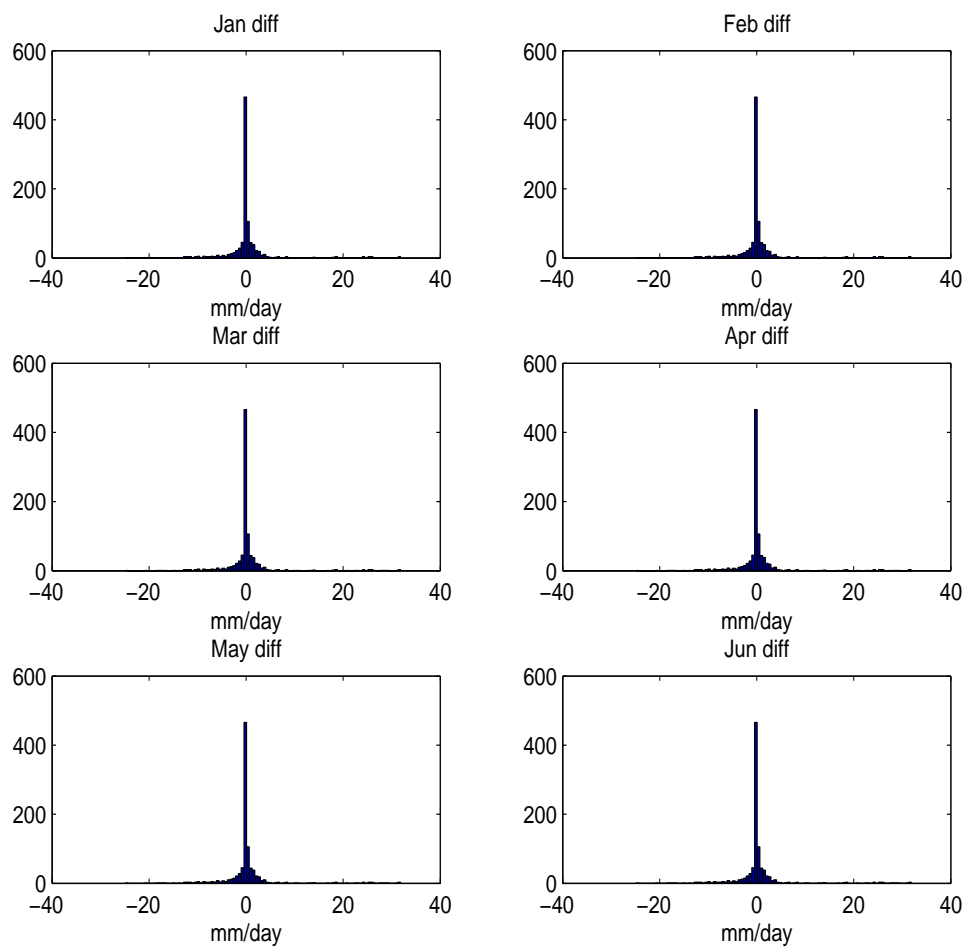


Figure 5.10: Jan to Jun: For Iowa the difference between the cells and the randomly drawn precipitation by month. The differences are all centred around 0.

5.4 Conclusions

1. A methodology to make an ensemble of gridded data for t_{min} and t_{max} for use by the CERES-Maize model is explained in Sections 5.1 and 5.2. The ensemble members are selected to reflect the sampling uncertainty in the gridded data by considering differences between cell 0 and the border cells. Perturbations are added to the initial gridded data using a Gaussian distribution.
2. A methodology to make an ensemble of gridded precipitation is discussed in Section 5.3. The method for adding ensemble members is explained in the flow chart from Figure 5.6.

Chapter 6

Summary

In this thesis we consider both the strengths and the main causes of errors in seasonal weather forecasting and crop modelling. A contribution from this thesis in Chapter 2 is the exploration of the effect model imperfection has on probabilistic forecasts for a dynamical system. Multi-model ensemble forecasts are made using a simple chaotic system as a proxy for seasonal weather. The simple one dimensional chaotic system modelled is the Moran Ricker Map. In this idealised world using models of the Moran Ricker Map, the only uncertainty is from model inadequacy as the observations are noise free. With three models of the system original experiments are set up to explore the limitations of small forecast-outcome archives. In Section 2.3 we demonstrate that small forecast-outcome archives, the size of DEMETER, can often overestimate the skill of the models. We also demonstrate why bootstrap resampling is a better way to estimate confidence intervals for

small samples. Additionally in Figure 2.10 we show that the order of skill for the models is consistent across different size archives.

Equally weighting forecasts from different seasonal weather models has been shown to improve the skill [26]. In Section 2.5 we demonstrate that this was not the case in our example with a large archive from the idealised world. Instead the forecast with the most skill in Table 2.8 uses two out of the three models. The multi-model ensemble forecast using all three models was in 4th place, behind single models.

We then explore new methods to improve the skill of a multi-model forecast by using climatology. One method which improves the skill is to include climatology as a separate model in an multi-model forecast, shown in Figure 2.15. Another method is blending individual model's forecasts with climatology, a technique known to improve the skill of seasonal weather forecasts [85]. To blend with climatology, the blending parameter is set in proportion to the skill of the model. In Section 2.6 we demonstrate for the first time in Table 2.10 that blending our three models of the Moran Ricker Map with climatology also improves their individual skill. In a further step these blended models are then equally weighted together which, in our example, provides forecasts with the most skill in Figure 2.16, an original contribution from this thesis. Further work beyond the scope of this thesis would be to weight the multi-model forecasts according to the skill of each individual model, as well as examining the skill of each model by lead time and location on the attractor [34].

In Chapter 3 crop modelling with the physical simulation model CERES-Maize [19,37] is examined using US maize yield. A probabilistic forecast is created from a singleton ensemble from the crop model by standard kernel dressing with Gaussian kernels in Section 3.1, this is a new contribution from this thesis. Also new in Section 3.2 is measuring the skill of the CERES-Maize model against a bench mark model. In crop modelling the bench mark model used is not climatology, which is not a strong enough test, but instead persistence. A multi-model forecast for crops is generated by creating empirical crop models based on US maize yield. The empirical crop models are the dynamic climatology model [85], not a new contribution, the ratio model, the asymmetric model and the gamma model which are new contributions, these are explained in Section 3.4. The asymmetric model uses a mixed Gaussian kernel to dress the estimates. The gamma model uses an asymmetric gamma kernel in kernel dressing. As far as we are aware using an asymmetric kernel for crop modelling is new. The skill of crop model forecasts, shown in Figure 3.19, improve if their kernel widths are set by minimising Ignorance, this is another contribution.

The crop model with the most skill for US maize yield forecasts, relative to persistence, is the CERES-Maize. By design empirical models are unable to predict large falls in yield from adverse weather events so they have less skill in these years than the physical simulation model CERES-Maize. The ratio model is the empirical model with the most skill. To examine if equally weighting crop models improves the skill [26], two crop models are equally weighted together. The CERES-Maize with the asymmetric model and the

CERES-Maize model with the ratio model both have more skill than the CERES-Maize model as shown in Figure 3.23. This is a surprising result as the asymmetric model has the least skill of all the crop models relative to persistence.

These experiments are repeated at state level for the top 8 maize producing states. The skill of the CERES-Maize relative to persistence at state level varied between states. It ranges between -1.13 for Illinois to -0.09 for South Dakota, where there is little skill above persistence. For each state the model or multi-model with the most skill relative to persistence is found. For Illinois it is the CERES-Maize model and for South Dakota it is the ratio model. For all but one of the remaining states it is from equally weighting CERES-Maize and ratio model, the exception is Iowa where it is from equally weighting CERES-Maize and asymmetric model. This demonstrates there is a role for simple empirical models to improve the skill of the CERES-Maize model.

The meteorological observations and maize yield are examined at county level for the state of Iowa, although the data itself is nothing new the presentation of it is. Also new is the consideration of the size of Iowa yield compared against the highest annual t_{max} and precipitation in Figure 4.12, which illustrates that it is not just drought conditions that cause lower than average maize yield. By comparing the residual yield against the number of days greater than 29°C it seems there is a negative impact on yield, also a new contribution. How to identify the date that the technical advancements start and how to strip the technical advancements from the yield time series

are also considered.

The CERES-Maize model uses one approximation of gridded weather to predict yield. How much the uncertainty in this gridded data impacts the yield forecasts is unknown. In Chapter 5 a methodology for creating an ensemble of initial conditions for the gridded data is proposed, a new contribution from this thesis. Gridded t_{max} and t_{min} ensembles are added by randomly drawing from a Gaussian distribution. Precipitation for ensemble members is conditional on whether or not there is precipitation in the gridded observation, the method is outlined in the flowchart in Figure 5.6. Once these gridded initial conditions are input into the CERES-Maize model it would be very interesting to compare the skill of a singleton ensemble from the CERES-Maize model with the skill from a 9 member ensemble. Unfortunately the runs from the CERES-Maize model have not yet been completed.

Appendix A

Gridded Meteorological data sets

Further details about the gridded meteorological observations used by the CERES-Maize model are discussed here.

A.1 Gridded Precipitation

The gridded precipitation data set is from Climate Prediction Centre (CPC) Unified Rain Gauge Database [33]. They estimate the gridded data set directly from the observations, as they have a large network of approximately 8,000 rain gauge meters covering the USA. The irregularly spaced precipitation observations are converted into a regularly spaced grid by a data interpolation scheme based on Cressman [33]. This interpolation allocates weights to the observation dependent on how close it is to a grid point.

A.2 Gridded Temperature

Gridded t_{min} and t_{max} are provided from reanalysis data by The National Centres of Environmental Prediction (NCEP) North American Reanalysis (NARR) [53]. The NCEP NARR data set provides the best estimate of the North American weather state, this is a high resolution sub set of the global NCEP reanalysis [39]. To create this multiple historical meteorological observations are collected, checked for errors and then compared against the physics from the medium-range forecast model in operation in April 2003 [74]. The NARR supplies data on a 0.3° by 0.3° grid.

A.3 Gridded solar radiation

The solar radiation is from the Surface Radiation Budget (SRB) found in Global Energy and Water Cycle Experiment (GWEX) by NASA. The data is available from 1998 to 2007 and is on a grid size of 1° by 1° . The variable used is downward shortwave radiation. To derive this variable an algorithm is used with cloud parameters from the International Satellite Cloud and reanalysis data from Global Modelling and Assimilation (GMAO) [84].

Appendix B

Imperfect Models for the Moran Ricker Map

Three imperfect models were designed to replicate the dynamical system. In this Thesis the dynamical system was the Moran Ricker Map [56, 70] with α as 3:

$$x_{i+1} = x_i^{\alpha(1-x_i)} \tag{B.1}$$

Iterations through this system produced values bounded between 0 and 2.46 (x_{max}), where $x_{max} = \frac{e^2}{3}$ as shown in Figure B.1. There were two fixed points at $x_i = 0$ and $x_i = 1$.

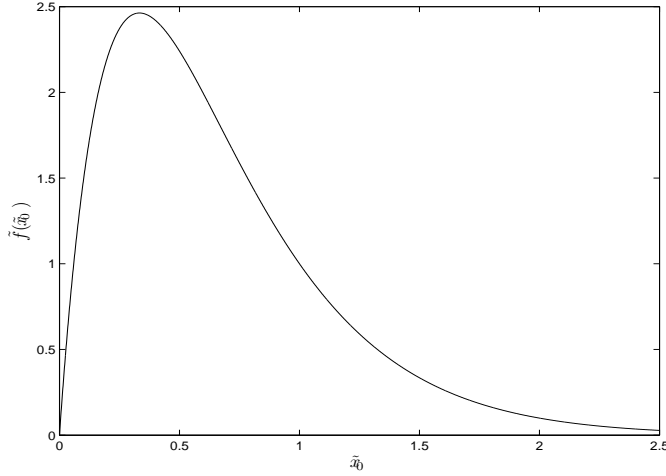


Figure B.1: The Moran Ricker Map when α is 3 is bounded between 0 and 2.46.

B.1 Model MR12: Taylor series expansion of the Moran Ricker Map

One imperfect model to replicate the Moran Ricker Map was built using Taylor's Series expansion for $e^{3(1-x_i)}$. Substituting this into Equation B.1 gives :

$$x_{i+1} = x_i \left(1 + 3(1 - x_i) + \frac{1}{2!}(3(1 - x_i))^2 + \cdots + \frac{1}{n!}(3(1 - x_i))^n \right) \quad (\text{B.2})$$

Equation B.2 is expanded to the 4th power (MR4) and compared to the Moran Ricker Map by setting x_0 as evenly spaced values between 0 and 2.5 and then iterating these values once through the model. Figure B.2a shows the values of x_1 are not bounded between 0 and x_{max} . In particular when x_0 is greater than 2, x_1 tends towards infinity. For each iteration through

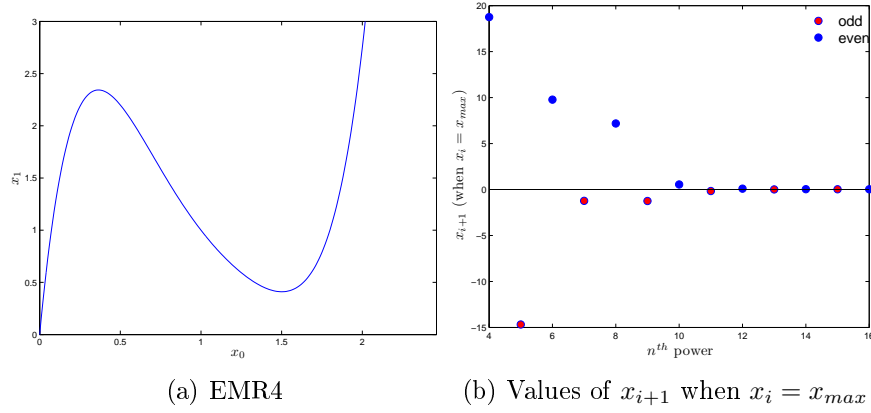


Figure B.2: Figure B.2a: MR4 tends to infinity for large values of x_0 . Figure B.2b: when the n^{th} expansion is odd x_{i+1} is negative.

model MR4 more points would move towards infinity.

For the values of x_{i+1} to lie between 0 and 2.46, to replicate the system, x_{i+1} when $x_i = x_{max}$ must be less than or equal to x_{max} . Figure B.2b shows the different values of x_{i+1} (when $x_i = x_{max}$) for Equation B.2 expanded to the n^{th} power. When the n^{th} power is even x_{i+1} is positive and when the n^{th} power is odd x_{i+1} is negative.

If x_{i+1} of x_{max} is negative any point iterated through the model will quickly move towards minus infinity. A Taylor Series expansion using an even power provides the simplest model to approximate the Moran Ricker Map. The cut off point where x_{i+1} (for $x_i = x_{max}$) is less than the x_{max} is shown in Figure B.2b as model MR10. The values of x_{i+1} are 0.5 for MR10 and 0.1 for MR12.

Model MR10 and MR12 both iterate x_{max} onto a number below x_{max} , so x_{i+1} will remain bounded for iterations through the system. Comparisons between the system and models MR10 and MR12 at x_1 are in Figure B.3.

B.1. Model MR12: Taylor series expansion of the Moran Ricker Map

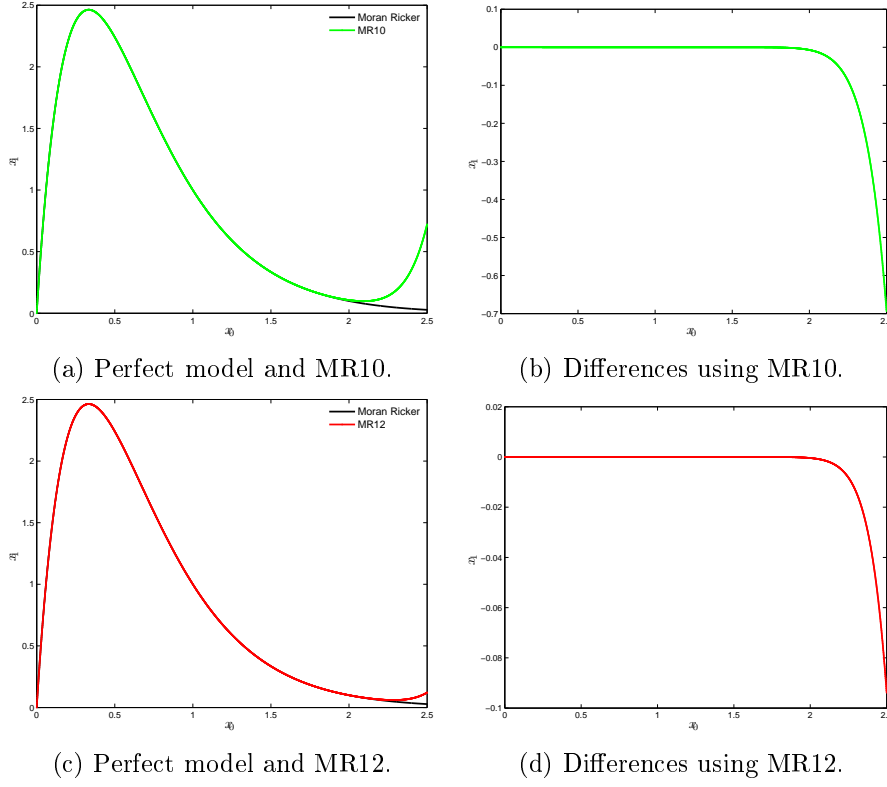


Figure B.3: Comparison between the Moran Ricker Map and the models MR10 and MR12. Model MR12 provides more accurate results than model MR10 for higher values of x_0 .

The Figure shows model MR12 provides a more accurate estimate for higher values of x_0 than model MR10.

If the initial conditions (x_0) lie between 0 and 2.675 for model MR10, and between 0 and 2.925 for model MR12 x_{i+1} remains bounded. Outside these values x_{i+1} quickly approaches infinity. The ensemble of initial conditions was selected so that they never fell outside these boundaries.

B.2 MRLM: Taylor series expansion of the log term

Another imperfect model of the system was created by taking the log of the Moran Ricker Map, and then making a Taylor series expansion for this log term:

$$\log(xe^{3(1-x)}) = \log(x) + \log(e^{3(1-x)}) \quad (\text{B.3})$$

$$\log(xe^{3(1-x)}) = \log(x) + 3 - 3x \quad (\text{B.4})$$

$$\log(xe^{3(1-x)}) = \log(x) + 3 - 3(e^{\log(x)}) \quad (\text{B.5})$$

The Taylor expansion for the log of the Moran Ricker Map is

$$\log(x_{i+1}) = \log(x_i) + 3 - 3 \left(1 + \log(x_i) + \frac{(\log(x_i))^2}{2!} - \frac{(\log(x_i))^3}{3!} + \dots \right) \quad (\text{B.6})$$

Equation B.6 is expanded to the 4th power to create MRLM4. For model MRLM4 when $x_i = 0$, x_{i+1} goes to infinity whereas for the system $x_i = 0$ is a fixed point. To account for this in the log model when $x_i = 0$, x_{i+1} was set to 0 too. Figure B.4 shows the results of one iteration through model MRLM4 using values of x_0 evenly spaced between 0 and 2.5. This graph clearly shows that the values of x_1 have a similar range to the system as they are bounded between 0 and a slightly lower maximum value of 2.40 however when x_0 is close to 0, the values differs for x_1 moving to 0.

If x_i , a set of points evenly spaced between 0 and x_{max} , is iterated through

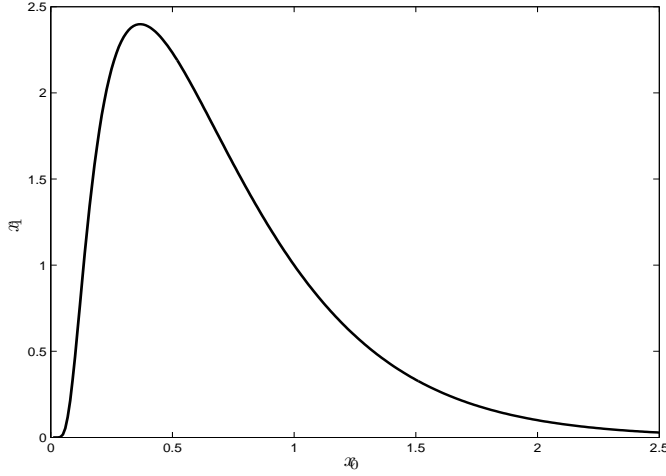


Figure B.4: MRLM4 at x_1 . Note this is close to the system except when x_0 is close to 0.

the system the only point in x_{i+1} that will ever reach 0 is the point in x_i equal to 0. For the model MRLM4 this is not the case, as shown in Figure B.4, if x_i is close to zero x_{i+1} will tend to zero especially as the number of iterations increase. To see the effect this has at x_{10} sorted values from different log models and the system are compared in Figure B.5. For model MRLM4 a large number of x_{10} are zero compared to the system. For MRLM6 there are less x_{10} at zero but overall the x_{10} values are lower than the system. For example in the Moran Ricker Map there are nearly 3000 points with values less than 0.5 but for MRLM6 there are about 4000 points. For model MRLM8 the x_{10} values are still lower but it is closer to the Moran Ricker Map. The log model used in this Thesis is MRLM8.

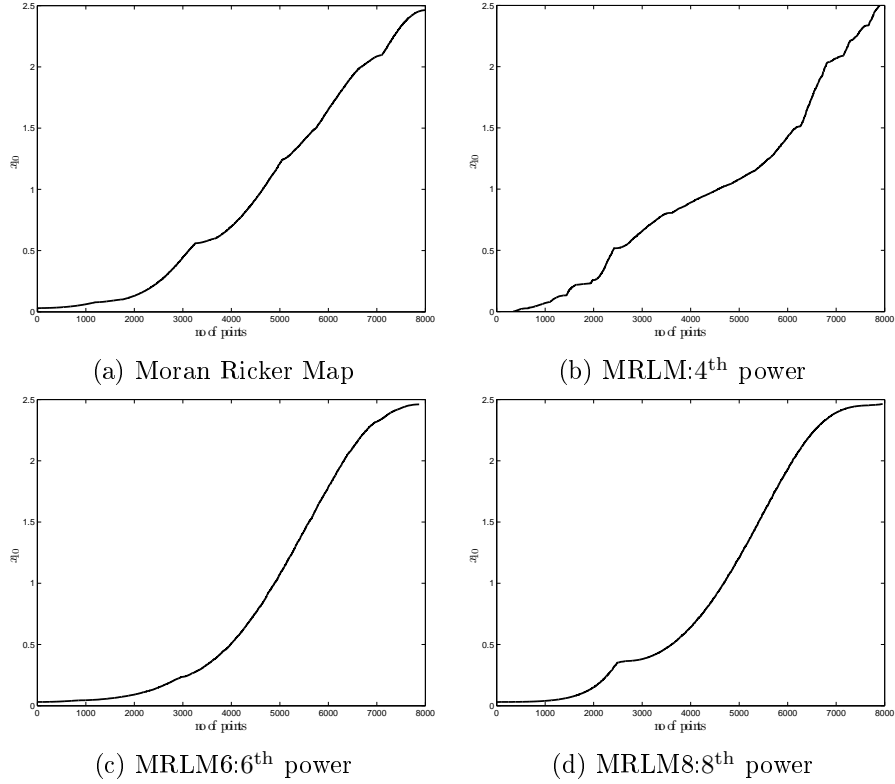


Figure B.5: Comparison of x_{10} values of Moran Ricker Map B.5a with log models to the 4th, 6th and 8th power. MRLM4 has more points at 0 than the system and MRLM8 has more points at the maximum value than the system.

B.3 Model MRFT: A Fourier transform approx of the Moran Ricker Map

The third imperfect model to replicate the Moran Ricker Map was created using the Fourier transform:

$$y = \frac{a_0}{\sqrt{2}} + \sum_{n=1}^{\infty} a_n \cos \left(2\pi n \frac{(x-a)}{(b-a)} \right) + b_n \sin \left(2\pi n \frac{(x-a)}{(b-a)} \right) \quad (\text{B.7})$$

B.3. Model MRFT: A Fourier transform approx of the Moran Ricker Map

As the Moran Ricker Map is bounded between 0 and 2.46 the constants can be calculated by numerically integrating the equation between 0 and 2.46. Equation B.7 becomes:

$$\begin{aligned} y = & \frac{2.8391}{\sqrt{2}} + 0.7551 \cos\left(\frac{1 * 2\pi * x}{2.46}\right) - 0.2872 \cos\left(\frac{2 * 2\pi * x}{2.46}\right) - \\ & 0.3414 \cos\left(\frac{3 * 2\pi * x}{2.46}\right) - 0.2642 \cos\left(\frac{4 * 2\pi * x}{2.46}\right) - \\ & 0.1960 \cos\left(\frac{5 * 2\pi * x}{2.46}\right) - 0.1476 \cos\left(\frac{6 * 2\pi * x}{2.46}\right) - \\ & 0.1139 \cos\left(\frac{7 * 2\pi * x}{2.46}\right) - 0.09 \cos\left(\frac{8 * 2\pi * x}{2.46}\right) - \\ & 0.0727 \cos\left(\frac{9 * 2\pi * x}{2.46}\right) - 0.0598 \cos\left(\frac{10 * 2\pi * x}{2.46}\right) + \\ & 1.8147 \sin\left(\frac{1 * 2\pi * x}{2.46}\right) + 0.9809 \sin\left(\frac{2 * 2\pi * x}{2.46}\right) + \\ & 0.4537 \sin\left(\frac{3 * 2\pi * x}{2.46}\right) + 0.2296 \sin\left(\frac{4 * 2\pi * x}{2.46}\right) + \\ & 0.1285 \sin\left(\frac{5 * 2\pi * x}{2.46}\right) + 0.0781 \sin\left(\frac{6 * 2\pi * x}{2.46}\right) + \\ & 0.0507 \sin\left(\frac{7 * 2\pi * x}{2.46}\right) + 0.0346 \sin\left(\frac{8 * 2\pi * x}{2.46}\right) + \\ & 0.0246 \sin\left(\frac{9 * 2\pi * x}{2.46}\right) + 0.0180 \sin\left(\frac{10 * 2\pi * x}{2.46}\right) \quad (\text{B.8}) \end{aligned}$$

Equation B.8 was expanded to the 4th term i.e. to the sine and cosine terms for $\frac{4 * 2\pi * x}{2.46}$, creating model MRFT4. One iteration through the model using x_0 is shown in Figure B.6. The Figure is a notably different shape from the system especially at the minimum and maximum values of x_1 . In particular the graph crosses the x_1 axis at a number greater than 0, it dips down to

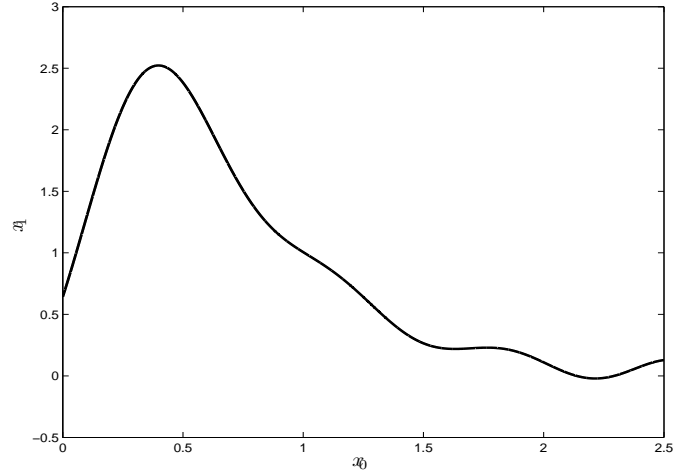
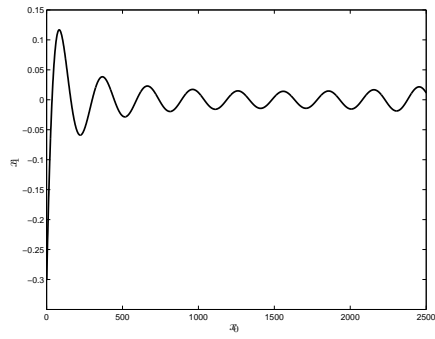


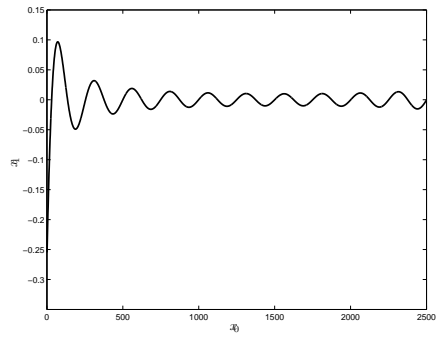
Figure B.6: MRFT4 at one iteration. This differs significantly from the system when x_0 is close to 0.

negative numbers between 2 and 2.5 and as x_0 moves closer to the maximum of 2.5 the values of x_1 start to increase. A further difference between the system and the Fourier Model as the model has no fixed points at $x = 0$ and $x = 1$.

To see which expansion of the Fourier transform to use comparisons were made at x_1 between the system and the Fourier transforms to the 10th and 12th expansion in Figure B.7. Model MRFT12 is a better fit than model MRFT10 for values close to 0. In this Thesis the Fourier transform model is to the 12th term.



(a) 10th term Fourier transform



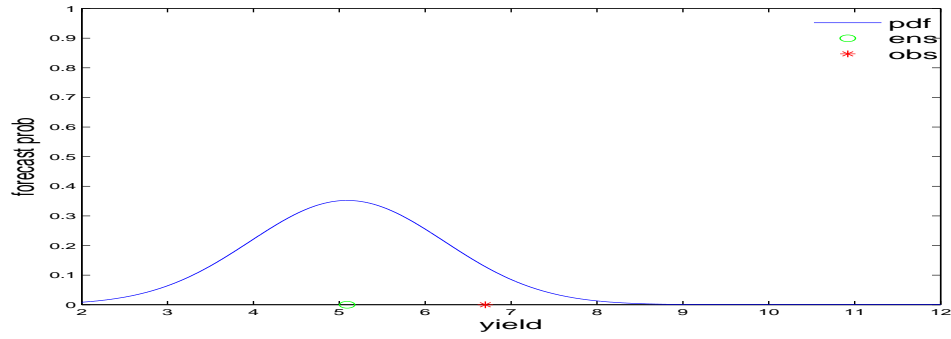
(b) 12th term Fourier transform

Figure B.7: Comparison of the system and (a) MRFT10 and (b) MRFT12 at x_1 . Notice how the differences close to $x_0 = 0$ decrease for MRFT12.

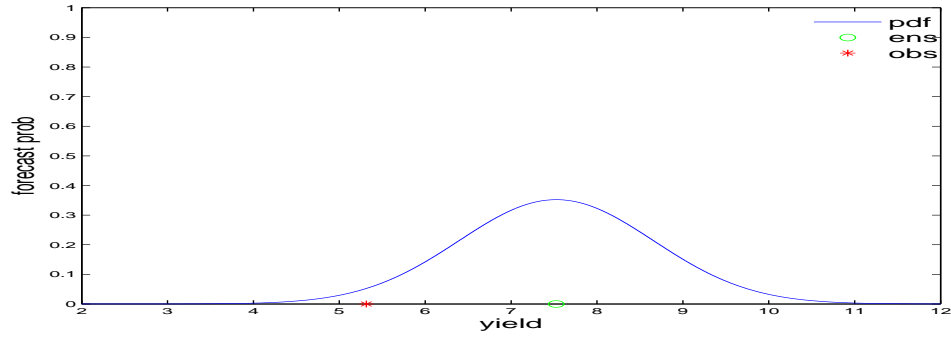
Appendix C

Figures for Crop Modelling

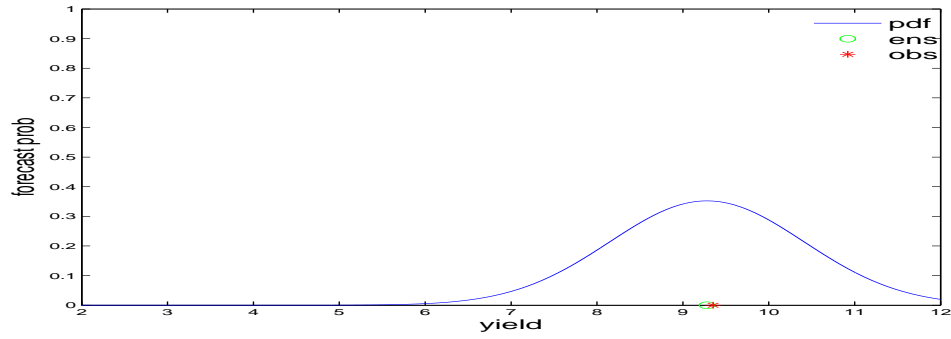
This appendix contains figures for the persistence, dynamic climatology and ratio models which are discussed in Chapter 3. Figures C.1, C.3 and C.5 show the forecast distributions from a kernel dressed ensemble for the years 1984, 1988 and 2006. The y -axis for Figure C.3 is on a different scale to the others as the forecast distribution is so flat. Figures C.2, C.4 and C.6 show how smooth the kernel width by ignorance is for the years 1984, 1988, 1993 and 2006.



(a) 1984



(b) 1988



(c) 2006

Figure C.1: The forecast distribution produced from dressing estimate with a Gaussian kernel of width σ . The forecast distribution from the persistence model is shown as a blue line, the estimates are shown as green circles and the outcomes are shown as red stars for selected years 1984, 1988 and 2006. Notice although the outcome does not always lie in the high probability areas of the forecast distribution for years 1984 and 1988 it is captured by the forecast distribution.

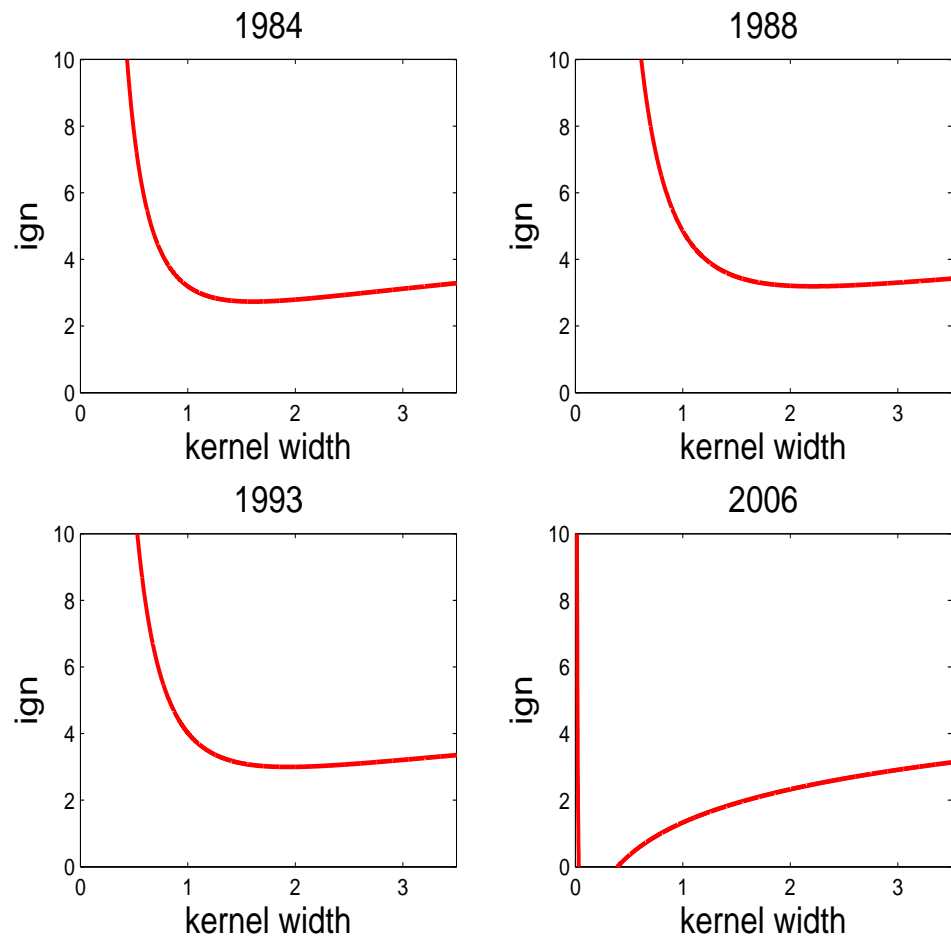
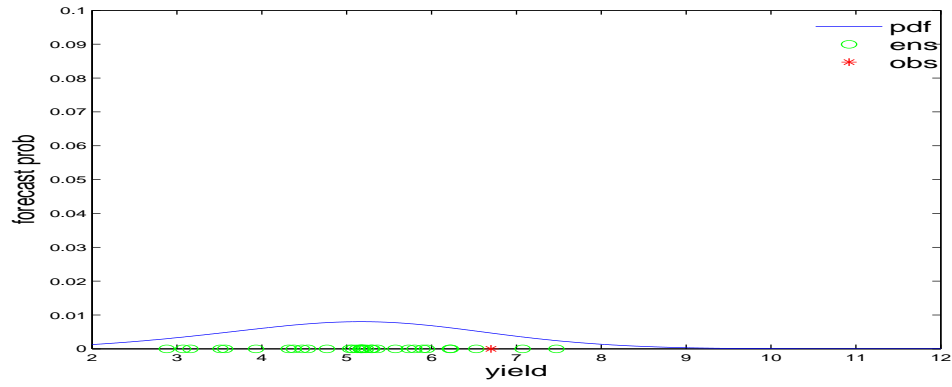
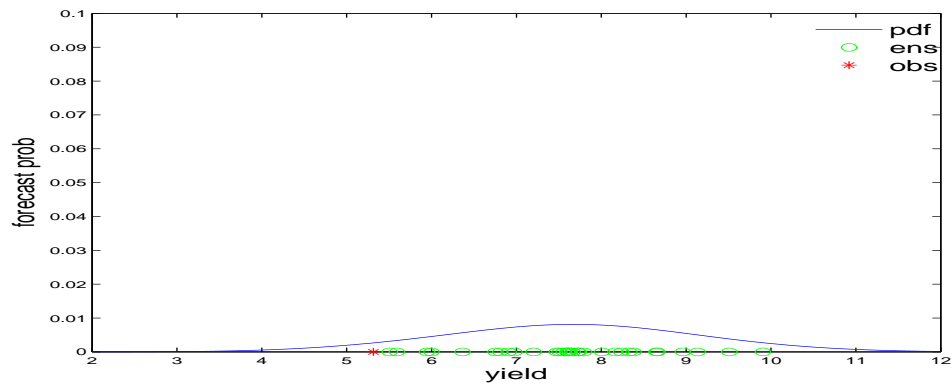


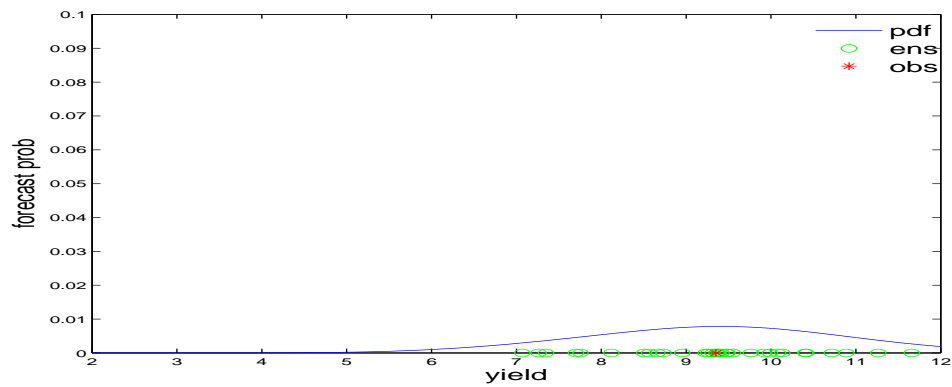
Figure C.2: The robustness of forecast distribution is checked by comparing the ignorance by kernel width for four selected years for the persistence model. In 2006 the outcome fell in a very high probability of the forecast distribution, so the narrower the kernel the better.



(a) 1984



(b) 1988



(c) 2006

Figure C.3: The forecast distribution for dynamic climatology. The forecast distribution from the dynamic climatology model is shown as a blue line, the ensemble of estimates are shown as green circles and the outcomes are shown as red stars for the years 1984, 1988 and 2006. The ensemble of estimates is so wide that the forecast distribution is very flat.

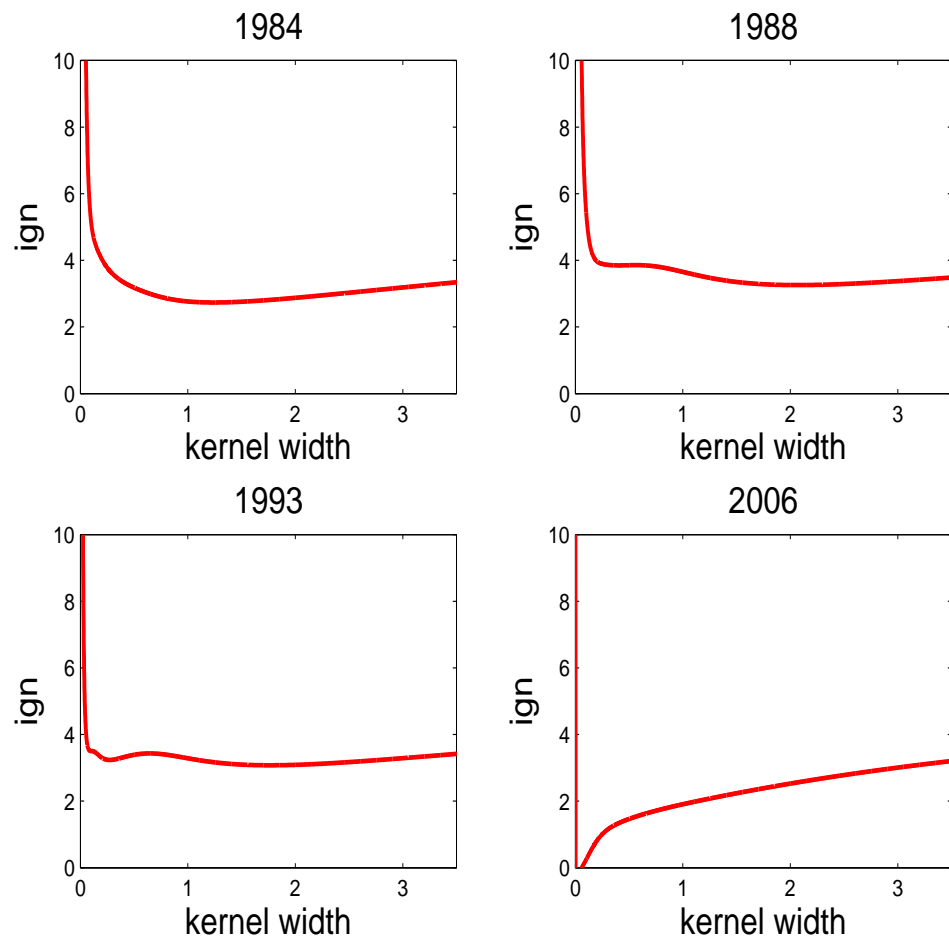
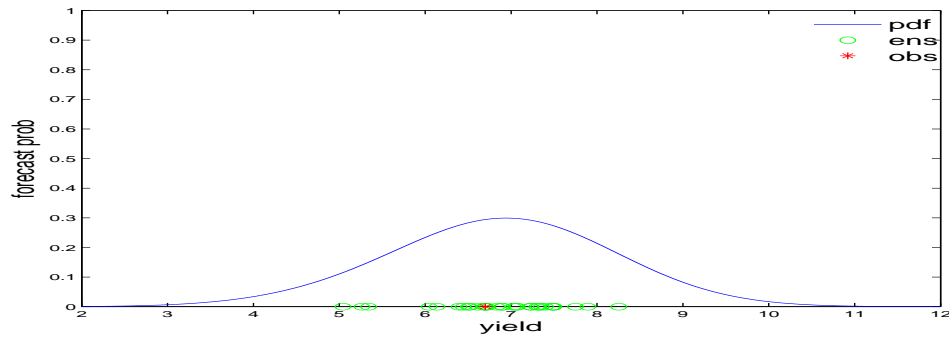
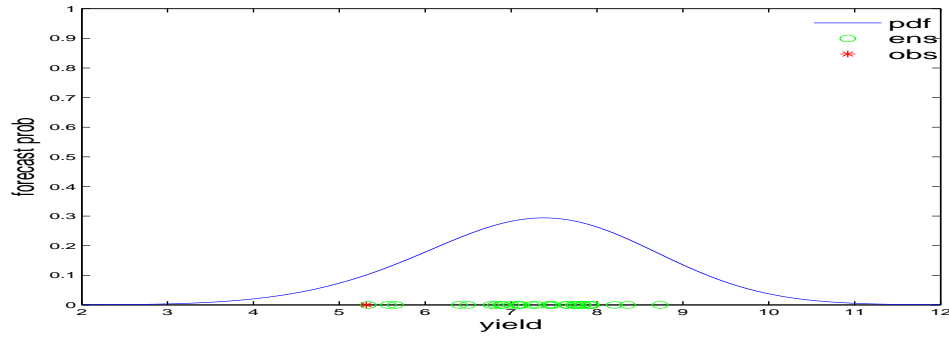


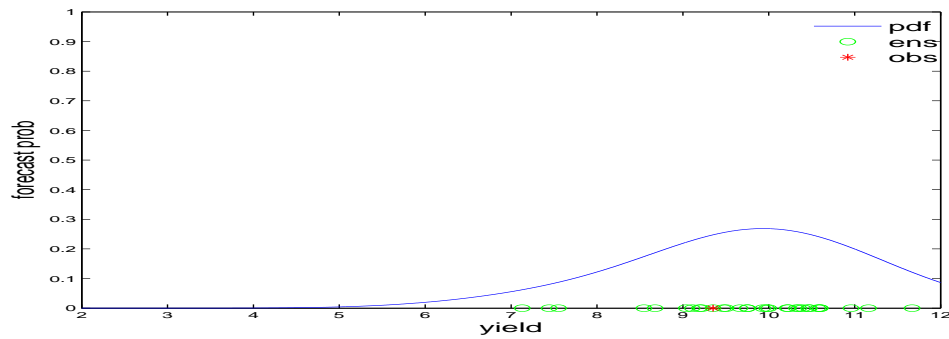
Figure C.4: Examining the ignorance by kernel width for selected years for the dynamic climatology model. The ignorance by kernel width for 1984, 1988 and 1993 have very similar shapes.



(a) 1984



(b) 1988



(c) 2006

Figure C.5: The forecast distribution produced from dressing the ensemble of estimates. The forecast distribution from the ratio model is shown as a blue line, the ensemble of estimates are shown as green circles and the outcomes are shown as red stars for the years 1984, 1988 and 2006. Notice how the outcome does not lie in the high probability area of the forecast distribution for 1984.

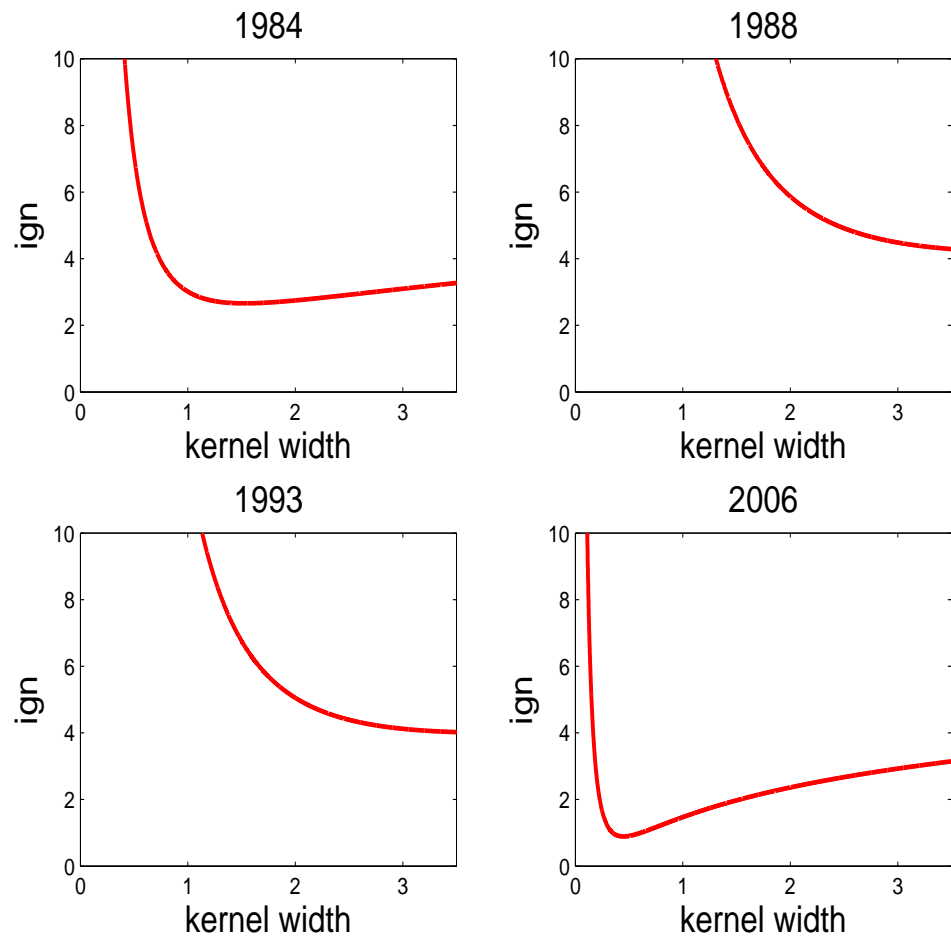


Figure C.6: Examining the robustness of the kernel width by looking at the ignorance by kernel width for four selected years for the persistence model. The kernel width needs to be much wider for years 1988 and 1993

Appendix D

Summary of state codes for weather stations

State Code	State Number	State
01	AL	Alabama
02	AZ	Arizona
03	AR	Arkansas
04	CA	California
05	CO	Colorado
06	CT	Connecticut
07	DE	Delaware
08	FL	Florida
09	GA	Georgia
10	ID	Idaho
11	IL	Idaho

Appendix D. Summary of state codes for weather stations

State Code	State Number	State
12	IN	Indiana
13	IA	Iowa
14	KS	Kansas
15	KY	Kentucky
16	LA	Louisiana
17	ME	Maine
18	MD	Maryland
19	MA	Massachusetts
20	MI	Michigan
21	MN	Minnesota
22	MS	Mississippi
23	MO	Missouri
24	MT	Montana
25	NE	Nebraska
26	NV	Nevada
27	NH	New Hampshire
28	NJ	New Jersey
29	NM	New Mexico
30	NY	New York
31	NC	North Carolina
32	ND	North Dakota
33	OH	Ohio
34	OK	Oklahoma

State Code	State Number	State
35	OR	Oregon
36	PA	Pennsylvania
37	RI	Rhode Island
38	SC	South Carolina
39	SD	South Dakota
40	TN	Tennessee
41	TX	Texas
42	UT	Utah
43	VT	Vermont
44	VA	Virginia
45	WA	Washington
46	WV	West Virginia
47	WI	Wisconsin
48	WY	Wyoming

D.0.1 Minimum temperature observations

Monthly \bar{t}_{min} for Iowa is shown in Figure D.1. Like t_{max} , t_{min} displays seasonality with on average higher temperatures between June and August and lower temperatures in January and December. The cooler months of January, February and December have the most variability. The monthly \bar{t}_{min} for the individual month of July is shown in Figure D.3 by weather

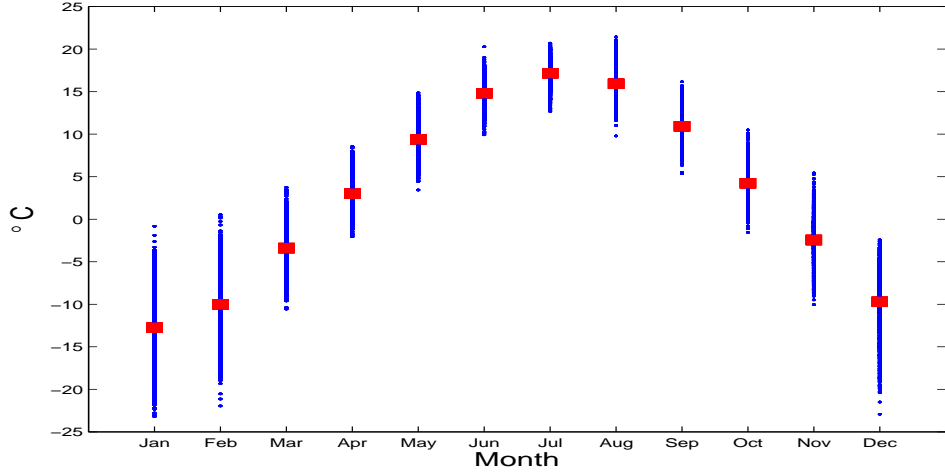


Figure D.1: Monthly \bar{t}_{min} in Iowa. Note that on average the higher temperatures are between June and August.

station and year. The weather station which tends to have lower \bar{t}_{min} across most of the years is again ○ (weather station 132684 from Fayette).

The monthly anomalies for t_{min} and t_{max} for Iowa are shown in Figure D.2. In this figure, zero is the annual mean temperature calculated from the overall t_{max} and t_{min} values. A green dot represents the anomaly between \bar{t}_{max} and the annual mean temperature for a year and a weather station and the red square is the monthly mean of these anomalies. A blue dot represents the anomaly between \bar{t}_{min} and the annual mean temperature. In the cooler months the anomalies from \bar{t}_{max} (green dots) and \bar{t}_{min} (blue dots) sometimes have similar values. This is not the case in the warmer months when \bar{t}_{max} and \bar{t}_{min} anomalies do not overlap.

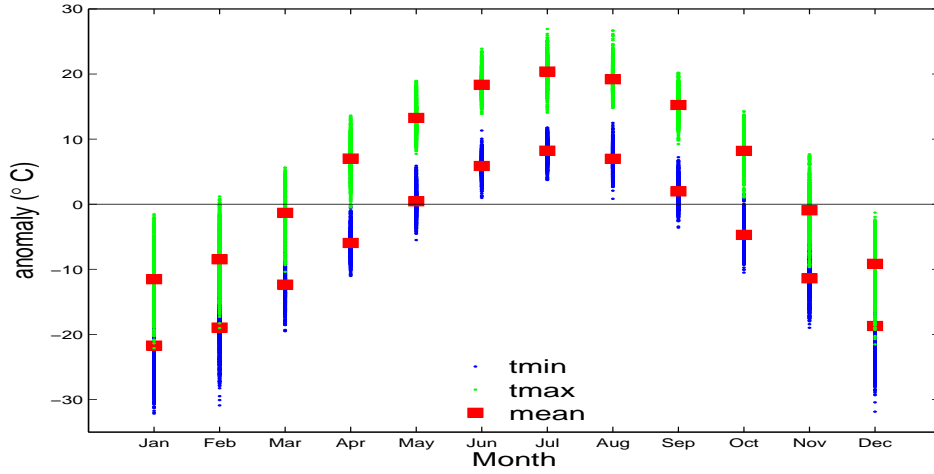


Figure D.2: Monthly anomalies for t_{min} (blue) and t_{max} (green) in Iowa.

D.0.2 Snowfall observations

Typically snow in Iowa falls between October and April, with the majority of snow falling in January and December. If snow falls in April or May it can delay the planting of maize. Looking at snowfall by month in Figure D.4 it can be seen there was significant snowfall in April (light blue) for the years 1973, 1982, 1983 and 1997.

D.1 Flags for Meteorological Observations

Further error flags contained in the USHCN database:

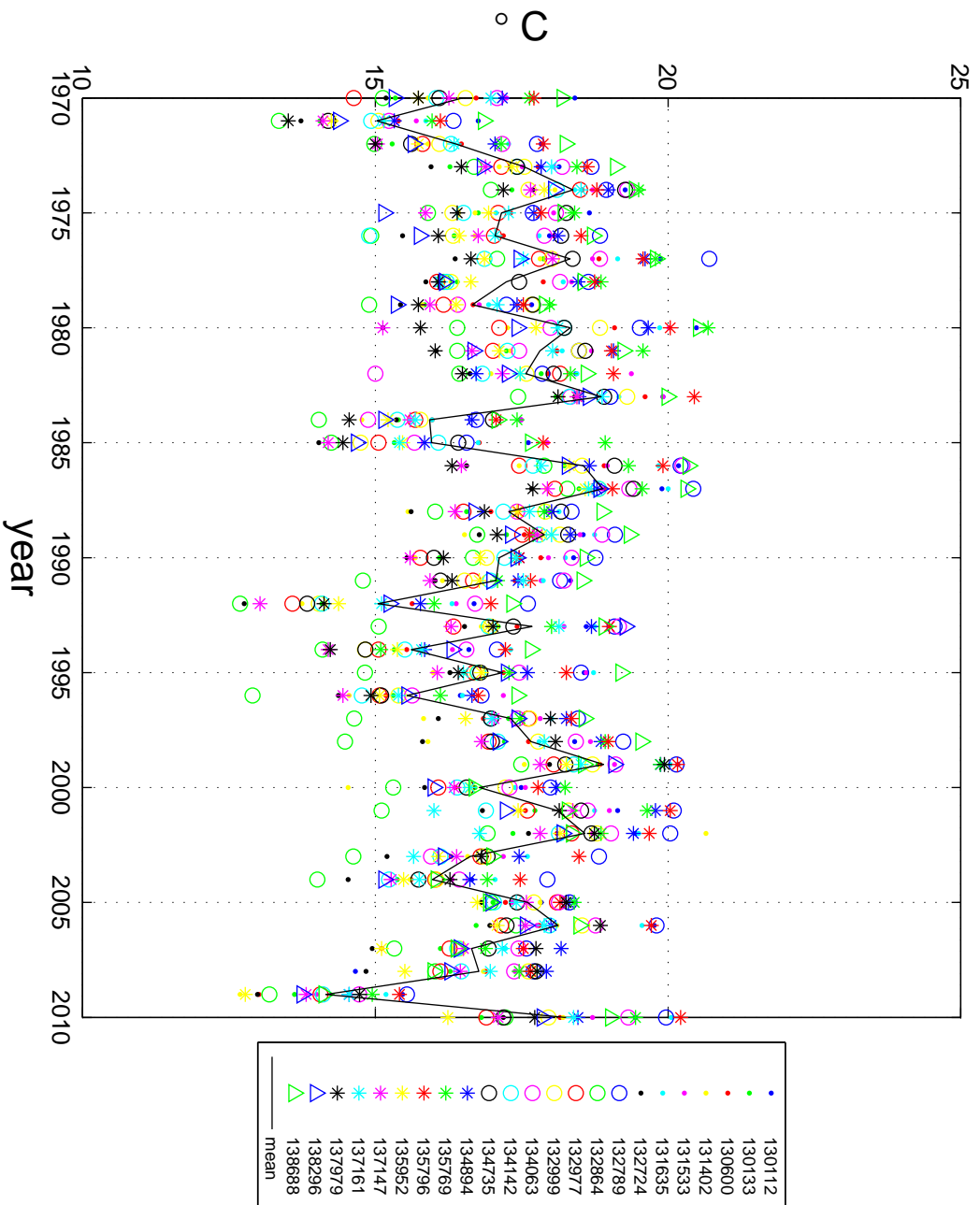


Figure D.3: Monthly \bar{t}_{min} for July. Note that 2009 was colder than usual in July for most of the weather stations across Iowa.

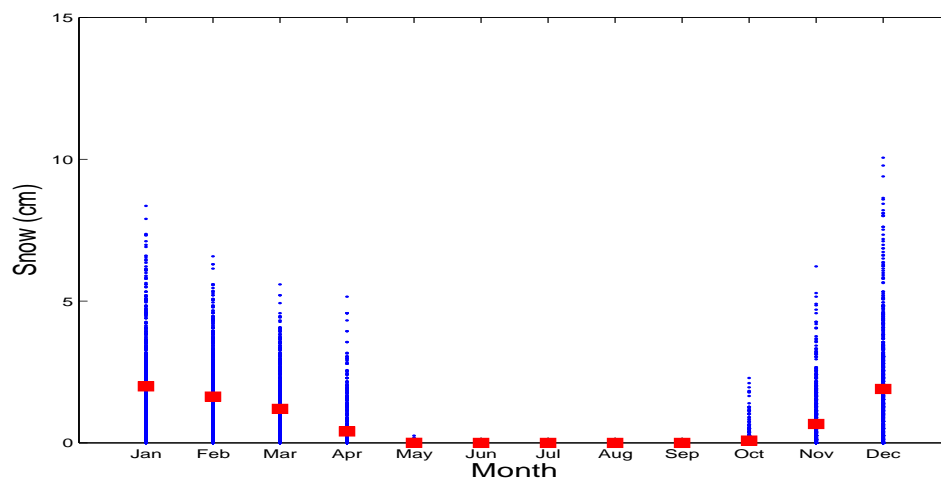


Figure D.4: Monthly snow in Iowa by year. On average more snow falls in the months of December and January.

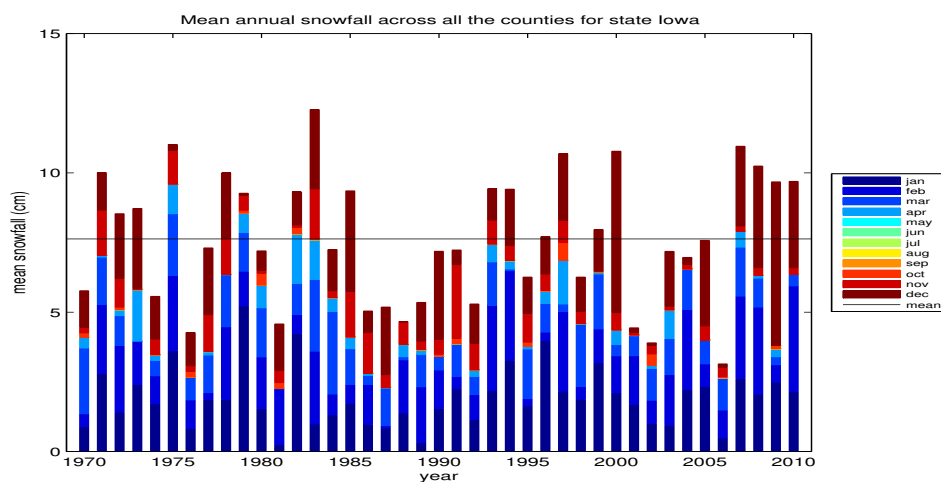


Figure D.5: Annual snowfall by year and month in Iowa. There is quite a lot of snow in April (light blue) for the years 1973, 1982, 1983 and 1997.

Table D.2: Summary of M Flags for Temperature Data

Flag Type	Code	Explanation
m	blank	no measurement information applicable
m	L	temperature appears to be lagged with respect to reported hour of observation

Table D.3: Summary of S Flags for Temperature Data

Flag Type	Code	Explanation
s	blank	no source (missing data)
s	0	data source U.S. Cooperative Summary of the Day (NCDC DSI-3200)
s	H	data source High Plains Regional Climate Center real-time data

Appendix E

Figures for creating initial conditions for gridded weather

This appendix contains figures for t_{max} and precipitation. Figures E.1 and E.2 show estimated monthly μ and σ for the Gaussian distribution that the t_{max} ensemble members are drawn from. The distributions are very similar to the t_{min} distributions. Figure E.3 compares the precipitation ensemble members against the gridded precipitation for the months of July to December. Most of the differences are close to 0.

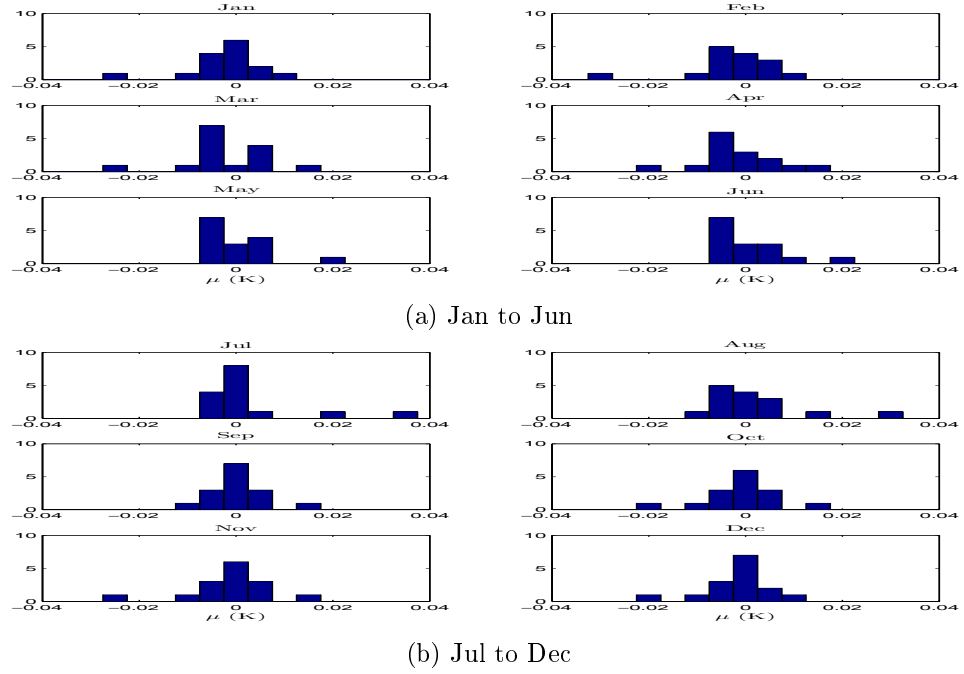
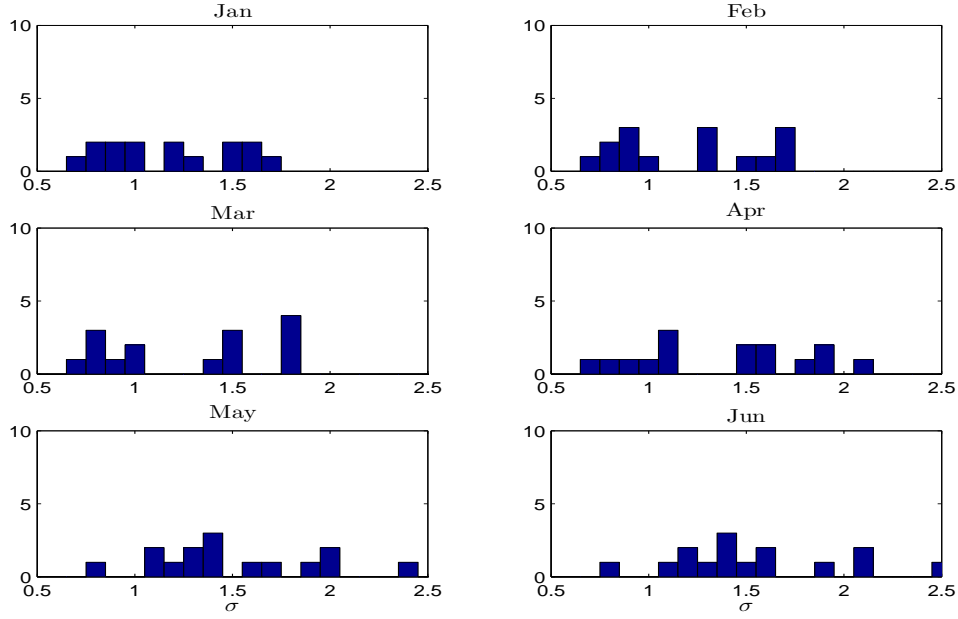
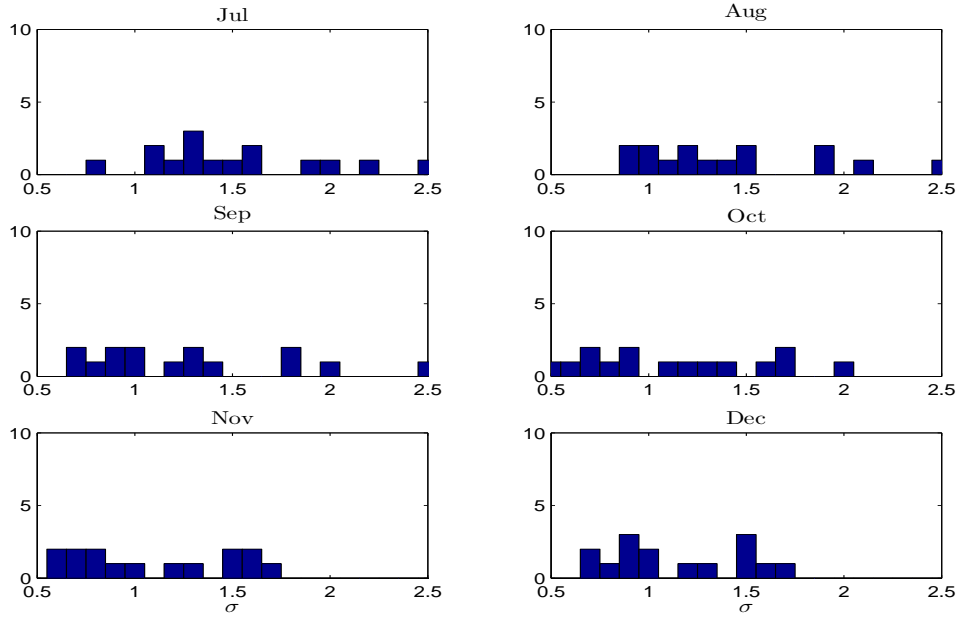


Figure E.1: μ parameters for t_{max} by month for each of the 16 sections. The differences are all centred around 0.



(a) Jan to Jun



(b) Jul to Dec

Figure E.2: σ parameters for t_{max} by month for each of the 16 sections. Notice how the parameters between t_{max} and t_{min} are similar.

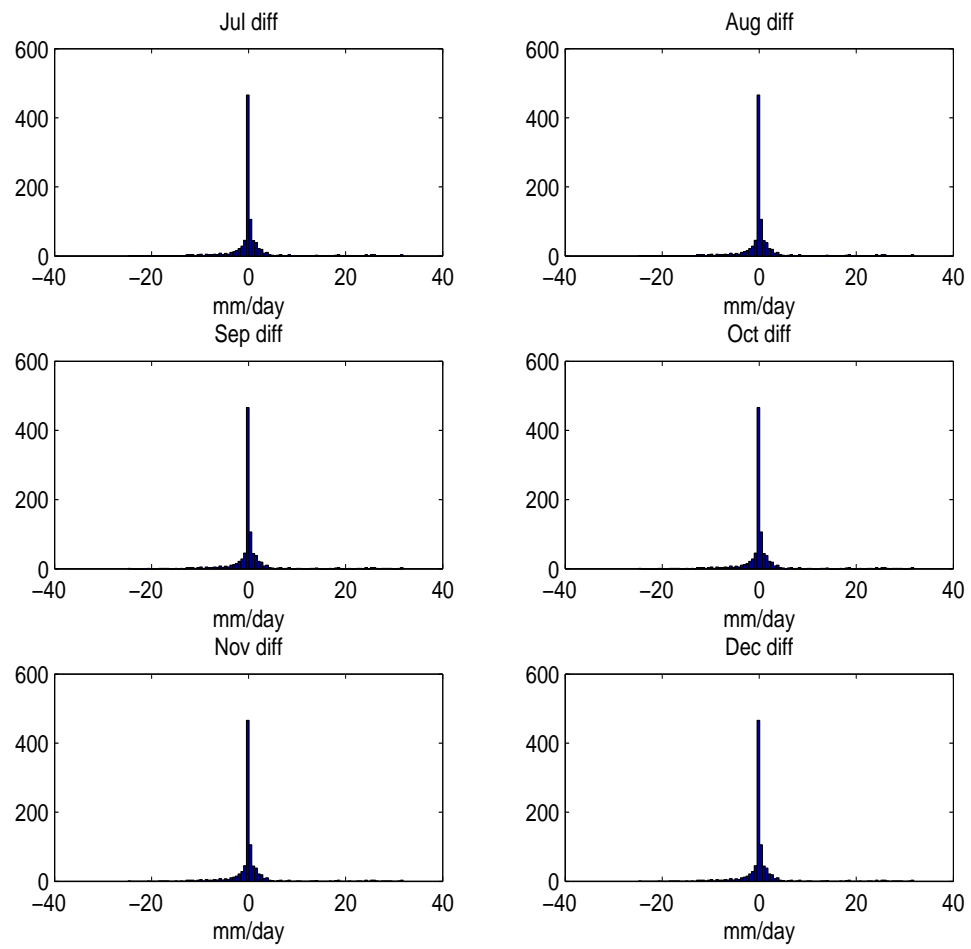


Figure E.3: Comparison of gridded precipitation and a 9 member ensemble for gridded precipitation from July to December.

Glossary

CERES-Maize a large scale physical simulation model to estimate maize yield [37], [19]. 8, 33, 34, 40–45, 61, 107–109, 111, 112, 117, 118, 120, 125, 128, 129, 146, 148, 150–152, 156, 158–161, 164, 185, 209, 223, 226, 229–232

DEMETER multi-model ensemble seasonal forecasts using seven different European models evaluated over the same historical period [60]. 94

Ignorance (IGN) skill score ($-\log_2(p)$) which measures the skill of a probabilistic forecast, the lower the ignorance the more skill a model has [25], [72]. 34, 56, 58, 77, 84, 85, 94, 99, 107, 108, 112, 116–118, 125, 128, 129, 134, 144, 146, 148, 156, 158, 161, 229

USHCN United States Historical Climatology Network manage a network of meteorological weather stations throughout the US [52]. 38, 39, 41, 163, 165, 171, 172, 178, 180

attractor a set of points towards which x_i integrate forward to over time. 52, 71

bench mark model a simple model that other models are measured against, with the expectation that this model will be beaten. 56, 57, 61, 64, 77, 107, 116, 117, 120, 150, 160, 229

blending when the imperfect model is weighted with the climatological distribution using α which can be between 0 and 1 [82]. 61, 98–100, 102

climatology distribution calculated by kernel dressing a large data set of outcomes. 33, 57, 61, 64, 65, 77, 79, 81–84, 94–96, 98–101, 105–107, 116–118, 120, 228, 229

dissipative a dynamical system where, on average, line segments shrink. 52

forecast bust where the outcome is vanishingly small probability mass as the outcome fell in an area of low probability in the forecast distribution. 95, 98, 112

forecast distribution from kernel dressing an ensemble of estimates, where in certain cases the ensemble could be a singleton ensemble [4]. 104, 125

forecast-outcome archive data set used to examine the accuracy of the models. Contains a series of ensemble of estimates (from the models) and outcomes. To create a forecast distribution the ensemble is kernel dressed with Gaussian kernels [4]. In DEMETER the ensemble contains nine ensemble members [60].. 33, 35, 57, 59, 61, 65, 77, 79, 80, 82–84, 86, 94, 105, 227

imperfect model a model which is an imperfect structural approximation of the system it is representing. 64, 234, 235, 238, 240

kernel dressing parameters the parameters are the width (σ) and offset (u) of the kernel used for kernel dressing. In this thesis the parameters are usually set by minimising ignorance [18]. The exception is in Sections 3.1 to 3.4.. 54, 56, 58, 80–84, 86, 99, 105, 134, 144

kernel dressing a method of converting an ensemble of estimates into a forecast distribution by, in this Thesis, putting a Gaussian kernel on each estimate [5]. 52, 54, 57, 61, 80, 107, 108, 111, 112, 116, 117, 123, 129, 131, 137, 210, 229

leave-one-out cross-validation a method where the forecast-outcome data is divided into two groups by leaving one point out. The kernel dressing parameters are estimated using the large group of forecast-outcome points and skill of the forecast is measured using the omitted forecast-outcome point. This is then repeated so that there is a set of kernel dressing parameters, the parameters are the median [82].. 57, 81, 84, 108, 111, 117, 120, 129, 131, 144, 146

meteorological observations there are many meteorological observations, in this thesis we mainly consider t_{min} , t_{max} and precipitation from the USHCN [52]. 33, 34, 37–41, 45, 60, 163, 165, 171, 172, 175, 178, 180, 206, 230, 232

model inadequacy errors in the model’s forecast caused by the model not capturing the system which generated the outcomes. 34, 58, 65, 80,

94, 227

multi-model forecast where the forecast is from more than one model.

95, 96, 98, 100

observational uncertainty imprecision caused by imperfections in the observations. 34, 37, 39, 58, 111, 163

realistic sample a sample that lies on or close to the attractor of the dynamical system. 71

relative Ignorance when the ignorance of models is bench marked against a “zero-skill” model. In this Thesis the “zero-skill” models used are climatology and persistence.. 56, 57, 82, 83, 95, 96, 98–102, 105, 107, 125, 134

skill score a measure of how accurate the probabilistic forecasts are at predicting the outcomes. 34, 56, 112

skill measure of how accurate a model is at approximating the system it is representing. 33, 34, 52, 56–58, 61, 62, 65, 81–84, 86, 94–96, 98–101, 107, 108, 112, 116–118, 120, 125, 129, 134, 137, 144, 146, 148, 150–152, 156, 158–161, 210, 228–231

technical advancements developments in crop growing methods that would cause yield to rise given identical enviromental conditions for example fertiliser, higher yielding cultivars and pest control. 33, 34, 37, 43, 45–48, 60, 164, 180, 194, 199, 201, 206, 208, 230

weighting the amount of weight each model's forecast is given. For example if the multiple models are equally weighted the weight given to each model's forecast is $\frac{1}{m}$ and where $\sum_{i=1}^m \alpha_i = 1$ where α is the weighting parameter and m is the number of models. 98, 99

Bibliography

- [1] Hirotugu Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974.
- [2] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79, 2010.
- [3] Donald G Baker. Effect of observation time on mean temperature estimation. *Journal of Applied Meteorology*, 14(4):471–476, 1975.
- [4] Jochen Bröcker and Leonard A Smith. Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting*, 22(2):382–388, 2007.
- [5] Jochen Bröcker and Leonard A Smith. From ensemble forecasts to predictive distribution functions. *Tellus A*, 60(4):663–678, 2008.
- [6] Roberto Buizza, PL Houtekamer, Gerald Pellerin, Zoltan Toth, Yuejian Zhu, and Mozheng Wei. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review*, 133(5):1076–1097, 2005.
- [7] Roberto Buizza, M Milleer, and TN Palmer. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 125(560):2887–2908, 1999.
- [8] Roberto Buizza and Tim N Palmer. Impact of ensemble size on ensemble prediction. *Monthly Weather Review*, 126(9), 1998.

- [9] Richard E Carlson, Dennis P Todey, and Sterling E Taylor. Midwestern corn yield and weather in relation to extremes of the Southern Oscillation. *Journal of production agriculture*, 9(3):347–352, 1996.
- [10] AJ Challinor, JM Slingo, TR Wheeler, and FJ Doblas-Reyes. Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles. *Tellus A*, 57(3):498–512, 2005.
- [11] AJ Challinor, TR Wheeler, PQ Craufurd, JM Slingo, and DIF Grimes. Design and optimisation of a large-area process-based model for annual crops. *Agricultural and forest meteorology*, 124(1):99–120, 2004.
- [12] Kevin KW Cheung. A review of ensemble forecasting techniques with a focus on tropical cyclone forecasting. *Meteorological Applications*, 8(03):315–332, 2001.
- [13] William D Collins, Cecilia M Bitz, Maurice L Blackmon, Gordon B Bonan, Christopher S Bretherton, James A Carton, Ping Chang, Scott C Doney, James J Hack, Thomas B Henderson, et al. The community climate system model version 3 (CCSM3). *Journal of Climate*, 19(11):2122–2143, 2006.
- [14] Neil E Day. Estimating the components of a mixture of normal distributions. *Biometrika*, 56(3):463–474, 1969.
- [15] Arthur T DeGaetano. Attributes of several methods for detecting discontinuities in mean temperature series. *Journal of Climate*, 19(5), 2006.
- [16] Arthur P Dempster, Nan M Laird, Donald B Rubin, et al. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal statistical Society*, 39(1):1–38, 1977.
- [17] Francisco J Doblas-Reyes, Renate Hagedorn, and TN Palmer. The rationale behind the success of multi-model ensembles in seasonal forecasting–II. Calibration and combination. *Tellus A*, 57(3):234–252, 2005.
- [18] Hailiang Du and Leonard A. Smith. Parameter estimation through ignorance. *Phys. Rev. E*, 86:016213, Jul 2012.
- [19] Joshua Elliott, David Kelly, Neil Best, Michael Wilde, Michael Glotter, and Ian Foster. The parallel system for integrating impact models and sectors (pSIMS).

- In *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery*, page 21. ACM, 2013.
- [20] Joshua Elliott, David Kelly, James Chryssanthacopoulos, Michael Glotter, Kanika Jhunjhnuwala, Neil Best, Michael Wilde, and Ian Foster. The Parallel System for Integrating Impact Models and Sectors (pSIMS). *Environmental Modelling & Software*, 2014.
 - [21] S. Fall, A. Watts, J. Nielsen-Gammon, E. Jones, D. Niyogi, J.R. Christy, and R.A. Pielke Sr. Analysis of the impacts of station exposure on the US Historical Climatology Network temperatures and temperature trends. *Journal of Geophysical Research*, 116(D14):D14120, 2011.
 - [22] Von Fischer, C Joseph, Larry L Tieszen, and David S Schimel. Climate controls on C3 vs. C4 productivity in North American grasslands from carbon isotope composition of soil organic matter. *Global Change Biology*, 14(5):1141–1155, 2008.
 - [23] Isla Gilmour, Leonard A. Smith, and Roberto Buizza. Linear regime duration: Is 24 hours a long time in synoptic weather forecasting? *Journal of the atmospheric sciences*, 58(22):3525–3539, 2001.
 - [24] Tilmann Gneiting, Adrian E Raftery, Anton H Westveld III, and Tom Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133(5), 2005.
 - [25] Irving John Good. *Good thinking: The foundations of probability and its applications*. U of Minnesota Press, 1983.
 - [26] Renate Hagedorn, Francisco J Doblas-Reyes, and TN Palmer. The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus A*, 57(3):219–233, 2005.
 - [27] Graeme L Hammer, Zhanshan Dong, Greg McLean, Al Doherty, Carlos Messina, Jeff Schussler, Chris Zinselmeyer, Steve Paszkiewicz, and Mark Cooper. Can changes in canopy and/or root system architecture explain historical maize yield trends in the US corn belt? *Crop Science*, 49(1):299–312, 2009.

- [28] James W Hansen, Alan W Hodges, and James W Jones. ENSO Influences on agriculture in the Southeastern United States. *Journal of Climate*, 11(3):404–411, 1998.
- [29] JW Hansen and JW Jones. Scaling-up crop models for climate variability applications. *Agricultural Systems*, 65(1):43–72, 2000.
- [30] Ed Hawkins, Thomas E Fricker, Andrew J Challinor, Christopher AT Ferro, Chun Kit Ho, and Tom M Osborne. Increasing influence of heat stress on French maize yields from the 1960s to the 2030s. *Global change biology*, 19(3):937–947, 2013.
- [31] Chris Hewitt. The ENSEMBLES project.
- [32] Chris D Hewitt. Ensembles-based predictions of climate changes and their impacts. *Eos, Transactions American Geophysical Union*, 85(52):566–566, 2004.
- [33] R Wayne Higgins and Climate Prediction Center. *Improved United States precipitation quality control system and analysis*. NOAA, National Weather Service, National Centers for Environmental Prediction, Climate Prediction Center, 2000.
- [34] S.M.W Higgins, H Du, and L.A Smith. On the design and use of ensembles of multi-model simulations for forecasting. *Nonlinear Processes in Geophysics*, submitted 2015.
- [35] David Caster Hoaglin, Frederick Mosteller, and John Wilder Tukey. *Understanding robust and exploratory data analysis*, volume 3. Wiley New York, 1983.
- [36] Ian T Jolliffe and David B Stephenson. *Forecast verification: a practitioner’s guide in atmospheric science*. John Wiley & Sons, 2012.
- [37] James W Jones, G Hoogenboom, CH Porter, KJ Boote, WD Batchelor, LA Hunt, PW Wilkens, U Singh, AJ Gijssman, and JT Ritchie. The DSSAT cropping system model. *European journal of agronomy*, 18(3):235–265, 2003.
- [38] JW Jones. *Soygro V5. 42: Soybean Crop Growth Simulation Model: User’s Guide*. University of Florida, Department of Agricultural Engineering, 1989.
- [39] Eugenia Kalnay, Masao Kanamitsu, Robert Kistler, William Collins, D Deaven, L Gandin, Mo Iredell, Suranjana Saha, Glenn White, John Woollen, et al. The

- NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, 77(3):437–471, 1996.
- [40] Christopher J Kucharik and Navin Ramankutty. Trends and variability in US corn yields over the twentieth century. *Earth Interactions*, 9(1):1–29, 2005.
- [41] M Leutbecher and TN Palmer. Ensemble forecasting. *Journal of Computational Physics*, 227(7):3515–3539, 2008.
- [42] Rachel Licker, Matt Johnston, Jonathan A Foley, Carol Barford, Christopher J Kucharik, Chad Monfreda, and Navin Ramankutty. Mind the gap: how do climate and agricultural management explain the yield gap of croplands around the world? *Global Ecology and Biogeography*, 19(6):769–782, 2010.
- [43] David B Lobell, Kimberly Nicholas Cahill, and Christopher B Field. Historical effects of temperature and precipitation on California crop yields. *Climatic Change*, 81(2):187–203, 2007.
- [44] David B Lobell and Christopher B Field. Global scale climate-crop yield relationships and the impacts of recent warming. *Environmental research letters*, 2(1):014002, 2007.
- [45] David B Lobell, Wolfram Schlenker, and Justin Costa-Roberts. Climate trends and global crop production since 1980. *Science*, 333(6042):616–620, 2011.
- [46] Edward N Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963.
- [47] Aleksandr Mikhailovich Lyapunov. The general problem of the stability of motion. *International Journal of Control*, 55(3):531–534, 1992.
- [48] The MathWorks. *Statistics Toolbox TM 7*, volume Revised Version for 7.5 (Release 2011a). CRC press, 2011.
- [49] Matthew J Menne and Claude N Williams Jr. Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, 22(7), 2009.
- [50] Matthew J Menne, Claude N Williams Jr, and Russell S Vose. The US Historical Climatology Network monthly temperature data, version 2. *Bulletin of the American Meteorological Society*, 90(7), 2009.

- [51] M.J. Menne, C.N. Williams Jr, and M.A. Palecki. On the reliability of the US surface temperature record. *Journal of Geophysical Research*, 115(D11):D11108, 2010.
- [52] M.J. Menne, CN Williams Jr, and RS Vose. United States Historical Climatology Network daily temperature, precipitation, and snow data. Technical report, United States Historical Climatology Daily Network, 2012.
- [53] Fedor Mesinger, Geoff DiMego, Eugenia Kalnay, Kenneth Mitchell, Perry C Shafran, Wesley Ebisuzaki, Dušan Jovic, Jack Woollen, Eric Rogers, Ernesto H Berbery, et al. North American regional reanalysis. *Bulletin of the American Meteorological Society*, 87(3):343–360, 2006.
- [54] JW Mjelde, TN Thompson, FM Hons, JT Cothren, and CG Coffman. Using Southern Oscillation information for determining corn and sorghum profit-maximizing input levels in east-central Texas. *Journal of production agriculture*, 10(1):168–175, 1997.
- [55] Derek W Moore and Edward A Spiegel. A thermally excited non-linear oscillator. *The Astrophysical Journal*, 143:871, 1966.
- [56] PAP Moran. Some remarks on animal population dynamics. *Biometrics*, 6(3):250–258, 1950.
- [57] USDA NASS. Crop Production Annual Summary. *US Department of Agriculture, National Agricultural Statistics Service*, 1936-2012.
- [58] USDA NASS. *Field Crops Usual Planting and Harvesting Dates*, agricultural handbook 628 edition, 2010.
- [59] TN Palmer. The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society*, 128(581):747–774, 2002.
- [60] TN Palmer, A Alessandri, U Andersen, P Cantelaube, M Davey, Pascale Delécluse, M Déqué, E Díez, FJ Doblas-Reyes, H Feddersen, et al. Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, 85(6), 2004.

Bibliography

- [61] TN Palmer, R Mureau, and F Molteni. The Monte Carlo forecast. *Weather*, 45(6):198–207, 1990.
- [62] Karl Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, pages 71–110, 1894.
- [63] Anders Persson and Federico Grazzini. User guide to ECMWF forecast products. *Meteorological Bulletin*, 3:2, 2005.
- [64] John R Porter and Mikhail A Semenov. Crop responses to climatic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1463):2021–2035, 2005.
- [65] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.
- [66] Adrian E Raftery, Tilmann Gneiting, Fadoua Balabdaoui, and Michael Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133(5), 2005.
- [67] Navin Ramankutty, Amato T Evan, Chad Monfreda, and Jonathan A Foley. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, 22(1), 2008.
- [68] Deepak K Ray, Navin Ramankutty, Nathaniel D Mueller, Paul C West, and Jonathan A Foley. Recent patterns of crop yield growth and stagnation. *Nature Communications*, 3:1293, 2012.
- [69] C. W. Richardson and D.A. Wright. *WGEN: A Model for Generating Daily Weather Variables*. USDA Agriculture Research Service, August 1984.
- [70] William E Ricker. Stock and recruitment. *Journal of the Fisheries Board of Canada*, 11(5):559–623, 1954.
- [71] Andrew W Robertson, Upmanu Lall, Stephen E Zebiak, and Lisa Goddard. Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. *Monthly Weather Review*, 132(12), 2004.
- [72] Mark S Roulston and Leonard A Smith. Evaluating probabilistic forecasts using information theory. *Monthly Weather Review*, 130(6):1653–1660, 2002.

- [73] M.S. Roulston and L.A. Smith. Combining dynamical and statistical ensembles. *Tellus A*, 55(1):16–30, 2003.
- [74] Suranjana Saha, Shrinivas Moorthi, Hua-Lu Pan, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, Robert Kistler, John Woollen, David Behringer, et al. The NCEP climate forecast system reanalysis. *BAMS*, 91(8):1015–1057, 2010.
- [75] Wolfram Schlenker and Michael J Roberts. Nonlinear effects of weather on corn yields. *Applied Economic Perspectives and Policy*, 28(3):391–398, 2006.
- [76] Mikhail A Semenov, Roger J Brooks, Elaine M Barrow, and Clarence W Richardson. Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. *Climate Research*, 10(2):95–107, 1998.
- [77] Mikhail A Semenov and JR Porter. Climatic variability and the modelling of crop yields. *Agricultural and forest meteorology*, 73(3):265–283, 1995.
- [78] USDA National Agricultural Statistics Service. Quick Stats. <http://quickstats.nass.usda.gov>, 2014. [Online; accessed 4-February-2014].
- [79] Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.
- [80] L.A. Smith, H. Du, E.B. Suckling, and Falk Niehörster. Probabilistic skill in ensemble seasonal forecasts. *Quarterly Journal of the Royal Meteorological Society*, tbc(tbc):tbc, 2014.
- [81] Leonard Smith. *Chaos: a very short introduction*. Oxford University Press, 2007.
- [82] Leonard A. Smith, Hailiang Du, Emma B. Suckling, and Nieh"0rster. Probabilistic skill in ensemble seasonal forecasts. *Royal Meteorological Society*, 2014 - in review.
- [83] Julien Clinton Sprott. *Chaos and time-series analysis*, volume 69. Oxford University Press Oxford, 2003.
- [84] Paul W Stackhouse Jr, Shashi K Gupta, Stephen J Cox, J Colleen Mikovitz, and Marc Chiacchio. New results from the NASA/GEWEX surface radiation budget project: Evaluating El Nino effects at different scales. 2002.
- [85] Emma B Suckling and Leonard A Smith. An evaluation of decadal probability forecasts from state-of-the-art climate models. *Journal of climate*, 26(23), 2013.

- [86] Earl R Swanson and James C Nyankori. Influence of weather and technology on corn and soybean yield trends. *Agricultural Meteorology*, 20, 1979.
- [87] Michael Tannura, Scott Irwin, and Darrel L Good. Weather, technology, and corn and soybean yields in the US corn belt. *Technology, and Corn and Soybean Yields in the US Corn Belt (February 1, 2008)*, 2008.
- [88] Zoltan Toth and Eugenia Kalnay. Ensemble forecasting at NMC: The generation of perturbations. *Bulletin of the American Meteorological Society*, 74(12):2317–2330, 1993.
- [89] Sakari M Uppala, PW Kållberg, AJ Simmons, U Andrae, V Bechtold, M Fiorino, JK Gibson, J Haseler, A Hernandez, GA Kelly, et al. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612):2961–3012, 2005.
- [90] USDA-NASS. Annual crop summary.
- [91] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 0-387-95457-0.
- [92] Russell S Vose, Claude N Williams, Thomas C Peterson, Thomas R Karl, and David R Easterling. An evaluation of the time of observation bias adjustment in the US Historical Climatology Network. *Geophysical research letters*, 30(20), 2003.
- [93] Bin Wang, June-Yi Lee, In-Sik Kang, J Shukla, C-K Park, A Kumar, J Schemm, S Cocke, J-S Kug, J-J Luo, et al. Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Climate dynamics*, 33(1):93–117, 2009.
- [94] Xuguang Wang and Craig H Bishop. Improvement of ensemble reliability with a new dressing kernel. *Quarterly Journal of the Royal Meteorological Society*, 131(607):965–986, 2005.
- [95] James Watson and Andrew Challinor. The relative importance of rainfall, temperature and yield data for a regional-scale crop model. *Agricultural and Forest Meteorology*, 170:47–57, 2013.
- [96] A Weisheimer, FJ Doblas-Reyes, TN Palmer, A Alessandri, A Arribas, M Déqué, Noel Keenlyside, M MacVean, A Navarra, and P Rogel. ENSEMBLES: A new multi-

- model ensemble for seasonal-to-annual predictions – Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophysical research letters*, 36(21), 2009.
- [97] Timothy R Wheeler, Peter Q Craufurd, Richard H Ellis, John R Porter, and PV Vara Prasad. Temperature variability and the yield of annual crops. *Agriculture, Ecosystems & Environment*, 82(1):159–167, 2000.
- [98] Daniel S Wilks. Comparison of ensemble-MOS methods in the Lorenz'96 setting. *Meteorological Applications*, 13(3):243–256, 2006.
- [99] Daniel S Wilks. *Statistical methods in the atmospheric sciences*, volume 100. Academic press, 2011.
- [100] D.S. Wilks. Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210, 1984.
- [101] Q Yu, H Hengsdijk, and JD Liu. Application of a progressive-difference method to identify climatic factors causing variation in the rice yield in the Yangtze Delta, China. *International Journal of Biometeorology*, 45(2):53–58, 2001.