Probabilistic Forecasting of Agricultural Yield

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

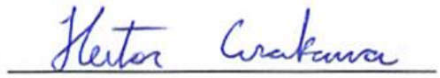
by

Heitor Haselmann Arakawa

August 2019

APPROVAL SHEET

This Dissertation
is submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Author Signature: 

This Dissertation has been read and approved by the examining committee:

Advisor: Roman Krzysztofowicz

Committee Member: William T. Scherer

Committee Member: James H. Lambert

Committee Member: Garrick Louis

Committee Member: Karen Kafadar

Committee Member: _____

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

August 2019

Abstract

Forecasting agricultural yield at a local or regional level is of utmost importance to decision makers in the food supply chain sector. Growers must make decisions based on projected yields. For example, they might be interested in selling their production in advance to cover part of their costs or to hedge potential price volatility. In Brazil, these stakeholders rely on public forecasts provided by the Companhia Nacional de Abastecimento (CONAB) and the Instituto Brasileiro de Geografia e Estatística (IBGE). However, the forecasts published by these sources have something in common: they are deterministic and discount or omit the uncertainty associated with their estimates. Moreover, forecasts for the same crop, region, and time may differ from source to source. This research develops a methodology to quantify the uncertainties associated with deterministic forecasts of soybean crop yields in the state of Mato Grosso, Brazil. The theory of Bayesian Processor of Forecasts (BPF) is reviewed and expanded to incorporate a judgmental prior distribution function modeled from the farmers' assessments. Farmers in Mato Grosso were interviewed and a set of quantiles of yields was assessed for each one. Individual prior distribution functions were modeled using these sets of quantiles and then combined into a single prior distribution function. The deterministic forecasts were collected from reports issued by CONAB and IBGE in October, February, and May annually between 1993 and 2017. The BPF model is able to merge these deterministic forecasts, and produce probabilistic forecasts of the yield. Various BPF models were developed for different lead times and using different prior information. The empirical and simulated results of this study exemplify the advantages of using the BPF theory and provide a guideline on how to apply this methodology to combine prior distribution functions, fuse information from different sources, and produce probabilistic forecasts.

Key words: agricultural yield; probabilistic forecasting; Bayesian forecaster; Bayesian Processor of Forecasts; data modeling; judgmental assessment; expert uncertainty.

Acknowledgements

This work was supported by the Brazilian agency Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) in partnership with Laspau, a nonprofit affiliated with Harvard University.

I would like to thank the University of Virginia (UVA) for the opportunity to work as a teaching assistant for the Accelerated Master Program in Systems Engineering.

I would like to thank Dr. Roman Krzysztofowicz for serving as my advisor and providing insightful guidance during this research. Drs. William T. Scherer, James H. Lambert, Garrick Louis, and Karen Kafadar not only served as members of my research committee, but also gave me helpful advice. I thank Drs. Lambert and Scherer specifically for all the fruitful discussions about my professional development. I owe particular gratitude to Dr. Kafadar for the invaluable discussions about statistics, teaching philosophy, and the academic life. In addition, Jayne Weber and Elizabeth Harrison provided an outstanding guidance through the administrative part of the program. I am also indebted to Fernanda Schwantes, Paulo Ozaki, and the many farmers from Mato Grosso who participated in this study.

The UVA community was also essential to the completion of this project. I am grateful for the friends I made in Charlottesville. I would like to thank Baozhen Xie and Adriana Vito for the counseling and help during the many rough patches of graduate school.

Finally, I am grateful for the support from my family, particularly from my mother Maria Cristina Haselmann Paulo. This project wouldn't have been possible without her presence, many times in person, and her emotional (and financial) support. I realize now that the tolls of an international degree fall not only in the students, but also in their families. Emily Miller also had an essential role by being constantly cheerful and helpful. Her positivity, along with the encouragement of her wonderful family, alleviated the stress of being so distant from home.

Table of contents

List of Acronyms

ASBG	Acompanhamento da Safra Brasileira de Grãos.
BMA	Bayesian Model Averaging.
BPF	Bayesian Processor of Forecasts.
CEPEA	Centro de Estudos Avançados em Economia Aplicada.
CONAB	Companhia Nacional de Abastecimento.
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária.
IBGE	Instituto Brasileiro de Geografia e Estatística.
LSPA	Levantamento Sistemático da Produção Agrícola.
NASS	National Agricultural Statistics Service.
NQT	Normal Quantile Transformation.
PAM	Produção Agrícola Municipal.
USDA	United States Department of Agriculture.
WASDE	World Agricultural Supply and Demand Estimates.

List of Distributions

LR1–LP	Log-ratio Laplace.
LC1–WB	Log-reciprocal type I Weibull.
LC1–IW	Log-reciprocal type I inverted Weibull.
LC1–LL	Log-reciprocal type I log-logistic.
LC2–IW	Log-reciprocal type II inverted Weibull.

List of Key Symbols

W	predictand.
W^F	net harvested yield of a field.
W^R	net harvested yield of a region.
$\mathcal{W}$	sample space of W.
w	realization of W.
$w(t)$	observed yield in year t.
$w^i(p)$	p-probability quantile of W^i, for $i = F, R$.
$w^i_j(p)$	p-probability quantile of W^i assessed from farmer j, for i = F, R.
X	predictor.
$\mathcal{X}$	sample space of X.
x	realization of X.
q	standard normal density function.
Q	standard normal distribution function.
Q^{-1}	inverse of the standard normal distribution function (quantile function).
g	prior density function.
G	prior distribution function.
G^{-1}	inverse of prior distribution function (quantile function).
G^H	historical prior distribution function.
G^i_t	judgmental prior distribution function at time t, for $i = F, R$.
$G^i_{t,j}$	judgmental prior distribution function assessed from farmer j at time t, for $i = F, R$.

$G_t^{i(-1)}$	inverse of judgmental prior distribution function.
κ	expected density function of X.
$\bar{K}$	initial estimate of marginal distribution function of X.
$\bar{K}_{t,l}^i$	marginal distribution function of X_l associated with judgmental prior distribution function at time t, for $i = F, R$.
f	likelihood function.
ϕ	posterior density function.
Φ	posterior distribution function.
Φ^{-1}	posterior quantile function.
Ψ	stochastic transformation.

List of Figures

List of Tables

1. Introduction

In agriculture, farmers are often required to make decisions under uncertain conditions, such as purchasing fertilizers, chemicals, seeds, and even selling their production ahead of harvest. Generally, their decision process for these issues is supported by information derived from multiple sources such as government agencies and private companies. However, there is uncertainty associated with these estimates and forecasts. This uncertainty is not always communicated by the organizations releasing such data. Even in cases when uncertainty is communicated, it may not be well understood by those not trained in probabilistic reasoning. For farmers, the uncertainty surrounding predictions of input prices, yield, and commodity markets, for example, can have huge impacts on their livelihoods.

1.1 Background for Research

1.1.1 Forecasts of Agricultural Yield

The United States Department of Agriculture (USDA) publishes several reports during the agricultural season estimating national and international production for the most significant crops. The USDA releases information through several sub-departments such as the Economic Research Service (ERS), Foreign Agricultural Service (FAS), National Agricultural Statistics Service (NASS), and the World Agricultural Outlook Board (WAOB).

These reports affect decision making in both business planning and policy making. Typically, agricultural models used in estimating acreage to be harvested and yield combine information from multiple sources, such as surveys, meteorological stations, and crop monitors. These models may involve complex estimations or forecasting algorithms, even though their outputs often appear as point estimates. Moreover, such estimates rarely acknowledge the uncertainty surrounding them.

The Brazilian government makes a similar effort to produce information for the agri-

culture sector through the Instituto Brasileiro de Geografia e Estatistica (IBGE) and the Companhia Nacional de Abastecimento (CONAB). The IBGE is associated with the Ministry of Planning, Budget and Management, and CONAB is associated with the Ministry of Agriculture and Supply. Both ministries are part of the executive branch.

Forecasting soybean production on a national scale is particularly important to countries such as Brazil and United States, whose climates are ideal for this vital crop. Approximately 21.6% of the Brazilian GPD originated from agribusiness in 2017 according to CEPEA (2018). Soybean production alone contributed 142.3 billion reais in 2017, equal to 25% of the total agricultural production in the country that year (IBGE, 2017b). Internationally, Brazil competes with the United States to be the world leader in production. In the 2017/2018 season, Brazil produced 120.80 million metric tons and the United States produced 120.07 million metric tons, corresponding to approximately 36% and 35%, respectively, of the global soybean production (USDA, 2019).

1.1.2 Challenges in the Brazilian Soybean Production

While Brazil enjoys the spoils of this lucrative industry, being one of the largest soybean producers presents many challenges. Approximately 45% of soybeans in Brazil were produced in the Center-West region of the country in the 2017/2018 season (IBGE, 2019). This region is composed of 3 states: Mato Grosso, Mato Grosso do Sul, and Goiás. The climate in this region is conducive to growing soybean, but there is a deficit of transportation alternatives to the main ports in Brazil for exportation.

Trucks are the transportation method of choice at the moment, but there are many downsides, such as costs associated with fuel, road conditions, wages, and so on. Certainly, the volatility of trucks as the main method of transportation leaves farmers vulnerable. This dilemma is already playing out on the national stage. Constant increases in the cost of transportation driven by increases in both tolls and the price of diesel led to a national truck driver's strike in May of 2018. This strike seriously impacted the exportation pos-

sible at that time (and ultimately, profit), given that there is limited storage capacity for unshipped goods with a shelf life. The strike ended with a settlement between the truckers' union and the government that imposed minimum freight rates and changes in the adjustments of the diesel price. These policies are still under review and are considered to be controversial by many agents in the supply chain. The total reliance on the road system for agricultural exportation is risky for both farmers and grain traders.

Other methods of transportation are being explored, though none have emerged as an obvious choice. Over the last few years, the government and private companies have invested in the railroad system, but the capacity is still limited. This system requires a large amount of investment, and consequently it is operated by a small number of companies. The rail freight rates usually compete with the road system, leaving a limited incentive to use trains depending on the size of the cargo. Waterways have a great potential in Brazil, but also depend heavily on public investments.

Brazil's logistical challenges are closely related to its storage capacity. The soybean spoilage process is relatively long as compared to other agricultural products, which allows them to be stored in silos and warehouses for up to 6 months under optimal moisture and temperature conditions, without any quality problems. Storing soybeans longer is possible, but additional efforts to maintain quality are required. However, the storage capacity in the country is still insufficient to alleviate the logistical problems.

These are some of the reasons why forecasting yield is important. Planning an entire season ahead while expecting to deal with many challenges requires reliable information. Furthermore, assessing the uncertainties associated with the forecasts will potentially improve decision making throughout the supply chain.

1.2 Research Objectives

The overall objective is to develop a methodology to produce probabilistic agricultural yield forecasts. This methodology will incorporate judgmentally assessed prior information from farmers into a yield forecasting model. Specifically, judgmentally assessed prior distribution function will be input to a Bayesian forecasting model that will output a probabilistic forecast of the yield. Two forecasting models will be formulated.

1. A *forecasting model using historical prior*, which will produce probabilistic yield forecasts for a region or a field. The predictors will be the deterministic forecasts for the region issued by government organizations. This model will allow growers and other decision makers in agriculture to quantify the uncertainty related to their recorded data set and also take advantage of multiple sources of deterministic forecasts.

2. A *forecasting model using a judgmental prior*, which will produce probabilistic yield forecasts for a region or a field. Similarly to the previous model, the *predictors* will be the deterministic yield forecasts for the region, issued by a government sources. However, a group of J farmers will contribute to the prior information. Each farmer will produce a prior distribution function of the yield of the region and selected field. The mixture of the J distribution functions will form a prior distribution function of the *predictand*.

The methodology developed in this research will: (1) improve the judgmental assessments currently being done in agricultural forecasting systems, (2) apply a new way to combine judgmentally assessed information using Bayesian Model Averaging, and (3) quantify and display the uncertainty about the agricultural yield using Bayesian Processor of Forecasts.

1.3 Overview

In order to achieve the objectives of this research, the next chapter reviews the methodologies currently applied to produce deterministic forecasts of agricultural yield by the main government organizations in Brazil and in the United States. Chapter 3 reviews the Bayesian forecasting framework, specifically the theory of Bayesian Processor of Forecasts (BPF) and expands it to incorporate the judgmental prior distributions.

Chapter 4 is devoted to the modeling of judgmental distribution functions. In addition, it addresses the topic of combining prior distribution functions using the Bayesian Model Averaging (BMA). Chapter 5 develops a framework for transforming judgmental assessments of the yield of a field into the yield of a region through a field-region stochastic transformation. This transformation is consistent with the Bayesian framework.

Chapter 6 constructs BPF models from observed data of the yield of Mato Grosso. This chapter serves as a guideline to produce probabilistic forecasts of agricultural yield. Subsequently, chapter 7 applies the methodology described in chapter 3 to incorporate the farmers' assessments into the BPF models. Lastly, chapter 8 summarizes the conclusions obtained from analyzing the models and the results.

The appendices contain three additional items. Appendix A constructs examples of the Bayesian Model Averaging algorithm to analyze its sensitivity to calibration and informativeness of forecasters. Appendix B summarizes the methodology for modeling distribution functions. This methodology is used in various parts of the research. Appendix C exemplifies the implementation of the models created in this research using R. It serves as a practical guide for future readers to execute the framework developed in this dissertation.

2. Deterministic Yield Forecasts

This section reviews the methodology currently being used by some government organizations to produce deterministic forecasts of agricultural yield. It starts with a description of the soybean production timeline in order to understand the activities related to different stages of the crop season and the economic value of forecasts issued at various lead times.

2.1 Soybean Production Timeline

Different varieties of soybeans reach maturity at different times, i.e., the time interval from planting to harvest varies depending on the variety. Moreover, external conditions such as soil moisture, temperature, solar irradiance, and day length have a great influence on the development of the crops. The planting season is determined by an attempt to combine the expected external conditions with the requirements of each stage of crop development (EMBRAPA, 2013).

The soybean crop seasons in the United States and Brazil are not the same; therefore, the forecasts are published during different times of year. In general, the soybean production cycle occurs from October until the end of May in Brazil. The same cycle occurs from April until November in the United States.

Farmers will take many factors into account as they plan when to plant, fertilize, and harvest for the upcoming year. The farmer will base his or her decisions about when to undertake these efforts on how the crop is developing so far and on forecasts regarding the production yield. Other factors, like the fact that grain storage facilities must be reserved prior to the harvest season, must also be taken into account. Because the planning process has many moving parts, farmers will often use tools like the Soybean Production Calendar developed by Lee et. al. (2007), for example.

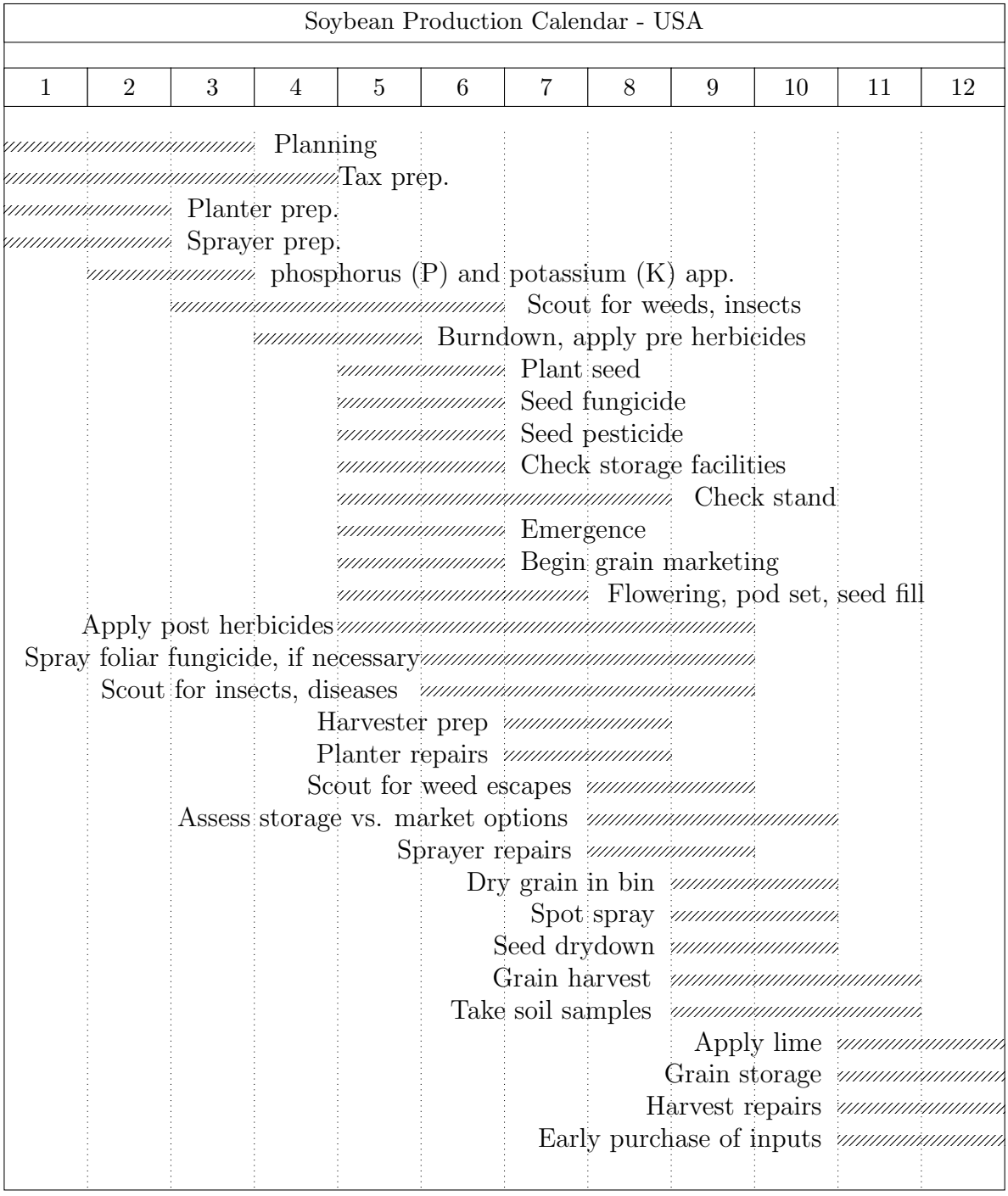


Figure 2.1 Soybean production calendar for the state of Kentucky (Lee et al., 2007).

This calendar lists an ideal time intervals for each agricultural operation during the season. The activities are divided into four types of operation: equipment maintenance/shop, crop management, grain storage, and business/marketing. The weather conditions and the growth requirements of the plant are the main factors that shape this calendar. Figure 2.1 shows the calendar and displays the activities and decisions made during the production cycle, exemplifying the complexity of a soybean production operation in Kentucky.

2.2 Brazilian Institute of Geography and Statistics (IBGE)

The Brazilian government produces statistics about its agricultural production through IBGE and CONAB . IBGE is part of the Ministry of Planning, Budget and Management in Brazil. The agricultural data is collected and analyzed by the Board of Research (Diretoria de Pesquisas - DPE) and, more specifically, supervised by the Agricultural Department (Coordenação de Agropecuária - DPE/COAGRO). IBGE‘s major publication regarding agricultural production is the Levantamento Sistemático da Produção Agrícola (LSPA). The LSPA report contains municipal data about estimated area, production, and yield for thirty five crops (including soybeans), and is published monthly (IBGE, 2002).

Private companies also produce agricultural information in some regions. For instance, the Instituto Mato-grossense de Economia Agropecuária (IMEA) located in the State of Mato Grosso collects information about the major crops in the center-west region of Brazil. The IMEA is a non-profit institution that publishes weekly and monthly reports about agricultural production.

2.2.1 Levantamento Sistemático da Produção Agrícola (LSPA)

The LSPA is a report issued monthly by IBGE from January to December, containing estimates of planted area, harvested area, production, and yield for Brazil for a number of agricultural crops, including soybeans. The IBGE does not contact the growers directly to

collect data about their production. Instead, three groups are formed to aggregate local data: the Grupo de Coordenação de Estatísticas Agropecuárias (GCEA), the Comissão Municipal de Estatísticas Agropecuárias (COMEA) and the Comissão Regional de Estatísticas Agropecuárias (COREA). The methodology used to elaborate the LSPA report is decribed in IBGE (2018).

These three groups are formed in several locations in Brazil every month. The COMEA gathers individuals connected to agricultural institutions and associations to review information at a local level. The GCEA and COREA gather individuals from agricultural unions, associations, companies, and banks to review information at a regional level. The participants of each meeting inform IBGE of their estimates regarding:

- Type of crop
- Area (in hectares): planted or to be planted, harvested or to be harvested, and harvested in the current month
- Production (in tons)
- Yield (in kilograms/hectare)
- Price per unit (R$/tons)
- % of production stored
- Irrigation: Yes/No, and % of current area
- Abnormal event, taking the values: (0) missing, (1) excessive rain, (2) insufficient rain, (3) frost, (4) hail, (5) pest, (6) disease, (7) eradication/abandoned, (8) other, (9) none
- Stage, taking the values: (0) missing, (1) planning, (2) soil management, (3) planting, (4) cultivation, (5) harvest in progress, (6) harvest complete, (7) marketing, (8) emergency, (9) flowering, (10) podding, (11) off season

The groups discuss and fill up paper forms with their estimates for the variables above. Once this information is collected, the GCEA, COMEA, and COREA send it to IBGE to compose the LSPA report. The reports issued in October, November, and December represent a prediction of the actual values of harvested area and production after the season is completed. These forecasts are deterministic.

The forecast of the current yield reported by the LSPA is calculated using the observed yields of the last 5 seasons. Let the continuous variate W be the yield of a certain state reported by the LSPA, and its realization be denoted by $w \in \mathcal{W}$, where $\mathcal{W}$ is the sample space. Let $w(t)$ be the observed yield in the state at year t. Let $w_L = min\{w(t)\}$ and $w_U = max\{w(t)\}$, for $t = 1, ..., 5$. The forecast of the yield for the current year, $w(6)$, is calculated as follows:

$$w(6) = \frac{\sum_{t=1}^{5} w(t) - w_L - w_U}{3}. \tag{2.1}$$

In other words, the LSPA eliminates the highest and the lowest observed yields in the last 5 years and takes the average of the remaining values to obtain a forecast for the current year. The estimated planted area is multiplied by the predicted yield to obtain a forecast about the production.

The final estimates are made by IBGE using the information collected from the groups and their analysis of conditions that may affect the production. During the planning and soil management stages, IBGE collects and analyzes data about input markets, such as fertilizers, seed, and soil correctors. This information is utilized to estimate the area intended for planting, while the expected yield is estimated based on historical yield.

Throughout the growing season, IBGE monitors the production conditions, such as weather and phytosanitary conditions, in order to update the planted area estimates and the yield forecasts. During the harvest stage, the actual data about harvested area and yield are obtained.

The IBGE validates the data before releasing the LSPA reports using both quantitative and qualitative methods. Quantitative validation refers to the process of looking for mistypes and missing data. The qualitative validation is a subjective analysis of the reliability of the estimates. The IBGE analyzes the occurrence of outliers by comparing the data collected in the previous season (for the same month) with the current data. The forms collected from the local groups (GCEA, COMEA, and COREA) are analyzed again for any explanation, in case there is some outlier.

2.2.2 Produção Agrícola Municipal (PAM)

The PAM is a report issued annually by IBGE containing estimates of planted area, harvested area, production, yield, and value of production for a number of agricultural crops. These estimates are separated by city, microregions, mesoregions, states, large regions, and country. This report is quite similar to the LSPA, distinguishable only by its periodicity.

In fact, for crops included in the LSPA, the PAM report simply aggregates the monthly data collected in that source. Soybeans are included in this category. In the case of crops monitored only by the PAM, the methodology for collecting information is similar to the LSPA's methodology. Questionnaires are applied by IBGE's field employees to growers, associations, stakeholders, etc. The collected data is validated similarly to the LSPA report (IBGE, 2018).

2.3 Companhia Nacional de Abastecimento (CONAB)

CONAB is a government organization commissioned to apply public policies related to warehousing services to small and mid-sized farms, and supply of agricultural products to low income families. As part of their mission, CONAB produces reports with estimates related to the agriculture sector, including statistics about soybean production. Their

report is called Acompanhamento da Safra Brasileira de Grãos (ASBG) .

The ASBG is a report issued monthly from October to September. Similar to the LSPA publication, the ASBG uses a quantitative and a qualitative analysis to produce production, area, and yield forecasts for several agricultural crops. The quantitative analysis involves statistical models and confidence intervals (CONAB, 2015). The specific statistical models used in this publication are not explicitly described in the methodology section of the report.

The qualitative analysis is made by interviewing agents in the supply chain, such as growers, cooperative staff, consulting and extension companies, local government agencies, and input resellers. First, the states are divided into large regions and then CONAB interview the subjects according to a stratified sampling method. The subjective information collect in the interviews refer to municipality level, or a group of municipalities (CONAB, 2015).

The ASBG report usually contains the estimates and forecasts of production, area, and yield for several crops along with a market analysis of each crop. This analysis involves explanations about variations in the estimates from one year to another related to weather conditions, diseases, insects and so on, as well as an assessment of marketing conditions, such as price, credit markets, and production costs.

2.4 United States Department of Agriculture (USDA)

USDA's National Agricultural Statistics Service (NASS) is the branch responsible for producing information about the USA's crop production, price paid and received by farmers, production factors such as labor and chemical usage, and variations in the demographics of rural areas. One of the forecasts published by the NASS each month is the yield of several crops, such as corn, soybeans, cotton, wheat, and potatoes (USDA, 2012).

The NASS forecasts and estimates the area to be harvested and yield per area. Es-

sentially, the crop production cycle is divided into three periods: (1) the beginning of the cycle, when the planted area is estimated, (2) during the cycle, when the area to be harvested and yield are forecasted, and (3) after the harvest, when the final harvested area and yield are estimated. In order to assess information about these variables, the NASS performs a subjective type of survey called the Agricultural Yield Survey, and an objective type of survey called the Objective Yield Survey, consisting of measurements and counts in random plots (USDA, 2012).

The methodological report published by USDA (2012) thoroughly describes the data collection and modeling of these two types of surveys. In summary, the following deterministic equation is used to compute the final net yield per acre:

$$Y = (F * W) - L, \tag{2.2}$$

where Y is the population net yield per acre, F is the population average number of fruit per acre, W is the population average net fruit weight per unit, in industry standard moisture, and L is the population average harvest loss per acre (USDA, 2012). The USDA collects a number of measurements to estimate each variate in equation (2.2). For instance, in order to forecast variate F, the following linear regression is utilized:

$$f = \beta_0 + \beta_1 x + \epsilon, \tag{2.3}$$

where f is the estimate of average weight of fruit per unit, β_0 and β_1 are the coefficients of the regression, x is a characteristic of the plant, and ϵ is the residuals. Similar approaches are applied to estimate the other variates in equation (2.2).

The limited collection capacity of primary information through interviews in other countries forces the USDA to base international forecasts on secondary sources, economic models, and environmental models. In addition, the USDA uses satellite images, weather analysis, expert reports, and reports of private and public companies as international

sources of information (USDA, 1999).

Another source of agricultural forecasts is the USDA's World Agricultural Supply and Demand Estimates (WASDE) report. It is a monthly publication that includes forecasts of national and international wheat, rice, coarse grain, oilseeds, and cotton production. This report is prepared by nine Interagency Commodity Estimates Committees (ICECs) formed by representatives of several USDA agencies. The ICECs are managed by the USDA World Agricultural Outlook Board (WAOB) analysts (USDA, 2006).

In order to produce the WASDE report, the ICECs compile and analyze information from: (1) the Foreign Agricultural Service (FAS) regarding international commodity markets, (2) the Economic Research Service (ERS) regarding national and international regional assessments, (3) the National Agricultural Statistics Service (NASS) regarding national crop and livestock estimates, and (4) the Farm Services Agency and the Agricultural Marketing Service regarding domestic policies and markets.

The WASDE report also contains an evaluation of the forecasts issued by this publication. This evaluation is in the form of the difference between projected and final estimates of production, exports, domestic use, and ending stocks.

2.5 Summary

The methodologies reviewed in this chapter and applied to produce deterministic forecasts of agricultural yield by the different government organizations have a subjective and a quantitative components. The subjective forecasts collected from growers and other agents in the supply chain provide a local and updated perspective of the agricultural production. However, these assessments are complemented with a statistical analysis involving information about factors that may impact agriculture, such as weather, diseases, and markets.

The LSPA, PAM, and ASBG reports issued by Brazilian government agencies do not

contain any assessments of uncertainty associated with their estimates and forecasts, while the WASDE reports issued by the USDA contains some statistics about the difference between the expected and actual values of production. In order to take advantage of random sampling in many statistical models, these methodologies allocate part of their resources in sampling methods. Some sampling methods involve more sophisticated technologies such as satellite images analysis and stratification of regions.

3. Bayesian Forecasting Framework

This chapter describes the Bayesian theory of probabilistic forecasting. First, the Bayesian framework is discussed, followed by the Bayesian Processor of Forecasts (BPF) theory, and the methodologies to construct Bayesian Meta-Gaussian models. This general framework will be used more specifically in the next chapters to construct Bayesian forecasting models that will output probabilistic forecasts of the yield.

3.1 Introduction

The Bayesian forecasting model developed here expands a well known Bayesian theory of probabilistic forecasting (e.g., Krzysztofowicz and Reese (1991); Kelly and Krzysztofowicz (1995); Kelly and Krzysztofowicz (1997); Krzysztofowicz (1999); Krzysztofowicz and Kelly (2000); Krzysztofowicz (2001); Maranzano (2006); Krzysztofowicz and Evans (2008); Maranzano and Krzysztofowicz (2008); Krzysztofowicz (2014)). This is the foundation for the models developed later in this research.

The goal of developing these models is to produce probabilistic forecasts of agricultural yield. The Bayesian framework is particularly convenient to address this problem since it allows the use of information from different sources. This chapter expands the BPF theory by using prior distribution functions assessed judgmentally by farmers to produce a probabilistic forecast.

In addition, the models developed in this chapter allow forecasting at different lead times. The same framework can be applied to other agricultural products, using different predictors.

3.1.1 Variates

Let the continuous variate W with sample space $\mathcal{W}$ be the *predictand*, i.e., a variate of interest to the forecast problem. Its realization is denoted $w \in \mathcal{W}$. Let the continuous variate X with sample space $\mathcal{X}$ be the *predictor*, i.e., a variate that potentially reduces the uncertainty of the forecast problem. Its realization is $x \in \mathcal{X}$.

This notation is expanded using subscripts and superscripts in the following chapters to define different predictands and predictors. The term predictand is more common in Bayesian statistics, while it is often referred in conventional statistical literature as response or dependent variable. The predictors are often called independent variates (James et al., 2017).

3.1.2 Bayesian Forecaster

The inputs to the model are a prior density function and a family of likelihood functions (Krzysztofowicz, 1999). Let function $g(w) = p(w)$ for all $w \in \mathcal{W}$ be the prior density function of W, where p is a generic density function. Function $G(w)$ is the prior distribution function of W.

The functions $f(x|w) = p(x|W = w)$ for all $x \in \mathcal{X}$ and $w \in \mathcal{W}$ constitute the family of likelihood functions. Specifically:

- $f(\cdot|w)$ for a fixed $w \in \mathcal{W}$ is the density function of X, conditional on the hypothesis that the actual realization of the *predictand* is $W = w$.

- $f(x|\cdot)$ for a fixed $x \in \mathcal{X}$ is the likelihood function of W, conditional on the realization of the *predictor* $X = x$.

The outputs from the model are an expected density function of the predictor and a family of posterior density functions (Krzysztofowicz, 1999). Let function $\kappa(x) = p(x)$ be the expected density function of X, for all $x \in \mathcal{X}$. The functions $\phi(w|x) = p(w|X = x)$

for all $x \in \mathcal{X}$ and $w \in \mathcal{W}$ constitute the family of posterior density functions of predictand W. Specifically:

- $\phi(\cdot|x)$ is the posterior density function of predictand W, conditional on a realization of the predictor $X = x$.

The Bayesian forecaster can be developed by applying the total probability law, for all $x \in \mathcal{X}$, to obtain the expected density function of X:

$$\kappa(x) = \int_{\mathcal{W}} f(x|w)g(w)dw. \tag{3.1}$$

The family of posterior density functions of W is derived by applying Bayes Theorem for all $w \in \mathcal{W}$:

$$\phi(w|x) = \frac{f(x|w)g(w)}{\kappa(x)}. \tag{3.2}$$

The posterior density function $\phi(\cdot|x)$ is conditional on the realization $x \in \mathcal{X}$ of X. It quantifies the uncertainty about predictand W, given realization $X = x$. It also constitutes a probabilistic forecast of W for a given realization of $X = x$.

The applications of the Bayesian forecaster are diverse. Krzysztofowicz (2001) discusses the risks of using deterministic forecasts in hydrological forecasting systems. Among these risks is the illusion of certainty that comes from the omission of information about the uncertainty associated with deterministic forecasts. On the other hand, probabilistic forecasts, such as the ones produced by the Bayesian forecaster, express the degree of certitude related to the forecasts. In addition, probabilistic forecasts can be used in accordance with decision analysis in order to make rational decisions. Other applications in hydrology were developed by Krzysztofowicz and Reese (1991), Kelly and Krzysztofowicz (1997), Krzysztofowicz (1999), and Krzysztofowicz and Evans (2008).

Maranzano and Krzysztofowicz (2008) used a Bayesian forecaster constructed for a binary *predictand* to forecast post-flight damage in one of the components of the shuttle

Challenger, called O-ring, in 1986. This study reanalyzes the prediction process of a failure event that happened before the space shuttle Challenger accident and suggests a new approach to produce probabilistic forecasts using prior probabilities of damage in the O-ring assessed judgmentally by experts. In this topic, one of the advantages of using the Bayesian forecaster is the quantification of uncertainty about the experts' assessments.

The challenge in using this methodology lies in modeling the distribution functions and analyzing the stochastic dependence between the variates. In this study, the Bayesian forecaster will be used to produce probabilistic forecasts of agricultural yield using prior distribution constructed from experts' assessments.

3.1.3 Prior Information

The conventional Bayesian Processor of Forecast (BPF) is constructed using a historical prior distribution function. This function is modeled using recorded data of W. This research expands the theory by applying judgmental prior distribution functions from farmers to the construction of the BPF. The framework to model judgmental prior distribution functions is describe in chapter 4.

The judgmental prior distribution function quantifies the uncertainty of W from a farmer's perspective. This research uses these functions as a way to incorporate information assessed from experts into the Bayesian forecaster. Currently, subjective forecasts obtained from experts are interpreted by analysts also in a subjective manner or they are incorporated into deterministic statistical models.

Let s_t denote the state of the agricultural system in year t at the time the farmer assesses his judgmental prior distribution function, for t = 1, 2, ..., T. Let $\mathcal{S}$ be the sample space of state such that $s_t \in \mathcal{S}$. The state of the system in a particular field or region may depend upon factors such as weather, soil conditions, disease incidence, and so on. Let $w(t)$ be the observed yield in year t.

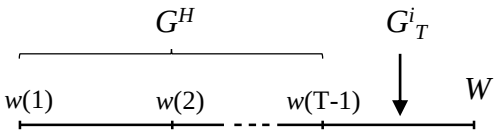


Figure 3.1 Timeline of the modeling of the prior distribution functions.

In summary, the two types of prior distributions involved are:

1. The *historical prior distribution function* G^H modeled using observed values of yield $\{w(t) : t = 1, ..., T-1\}$. This distribution is stationary, since it is independent of the current state of the system s_T. The superscript H will be omitted for simplicity, so that $G = G^H$.

2. The *judgmental prior distribution function* G^i_t, for $i = F, R$, modeled from farmer's assessments and discussed in chapters 4 and 5. This distribution is assessed once a year and it is conditional on s_t:

$$G^i_t(w) = G^i(w|s_t). \tag{3.3}$$

   Therefore, this distribution is nonstationary. Figure 3.1 shows the timeline considered to estimate G^H and G^i_T.

This research assessed the judgmental prior distribution G^i_T for year $T = 2018$. The next subsection describes the theory of the Bayesian Processor of Forecasts. The conventional BPF uses the historical prior distribution function G, and the expanded version of the BPF uses both G and the judgmental prior distribution function G^i_T.

3.2 Bayesian Processor of Forecasts

The formulation of the BPF was presented by Krzysztofowicz (1999) under the section called Bayesian Post-Processing. This section describes the Bayesian meta-Gaussian model using one predictor and multiple predictors. This framework is applied in chapter 6 to construct the Bayesian forecasters for the yield of regions in Mato Grosso, Brazil.

3.2.1 Purpose and Structure

Figure 3.2 shows the principle of the BPF developed in this research. The realization of X is a deterministic forecast of W with a fixed lead time. The prior distribution function quantifies the uncertainty about W from observed data. The BPF merges this quantification with the assessment of uncertainty regarding X. The BPF has been successfully applied in meteorology and hydrology (e.g., Krzysztofowicz and Reese (1991); Krzysztofowicz (1999); Krzysztofowicz and Evans (2008)).

Krzysztofowicz and Evans (2008) designed a meta-Gaussian BPF to assess the uncertainty related to deterministic forecasts produced by the National Weather Service (NWS). These deterministic forecasts were produced judgmentally by human forecasters with the help of a software systems and a subjective analyses of information from multiple sources. The meta-Gaussian BPF was able to successfully produce probabilistic forecasts of daily maximum temperature.

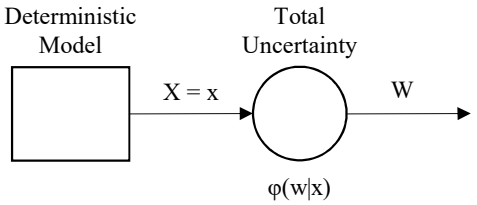


Figure 3.2 Scheme of the Bayesian Processor of Forecasts.
Source: Krzysztofowicz (1999)

3.2.2 Information Fusion

Frequently, data from multiple sources must be analyzed simultaneously to solve a problem or make a decision. Analysts often provide their subjective assessments by analyzing multiple sources one by one. However, the formal framework to systematically combine information from different sources in a useful way is referred as information fusion.

In this study, Bayes theorem expressed in equation (3.2) is used to fuse information about the predictand issued by IBGE, and information about the predictive performance of deterministic forecasts issued by the IBGE and CONAB. According to Krzysztofowicz and Evans (2008), a proper application of Bayes theorem can solve problems in merging information, such as when a joint sample of X and W has smaller size than the prior sample of W.

In conventional statistical analysis, disparities in the size of the joint and prior sample may result in the discard of data. In other words, only the joint sample is utilized. Therefore, the Bayesian approach provides an advantage related to the efficiency in using data. This advantage might turn out to be useful in cases with small sample sizes.

The following methodology focuses on the implementation of the BPF. The Bayesian meta-Gaussian models using one predictor and using multiple predictors are described in order to construct the BPF.

3.2.3 Bayesian Meta-Gaussian Model Using One Predictor

In order to construct a BPF for agricultural yield, this research examines the Bayesian meta-Gaussian Forecaster (BMGF) for its convenient properties. The BMGF has been developed by Kelly and Krzysztofowicz (1997), and extensively applied in research by Maranzano (2006), and Krzysztofowicz and Evans (2008).

This general framework is described here for the BMGF using one predictor, and using multiple predictors in the following sub-section. Chapter 6 uses this framework

to construct BPF models for common forecasting problems in agriculture. The theory described is based on the research from Maranzano (2006).

The construction and estimation of parameters of the Bayesian meta-Gaussian Forecaster involves modeling in the original space of variates X and W, modeling in a transformed space through the use of the Normal Quantile Transformation (NQT), and finally return to the original space.

The prior distribution function G, and the initial estimate $\bar{K}$ of the marginal distribution function of X, are modeled using the methodology described in Appendix B. Once these functions are modeled, it is possible to transform the variates using the Normal Quantile Transformation (NQT),

$$V = Q^{-1}(G(W)), \tag{3.4a}$$

$$Z = Q^{-1}(\bar{K}(X)), \tag{3.4b}$$

where Q^{-1} is inverse of the standard normal distribution function.

Once the variates are transformed, it is possible to model the family of likelihood functions. The model assumes that the stochastic dependence between V and Z is normal-linear. Therefore, the likelihood parameters can be estimated by the relationship

$$Z = aV + b + \Xi, \tag{3.5}$$

where a and b are unknown parameters, and Ξ is the residual from the linear regression. The conditional mean and variance are

$$E(Z|v) = av + b, \tag{3.6a}$$

$$Var(Z|v) = \sigma^2. \tag{3.6b}$$

The parameters $a, b,$ and σ^2 should be estimated using the method of maximum likelihood (ML), which for a and b is equivalent to the method of least squares. The estimation of these parameters uses the historical joint sample of (X, W). The ML estimator of σ^2 is

$$\bar{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}(\Xi_n - \bar{\mu}_\Xi)^2, \tag{3.7}$$

where the ML estimator of the mean μ_Ξ, is

$$\bar{\mu}_\Xi = \frac{1}{N}\sum_{n=1}^{N}\Xi_n. \tag{3.8}$$

The assumptions regarding the linearity of the stochastic dependence between V and Z, homoscedasticity, and normality of the residuals must be validated here. The first step to validate these assumptions is to plot the sample $\{(z, v)\}$ and visually analyze the dependence structure between these two variates. The second step is to analyze the residuals.

Given a satisfactory validation of these assumptions, the coefficients from the linear regression can be used in the conditional density functions $f_Q(z|v)$ as follows.

$$f_Q(z|v) = \frac{1}{\sigma}q\left(\frac{z - av - b}{\sigma}\right), \tag{3.9}$$

where q is the standard normal density function.

According to Krzysztofowicz (1999), the normal-linear likelihood function allows the complete description of the predictive capability of the predictor in terms of three parameters: a, b, and σ.

The posterior density function of W, conditional on $X = x$, is

$$\phi(w|x) = \frac{g(w)}{Tq(Q^{-1}(G(w)))}q\left(\frac{Q^{-1}(G(w)) - c_1Q^{-1}(\bar{K}(x)) - c_0}{T}\right), \tag{3.10}$$

and the posterior distribution function of W, conditional on $X = x$, is

$$\Phi(w|x) = Q\left(\frac{Q^{-1}(G(w)) - c_1 Q^{-1}(\bar{K}(x)) - c_0}{T}\right), \tag{3.11}$$

where Q is the standard normal distribution function and c_1, c_0, and T^2 are the posterior parameters that are calculated as follows:

$$c_1 = \frac{a}{a^2 + \sigma^2}, \tag{3.12a}$$

$$c_0 = \frac{-ab}{a^2 + \sigma^2}, \tag{3.12b}$$

$$T = \left(\frac{\sigma^2}{a^2 + \sigma^2}\right)^{1/2}. \tag{3.12c}$$

The posterior quantile function of W, conditional on a realization $X = x$, can be found by inverting the posterior distribution in equation (3.11) as follows.

$$w(p|x) = \Phi^{-1}(p|x) = G^{-1}(Q(c_1 Q^{-1}(\bar{K}(x)) + c_0 + TQ^{-1}(p))). \tag{3.13}$$

for any p $(0 < p < 1)$, where G^{-1} denotes the inverse of the function G. For $p = 0.5$, equation (3.13) produces the Bayesian meta-Gaussian median posterior regression. A 90% central credible interval around the median posterior regression can be computed using $w(0.05|\cdot)$ and $w(0.95|\cdot)$.

3.2.4 Bayesian Meta-Gaussian Model Using Multiple Predictors

Similarly to the previous section, this framework has been developed by Kelly and Krzysztofowicz (1997), Maranzano (2006), and Krzysztofowicz and Evans (2008). Let X_l be the predictor, with sample space $\mathcal{X}_l$, and realization $x_l \in \mathcal{X}_l$, for $l = 1, ..., L$. Let $\boldsymbol{X}$ be the vector of predictors. In summary, the following samples are used in this model:

Historical prior sample of W – $\{w(n) : n = 1, ..., N-1\}$, (3.14a)

Marginal sample of X_l – $\{x_l(n) : n = 1, ..., N_l - 1\}, \quad l = 1, ..., L,$ (3.14b)

Joint sample of X_l and W – $\{(x_l(n), w(n)) : n = 1, ..., N' - 1\}, \quad l = 1, ..., L,$ (3.14c)

where $N_l - 1$ is the sample size of predictor X_l. The sizes of the historical sample of W and the marginal samples of X_l do not have to be equal. In fact, due to the nature of the problem, it is common to have the sample size of the W larger than the sample sizes of X_l. In addition, the sample sizes of X_l do not necessarily have to be equal. The size of the joint sample of X_l and W is bounded by the smallest sample size of X_l. Therefore, it is possible that $N \geq N'$, and $N_l \geq N'$.

Let $\bar{K}_l$ denote the initial estimate of the marginal distribution function of X_l, obtained from the marginal sample. The variates $X_1, ..., X_L, W$ are transformed using the Normal Quantile Transformation (NQT) as follows:

$$V = Q^{-1}(G(W)), \tag{3.15a}$$

$$Z_l = Q^{-1}(\bar{K}_l(X_l)), \quad \text{for } l = 1, ..., L. \tag{3.15b}$$

Each variate is assumed to be normally distributed: variate V with mean μ_0 and variance σ_0^2, and variate Z_l with mean μ_l and variance σ_l^2. Consider the two L-dimensional column vectors

$$\boldsymbol{\mu} = (\mu_1, ..., \mu_L), \tag{3.16a}$$

$$\boldsymbol{\sigma} = (\sigma_{10}, ..., \sigma_{L0}). \tag{3.16b}$$

The stochastic dependence between the variates in the transformed space is assumed to be from the multivariate Normal family. Considering the vector $(Z_1, ..., Z_L, V)$ to be

$\text{MVN}(\boldsymbol{\mu}_Q, \boldsymbol{\Sigma}_Q)$, where

$$\boldsymbol{\mu}_Q = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_L \\ \mu_0 \end{bmatrix}, \quad \boldsymbol{\Sigma}_Q = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1L} & \sigma_{10} \\ \vdots & & \vdots & \vdots \\ \sigma_{1L} & \cdots & \sigma_L^2 & \sigma_{L0} \\ \sigma_{10} & \cdots & \sigma_{L0} & \sigma_0^2 \end{bmatrix}. \tag{3.17}$$

Consequently, the multivariate density function $f_Q(\cdot|v)$ is also $\text{MVN}(\boldsymbol{\mu}_{f_Q}, \boldsymbol{\Sigma}_{f_Q})$, where

$$\boldsymbol{\mu}_{f_Q} = \begin{bmatrix} \mu_1 + \frac{\sigma_{10}}{\sigma_0^2}(v - \mu_0) \\ \vdots \\ \mu_L + \frac{\sigma_{L0}}{\sigma_0^2}(v - \mu_0) \end{bmatrix}, \quad \boldsymbol{\Sigma}_{f_Q} = \begin{bmatrix} \sigma_1^2 - \frac{\sigma_{10}\sigma_{10}}{\sigma_0^2} & \cdots & \sigma_{1L} - \frac{\sigma_{10}\sigma_{L0}}{\sigma_0^2} \\ \vdots & & \vdots \\ \sigma_{1L} - \frac{\sigma_{10}\sigma_{L0}}{\sigma_0^2} & \cdots & \sigma_L^2 - \frac{\sigma_{L0}\sigma_{L0}}{\sigma_0^2} \end{bmatrix}. \tag{3.18}$$

According to Maranzano (2006), the posterior density function ϕ_Q in the transformed space takes the form:

$$\phi_Q(v|\boldsymbol{z}) = \frac{1}{T} q\left(\frac{v - \Sigma_{l=1}^L c_l z_l - c_0}{T}\right). \tag{3.19}$$

The posterior distribution function Φ_Q in the transformed space takes the form:

$$\Phi_Q(v|\boldsymbol{z}) = Q\left(\frac{v - \Sigma_{l=1}^L c_l z_l - c_0}{T}\right), \tag{3.20}$$

where

$$\boldsymbol{c}^T = \frac{T^2}{\sigma_0^2} \boldsymbol{\sigma}^T \boldsymbol{\Sigma}_{f_Q}^{-1}, \tag{3.21a}$$

$$c_0 = \boldsymbol{c}^T \left(\frac{\mu_0}{\sigma_0^2}\boldsymbol{\sigma} - \boldsymbol{\mu}\right), \tag{3.21b}$$

$$T = \left(\frac{\sigma_0^4}{\boldsymbol{\sigma}^T \boldsymbol{\Sigma}_{f_Q}^{-1} \boldsymbol{\sigma} + \sigma_0^4}\right)^{\frac{1}{2}}. \tag{3.21c}$$

and $\boldsymbol{c}^T = [c_1, ..., c_L]$ is an L-dimensional row vector (the transpose of $\boldsymbol{c}$).

The expressions for the conditional density function f and posterior density function ϕ in the original space can be derived from the relationship originated from the NQT in equations (3.15a) and (3.15b). The density function $f(\boldsymbol{x}|w)$ of $\boldsymbol{X}$, conditional on the hypothesis that $W = w$ has the expression

$$f(\boldsymbol{x}|w) = \frac{1}{(2\pi)^{L/2}|\boldsymbol{\Sigma}_{f_Q}|^{1/2}} \prod_{l=1}^{L} \frac{\bar{\kappa}_l(x_l)}{q(Q^{-1}(\bar{K}_l(x_l)))} e^{-(\boldsymbol{z}-\boldsymbol{\mu}_{f_Q})'\boldsymbol{\Sigma}_{f_Q}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_{f_Q})/2}. \tag{3.22}$$

The posterior density function of W, conditional on the realization of the vector of predictors $\boldsymbol{X} = \boldsymbol{x}$, is

$$\phi(w|\boldsymbol{x}) = \frac{g(w)}{Tq(Q^{-1}(G(w)))} q\left(\frac{Q^{-1}(G(w)) - \Sigma_{l=1}^{L} c_l Q^{-1}(\bar{K}_l(x_l)) - c_0}{T}\right). \tag{3.23}$$

The posterior distribution function of W, conditional on the realization of the vector of predictors $\boldsymbol{X} = \boldsymbol{x}$, is

$$\Phi(w|\boldsymbol{x}) = Q\left(\frac{Q^{-1}(G(w)) - \Sigma_{l=1}^{L} c_l Q^{-1}(\bar{K}_l(x_l)) - c_0}{T}\right). \tag{3.24}$$

The posterior quantile function of W is

$$w(p|\boldsymbol{x}) = \Phi^{-1}(p|\boldsymbol{x}) = G^{-1}(Q(\sum_{l=1}^{L} c_l Q^{-1}(\bar{K}_l(x_l)) + c_0 + TQ^{-1}(p))), \tag{3.25}$$

for p such that $0 < p < 1$. The posterior quantile function in equation (3.25) can be used to produce a 90% central credible interval around the median posterior regression using $w(0.05|\cdot)$ as a lower bound and $w(0.95|\cdot)$ as an upper bound.

The posterior density, distribution, and quantile functions produce probabilistic forecasts of W, conditional on the realization of the deterministic forecast $X_l = x_l$ for $l = 1, ..., L$. This systematical approach can be used consistently with a decision model,

as opposed to subjectively determining which deterministic forecast is more reliable.

3.3 Bayesian Processor of Forecasts Using Judgmental Prior

This section develops the framework to use judgmental prior distribution functions in the Bayesian Processor of Forecasts. This framework also provides the tools to analyze whether the farmers are able to contribute with insightful information about future yields. The judgmental prior distribution functions quantify the uncertainty related to W under the expert's perspective. The following subsections adapt the Bayesian meta-Gaussian models for one and multiple predictors using this new rationale.

3.3.1 Bayesian Meta-Gaussian Model Using One Predictor

Modeling the Bayesian meta-Gaussian forecaster using the judgmental prior distribution is similar to the framework discussed in subsection 3.2.3 up to the derivation of the posterior distribution and density functions. This model uses both the historical and judgmental prior distribution functions.

Figure 3.3 shows a scheme of the BPF using judgmental prior distributions. This model expands the conventional BPF presented in figure 3.2 by quantifying the uncertainty about W for the current year from the farmer's perspective. The posterior parameters are obtained using the historical prior distribution function, but the posterior functions are constructed using the judgmental prior distribution function.

The variates are transformed using the NQT through the historical prior distribution function $G = G^H$ and the initial estimate $\bar{K}$:

$$V = Q^{-1}(G(W)), \tag{3.26a}$$

$$Z = Q^{-1}(\bar{K}(X)). \tag{3.26b}$$

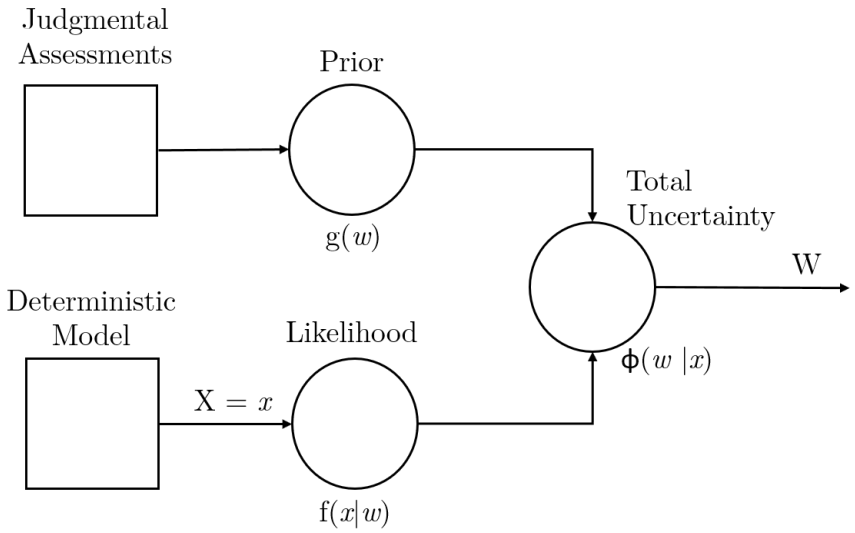


Figure 3.3 Scheme of the Bayesian Processor of Forecasts using judgmental priors.
Source: Adapted from Krzysztofowicz (1999)

Similar to subsection 3.2.3, the stochastic dependence between V and Z is assumed to be normal-linear and the likelihood parameters can be estimated using the relationship

$$Z = aV + b + \Xi. \tag{3.27}$$

The conditional mean and variance are

$$E(Z|v) = av + b, \tag{3.28a}$$

$$Var(Z|v) = \sigma^2. \tag{3.28b}$$

The conditional density functions $f_Q(z|v)$ can be derived as follows, given that the assumptions related to the stochastic dependence between V and Z have been validated:

$$f_Q(z|v) = \frac{1}{\sigma} q\left(\frac{z - av - b}{\sigma}\right). \tag{3.29}$$

Thus far, the modeling steps have not been modified. The posterior distribution and

density functions are constructed using a judgmental prior distribution. However, some considerations must be made in order to use function G_t^i.

According to equation (3.1), a new prior density function of W results in a new expected marginal density function of X, $\kappa(x)$. Let $\bar{K}$ be the initial estimate of the marginal distribution function of X. Then, there exists $w \in \mathcal{W}$ and $x \in \mathcal{X}$ such that

$$G(w) = \bar{K}(x). \tag{3.30}$$

Solving equation (3.30) for w, we have

$$w = G^{-1}(\bar{K}(x)). \tag{3.31}$$

Theoretically, using the same approach as in equation (3.30), the new expected density function $\bar{K}_t^i(x)$ of X, which corresponds to a judgmental prior distribution function G_t^i, can be obtained as follows:

$$\begin{aligned} \bar{K}_{t,l}^i(x) &= G_t^i(w) \\ &= G_t^i(G^{-1}(\bar{K}(x))). \end{aligned} \tag{3.32}$$

Finally, the posterior density function of W, conditional on $X = x$, is

$$\phi(w|x) = \frac{g_t^i(w)}{Tq(Q^{-1}(G_t^i(w)))} q\left(\frac{Q^{-1}(G_t^i(w)) - c_1 Q^{-1}(\bar{K}_t^i(x)) - c_0}{T}\right), \tag{3.33}$$

and the posterior distribution function of W, conditional on $X = x$, is

$$\Phi(w|x) = Q\left(\frac{Q^{-1}(G_t^i(w)) - c_1 Q^{-1}(\bar{K}_t^i(x)) - c_0}{T}\right), \tag{3.34}$$

where c_1, c_0, and T^2 are the posterior parameters that are calculated as follows:

$$c_1 = \frac{a}{a^2 + \sigma^2}, \tag{3.35a}$$

$$c_0 = \frac{-ab}{a^2 + \sigma^2}, \tag{3.35b}$$

$$T = \left(\frac{\sigma^2}{a^2 + \sigma^2}\right)^{1/2}. \tag{3.35c}$$

The posterior quantile function of W, conditional on a realization $X = x$, can be found by inverting the posterior distribution in equation (3.34) as follows.

$$w(p|x) = \Phi^{-1}(p|x) = G_t^{i(-1)}(Q(c_1 Q^{-1}(\bar{K}_t^i(x)) + c_0 + TQ^{-1}(p))). \tag{3.36}$$

for any p $(0 < p < 1)$, where $G_t^{i(-1)}$ denotes the inverse of the function G_t^i. For $p = 0.5$, equation (3.36) produces the Bayesian meta-Gaussian median posterior regression. A 90% central credible interval around the median posterior regression can be computed using $w(0.05|\cdot)$ and $w(0.95|\cdot)$.

The BPF constructed in this section: (1) produces probabilistic forecast of W, conditional on the realization of $X = x$, through the posterior distribution function $\Phi(w|x)$, (2) quantifies the uncertainty associated with the deterministic forecasts of yield, and (3) incorporates the judgmental prior distribution function assessed from a farmer.

3.3.2 Bayesian Meta-Gaussian Model Using Multiple Predictors

Similarly to subsection 3.2.4, this subsection describes BMGF for multiple predictors, but using the judgmental prior distribution function. Let $\bar{K}_l$ denote the initial estimate of the expected marginal distribution function of X_l, obtained from the marginal sample. The variates $W, X_1, ..., X_L$ are transformed using the Normal Quantile Transformation (NQT)

as follows:

$$V = Q^{-1}(G(W)), \tag{3.37a}$$

$$Z_l = Q^{-1}(\bar{K}_l(X_l)), \quad \text{for } l = 1, ..., L. \tag{3.37b}$$

The posterior parameters $T, \boldsymbol{c}^L$, and c_0 can be estimated similarly to subsection 3.2.4, using equations (3.16)-(3.21). However, the new posterior density and distribution functions use the judgmental prior distribution function in this framework. In the following equations (3.39), (3.40), and (3.41), the judgmental prior distribution function G_t^i and its inverse $G_t^{i(-1)}$ are applied. In parallel to equation (3.32), the new expected density function $\bar{K}_{t,l}^i$ resulting from using the nonstationary prior distribution function G_t^i can be obtained as follows:

$$\begin{aligned} \bar{K}_{t,l}^i(x_l) &= G_t^i(w) \\ &= G_t^i(G^{-1}(\bar{K}_l(x_l))), \quad \text{for } l = 1, ..., L. \end{aligned} \tag{3.38}$$

The posterior density function of W, conditional on the realization of the vector of predictors $\boldsymbol{X} = \boldsymbol{x}$, is

$$\phi(w|\boldsymbol{x}) = \frac{g_t^i(w)}{Tq(Q^{-1}(G_t^i(w)))} q\left(\frac{Q^{-1}(G_t^i(w)) - \Sigma_{l=1}^L c_l Q^{-1}(\bar{K}_{t,l}^i(x_l)) - c_0}{T}\right). \tag{3.39}$$

The posterior distribution function of W, conditional on the realization of the vector of predictors $\boldsymbol{X} = \boldsymbol{x}$, is

$$\Phi(w|\boldsymbol{x}) = Q\left(\frac{Q^{-1}(G_t^i(w)) - \Sigma_{l=1}^L c_l Q^{-1}(\bar{K}_{t,l}^i(x_l)) - c_0}{T}\right). \tag{3.40}$$

Similarly to the Bayesian forecaster using a single predictor, the model developed for multiple predictors incorporates prior distributions assessed judgmentally based on

the premise that experts, such as farmers, posses a greater understanding about the production conditions of their fields and can potentially add insightful information to the forecast. The posterior quantile function of W is

$$w(p|\boldsymbol{x}) = \Phi^{-1}(p|\boldsymbol{x}) = G_t^{i(-1)}(Q(\sum_{l=1}^{L} c_l Q^{-1}(\bar{K}_{t,l}^{i}(x_l)) + c_0 + TQ^{-1}(p))), \tag{3.41}$$

for p such that $0 < p < 1$.

The posterior quantile function in equation (3.41) can be used to produce a 90% central credible interval around the median posterior regression using $w(0.05|\cdot)$ as a lower bound and $w(0.95|\cdot)$ as an upper bound.

3.4 Summary

The framework described and developed in this chapter is capable of producing probabilistic forecasts of agricultural yield. The user is able to quantify uncertainty about the yield, whether using historical data or through an expert perspective, and merge it with the quantification of uncertainty about deterministic forecasts. In addition, the models developed for multiple predictors allow the user to take advantage of several deterministic forecasts.

Modeling the BPF is relatively tractable in the sense that only the judgmental distribution functions must be estimated every season. The analyst can model them and update the posterior functions at any time desired. The probabilistic forecasts released to the decision makers do not have to contain the complexities associated with the structure of the forecaster.

Chapter 4 addresses the problem of judgmentally assessing prior distribution functions from individuals and the problem of combining multiple functions into a single one. This topic is compelling to the successful construction of a BPF model, since it can directly impact the derivation of the posterior functions.

4. Judgmental Distribution Functions

Experts are often summoned in decision analysis to provide insightful information. Through experience, an expert can subjectively forecast a realization of a continuous *predictand* by considering possible scenarios, and the uncertainty related to each of them. This chapter reviews the predominant methods to assess a distribution function judgmentally, and develops an approach to combine assessments from different experts.

4.1 Why Are Farmers Experts?

DeGroot (1988) presents two definitions of expert, each one being on a different extreme. First, DeGroot summarizes the definition of expert presented by other authors (e.g. Morris (1974, 1977)) as "to be anyone or any system that will give you a prediction". Second, he defines an expert as "to be someone whose prediction you will simply adopt as your own posterior probability without modification".

Farmers think about expected production many times before and during the growing season. For instance, deciding the amount of lime to be applied in the soil to correct the pH is a complex task. The optimal quantity will depend on the type of soil of a particular field, the initial pH, the type of crop, and the depth of soil to be neutralized (Buchholz, 2004; Mallarino et al., 2013). However, farmers can consider the trade-off between applying smaller amounts of lime and providing sub-optimal conditions of production. This is one example among many associated with expected production that farmers must considerate.

Returning to the definition of expert by Morris (1977) as "anyone with special knowledge about an uncertain quantity or event", farmers evidently qualify as experts of their own production, or even on a larger scale, of the regional production. They are certainly capable of providing predictions to those who don't participate in the production process.

4.2 Assessing Judgmental Distribution Functions

4.2.1 Overview

The elicitation of judgmental point estimates, probabilities, and distribution functions from experts is subject of a extensive discussion. The effectiveness of judgmental assessment methods may depend on the experts' understanding of the problem or variate of interest, but also on their knowledge about statistical concepts such as probability or quantile. However, there are some advantages of assessing quantiles, like the facility to construct parametric distribution functions.

Alpert and Raiffa (1982) described the method of direct quantile assessments, and tested it among MBA and graduate students at Harvard University between 1968 and 1969. Briefly, this method consists of requesting the participants to estimate median and other quantiles related to some variable. The researcher directly requests the quantiles.

Keeney et al. (1984) applied the method of direct quantile assessments to 14 health specialists working for the United States Environmental Protection Agency (EPA) to assess their judgments about the health risks associated with different levels of CO exposure in the air. This problem was particularly difficult to tackle exclusively with experimental data due to the possibility of unethical procedures. However, Keeney et al. successfully developed a risk assessment model to expose the health risks of several levels of CO in the air caused by pollution.

Winterfeldt and Schweitzer (1998) assessed the quantiles of technical staff members of the United States Department of Energy (DOE), contractors, and consultants regarding a proposed schedule of tritium production for nuclear weapons. At that time, Winterfeldt and Schweitzer described the lack of national tritium production as a vulnerability to the nuclear program and to the national security. Given multiple solutions, the experts provided their estimates about the schedule of each option.

4.2.2 Framework

Let the continuous variate W^i for $i = F, R$ with sample space $\mathcal{W}^i$ be the *predictand*, i.e., a variate of interest to the forecast problem. Its realization is denoted $w^i \in \mathcal{W}^i$, where $\mathcal{W}^i = \{w^i : \eta_L < w^i < \eta_U\}$, where η_L is a lower bound and η_U is an upper bound. The superscripts F and R are the initials for Field and Region, respectively. The following definition is adapted from Krzysztofowicz (2016).

Definition. A *judgmental prior distribution function* G_t^i of the continuous predictand W^i is the numerical measure of the *degree of certainty* about the occurrence of all events $\{W^i \leq w^i : w^i \in \mathcal{W}^i\}$, given information (I), knowledge (K), and experience (E) possessed by the expert at the time of forecast preparation. At every point $w^i \in \mathcal{W}^i$, function G^i specifies the probability assigned by the expert to event $\{W^i \leq w^i\}$:

$$G_t^i(w^i) = P(W^i \leq w^i | I, K, E), \text{ for } i = F, R.$$

According to Krzysztofowicz (2016), assessing the judgmental distribution function from the definition above is unfeasible due to the infinite possible realizations of W^i. However, an approximation to function G_t^i of W^i can be constructed by assessing quantiles from the experts. The following definition was also adapted from to Krzysztofowicz (2016).

Definition. A *judgmental quantile* of W^i corresponding to probability $p \in (0, 1)$ is a realization $w^i(p) \in \mathcal{W}^i$ such that event $\{W^i \leq w^i(p)\}$ is judged to have the probability

$$G(w^i(p)) = P(W^i \leq w^i(p) | I, K, E) = p.$$

The quantiles $w^i(p)$ were assessed for five values of p: 0.1, 0.25, 0.5, 0.75, 0.9. Therefore, a set of points $\{(w^i(p), p)\}$ was obtained for $i = F$ and $i = R$.

4.2.3 Assessment Procedures

The procedures to assess judgmental quantiles from the farmers were developed based on the research by Alpert and Raiffa (1982), and Krzysztofowicz (2016). The goal is to assess values of $w_j^i(p)$ for $i = F, R$, from farmer j and for several predetermined values of p. However, the methodology developed had to be adapted to subjects that are not familiar with statistical concepts, such as quantiles. For that reason, the subjects were divided into two groups, a pilot and a research group. The groups were as follows.

> *Pilot Group*: 1 soybean grower from Rondonópolis with 5,000 hectares of land and 1 soybean grower from São Desidério with 10,000 hectares. Both locations are in the state of Mato Grosso.
>
> *Research Group*: 6 soybean growers in total from Sorriso (550 and 4,000 hectares), Paranantinga (20,000 hectares), Sapezal (1,200 hectares), Diamantino (2,000 hectares), and Canarana (150 hectares), in Mato Grosso.

Alpert and Raiffa (1982) developed the method of direct quantile assessments [1] to quantify the uncertainty of a number of students regarding some general-knowledge variables. They analyzed 4 groups of students enrolled in different programs of Harvard, but with an advanced knowledge in probability theory. This particular aspect of the students allowed Alpert and Raiffa (1982) to directly request quantile values using a survey with 2 pages divided into 4 sections. The first page displayed the instructions, the questions, and a list of uncertainties. The second page was reserved for the student's quantile assessments.

The instructions provided by Alpert and Raiffa (1982) explained the purpose of the survey, mathematical definitions of quantiles, and a guideline to the following sections in their study. The questions section collected some data from the students, such as

[1] *direct fractile assessments* in the original publication.

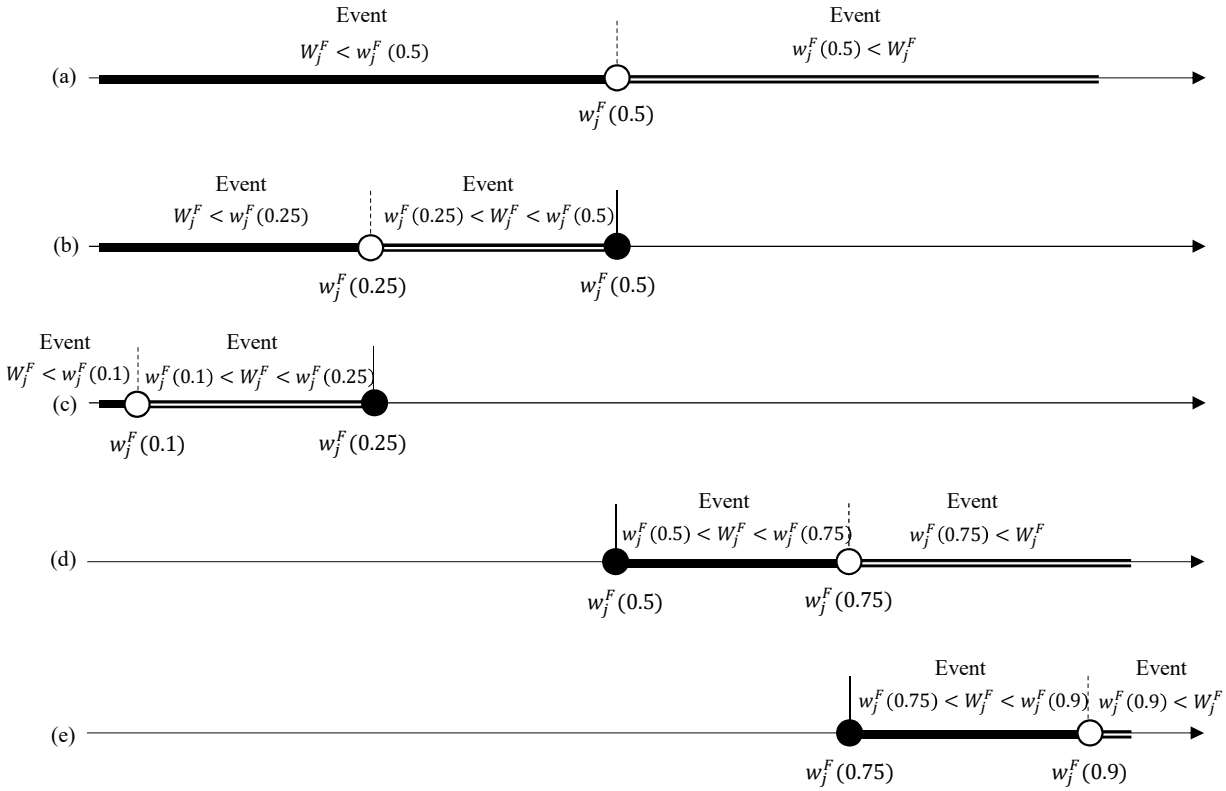


Figure 4.1 Scheme for assessing the judgmental quantiles, $w_j^F(p)$, for: (a) $p = 0.5$, (b) $p = 0.25$, (c) $p = 0.1$, (d) $p = 0.75$, and (e) $p = 0.9$.
Source: Adapted from Krzysztofowicz (2016)

preferences for drinking, draft deferments for graduate students, and gambling. The list of uncertainties contained the variables assessed in the question section and 6 other variables, such as the total egg production in millions in the U.S. in 1965.

The methodology developed by Krzysztofowicz (2016) overcomes the need to directly request quantiles by designing questions using the definition of quantile and a more inclusive language. Figure 4.1 shows the rationale for assessing judgmental quantiles adapted from Krzysztofowicz (2016). Each line in Figure 4.1 represents the assessment of one quantile. The white circles represent a quantile to be assessed and the black circles represent a quantile already assessed.

The assessment of each quantile starts by defining a subset of the domain to be considered by the subject. For instance, the subject must consider the complete sample space

for assessing the median, $w_j^F(0.5)$, in figure 4.1-(a). In order to assess $w_j^F(0.25)$, the subject is asked to consider a subset of the sample space with $W_j^F < w_j^F(0.5)$ as shown in figure 4.1-(b). The same rationale applies to the assessment of the other quantiles.

Krzysztofowicz (2016) proposes next a format in which the subject provides a possible realization of W_j^F, $w_j^F(p)$, for a certain p that in his judgment, the following events are equally likely, given the subset considered:

Realization of W_j^F will be below your estimate: $W_j^F < w_j^F(p)$	Realization of W_j^F will be above your estimate: $w_j^F(p) < W_j^F$

Subsequently, Krzysztofowicz (2016) proposes a validation question to verify if the estimated $w_j^F(p)$ has the same properties as stated by the definition of quantile according to the subject. This step is made by suggesting a reference event that is commonly known by the subject and that has the probability of outcomes equal to $\frac{1}{2}$, such as a coin toss. The subject is then asked to compare the properties of the reference event with the properties of the assessed quantile, in terms of probabilities.

Suitable results using this approach will depend upon the subject's understanding of events that are *equally likely* to occur and the researcher's ability to adapt the method to be as inclusive as possible. Goldstein and Rothschild (2014) compare the assessment of subjective probability using standard (direct) and graphical methods. They point out that using a graphical interface may improve accuracy when assessing subjective probabilities from laypeople. For this reason, the validation question was modified firstly to provide a visual representation of outcomes that are *equally likely* to happen.

Clemen (1996) suggests an approach to assess subjective probability that bypasses the problem of directly requesting a number. In this approach, the subject is asked to pick between two lottery-like games. Each game has the same two outcomes Prize A and Prize B. The researcher must set Prize A to be much more valuable than Price B, so that

A would be preferable to B. The subject must pick one out of the two following games, considering the subset in question for each quantile assessment:

$$\text{Win Prize A if } W_j^F < w_j^F(p).$$
$$\text{Win Prize B if } w_j^F(p) < W_j^F.$$

or

$$\text{Win Prize A with known probability } \tfrac{1}{2}.$$
$$\text{Win Prize B with known probability } \tfrac{1}{2}.$$

This approach was constructed to be applied in the pilot group for validating the assessed $w_j^F(p)$. Clemen (1996) keeps the conditions of the first lottery constant, but changes the probability of Prize A and B in the second game until the subject is indifferent in order to assess these probabilities. However, in this study the values of $w_j^F(p)$ can be reassessed until the subject is indifferent between both games, keeping the same probability $\frac{1}{2}$ for each outcome in the second lottery. The indifference point indicates that the subject consider the events $W_j^F < w_j^F(p)$ and $w_j^F(p) < W_j^F$ to be equally likely.

The approach proposed by Krzysztofowicz (2016) was adapted to this research and applied to the pilot group. Each interviewee indirectly provided his assessment of $w_j^i(p)$ for p: 0.1, 0.25, 0.5, 0.75, 0.9. The soybean growers were previously contacted by phone or email in order to explain the purposes of the research and to provide a brief overview on how the data collected would be used. Additionally, a brief conversation about the current production conditions in Mato Grosso was had with the subjects contacted by phone to engage them in the mindset of the interview. The online survey was applied using Qualtrics — an online platform that can be use to create and apply surveys.

The individual interviews were made available via desktop or mobile devices. The hypothetical prizes presented to the pilot group were a trip to Hawaii (Prize A) and a ticket to the movies (Prize B). The assessment of the quantiles proceeded as follows.

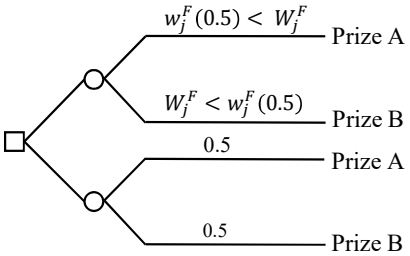
Assessing the 0.5-Quantile, $w_j^F(0.5)$

Step 1. The subject was asked to consider all possible realizations of W_j^F.

Step 2. The subject was then asked to provide an estimate of W_j^F such that in his judgment the following events would be equally likely:

Realization of W_j^F will be below your estimate: $W_j^F < w_j^F(0.5)$	Realization of W_j^F will be above your estimate: $w_j^F(0.5) < W_j^F$

Step 3. The subject was asked to pick one out of two of the following games. The first game would result in Prize A if $w_J^F(0.5) < W_j^F$ or Prize B if $W_j^F < w_J^F(0.5)$. The second game would result in Prize A with probability 0.5 or Prize B with probability 0.5.



Step 4. The assessed value of $w_J^F(0.5)$ is validated when the subject in indifferent between the first game and the second game. The estimate is reassessed if the subject clearly prefers one game over the other.

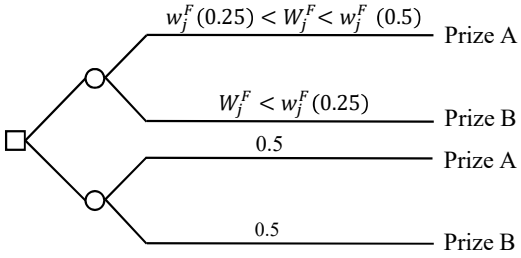
Assessing the 0.25-Quantile, $w_j^F(0.25)$

Step 1. The subject was asked to consider possible realizations of W_j^F below $w_j^F(0.5)$.

Step 2. The subject was then asked to provide an estimate of W_j^F such that in his judgment the following events would be equally likely:

Realization of W_j^F will be below your estimate: $W_j^F < w_j^F(0.25)$	Realization of W_j^F will be above your estimate: $w_j^F(0.25) < W_j^F$

Step 3. The subject was asked to pick one out of two of the following games. The first game would result in Prize A if $w_J^F(0.25) < W_j^F < w_J^F(0.5)$ or Prize B if $W_j^F < w_J^F(0.25)$. The second game would result in Prize A with probability 0.5 or Prize B with probability 0.5.



Step 4. The assessed value of $w_J^F(0.25)$ is validated when the subject in indifferent between the first game and the second game. The estimate is reassessed if the subject clearly prefers one game over the other.

Assessing the 0.01-Quantile, $w_j^F(0.01)$

Step 1. The subject was asked to consider possible realizations of W_j^F below $w_j^F(0.25)$.

Step 2. The subject was then asked to provide an estimate of W_j^F such that in his judgment this estimate would be an extreme under the worst possible production conditions.

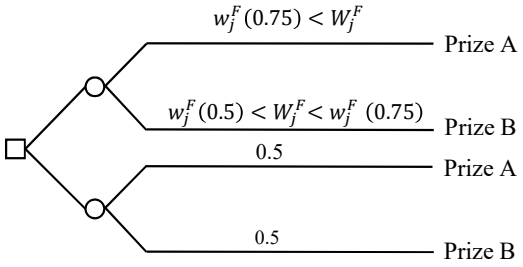
Assessing the 0.75-Quantile, $w_j^F(0.75)$

Step 1. The subject was asked to consider possible realizations of W_j^F above $w_j^F(0.5)$.

Step 2. The subject was then asked to provide an estimate of W_j^F such that in his judgment the following events would be equally likely:

Realization of W_j^F will be below your estimate: $W_j^F < w_j^F(0.75)$	Realization of W_j^F will be above your estimate: $w_j^F(0.75) < W_j^F$

Step 3. The subject was asked to pick one out of two of the following games. The first game would result in Prize A if $w_J^F(0.75) < W_j^F$ or Prize B if $w_J^F(0.5) < W_j^F < w_J^F(0.25)$. The second game would result in Prize A with probability 0.5 or Prize B with probability 0.5.



Step 4. The assessed value of $w_J^F(0.75)$ is validated when the subject in indifferent between the first game and the second game. The estimate is reassessed if the subject clearly prefers one game over the other.

Assessing the 0.9-Quantile, $w_j^F(0.9)$

Step 1. The subject was asked to consider possible realizations of W_j^F above $w_j^F(0.75)$.

Step 2. The subject was then asked to provide an estimate of W_j^F such that in his judgment this estimate would be an extreme under the best possible production conditions.

Two important results came from applying the methodology described above to the pilot group. First, the unit being used to measure yield had to be changed. The official reports by IBGE and CONAB express yield in terms of *tons per hectares*, but many farmers are used to deal with *bags per hectares*. Therefore, the survey had to adopt this unit.

Second, the farmers could not finish the survey. There were complications about understanding the concept of events that are *equally likely* to occur. Farmers may be used to mentally formulate point-estimates of their yield. For that reason, it was possibly difficult for them to think in terms of probability. In addition, the validation question could not be interpreted adequately by them, generating a lot of uncertainty. These complications lead to the adaptation of a survey.

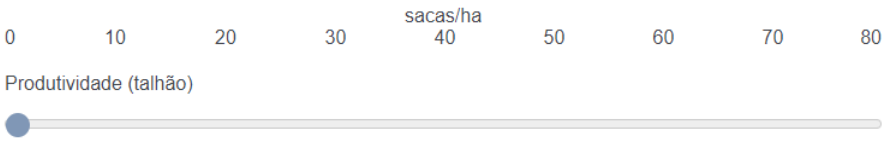
The second round of surveys, applied to the research group, was changed in the following topics.

(a) Units. The yield measurements in the new survey were in terms of *bags per hectare*

(b) Language. The questions were reformulated to circumvent use of the words *equally likely*, and probability.

(c) Validation. The validation was incorporated in the assessment question.

(d) Graphical format. The subject could pick his estimate by sliding a marker through a ruler, using Qualtrics.



The subject was first asked to assess his estimate of yield given current production conditions. Based on this estimate, different scenarios were introduced and the subject was asked to provide a new estimate. Each scenario was linked to a quantile of W_j^F. The new estimate was bounded by the previous answers, i.e., the set of possible values for new scenarios were conditional on their previous assessments. The probability p was connected to the different scenarios and subjectively determined to avoid defining concepts such as quantiles, or probability.

Table 4.1 Judgmental quantiles $w_j^i(p)$ [bags/ha] of W^i assessed by farmers in Mato Grosso, Brazil, in February 2018.

j	City	$w_j^F(0.1)$	$w_j^F(0.25)$	$w_j^F(0.5)$	$w_j^F(0.75)$	$w_j^F(0.9)$	$w_j^R(0.1)$	$w_j^R(0.25)$	$w_j^R(0.5)$	$w_j^R(0.75)$	$w_j^R(0.9)$
1	Sorriso	50.3	60.9	70.5	73.8	76.1	45.2	50.1	60.6	65.0	67.6
2	Sorriso	50.2	58.4	65.6	67.5	70.3	49.6	50.5	55.2	58.0	58.7
3	Paranatinga	44.9	57.5	65.0	69.9	80.0	48.0	50.3	53.4	56.0	57.5
4	Sapezal	30.1	45.5	70.0	75.0	80.0	50.2	55.2	58.0	62.7	67.5
5	Diamantino	34.9	44.0	52.5	60.2	73.9	43.8	45.6	51.5	58.6	66.4
6	Canarana	45.4	52.1	60.6	65.5	72.0	44.1	48.5	54.4	56.5	58.6

The same approach was applied to assess the quantiles $w_j^R(p)$ of W_j^R. Table 4.1 shows the assessed quantiles. Each farmer j provided quantiles $w_j^i(p)$ for $i = F, R$, and $p = 0.1, 0.25, 0.5, 0.75,$ and 0.9. The following sub-sections use these quantiles to model distribution functions.

4.3 Parametric Models

The judgmentally assessed quantiles $w_j^F(p)$ of W_j^F and $w_j^R(p)$ of W_j^R in figure 4.1 are utilized to model two parametric distribution functions per farmer j: $G_{t,j}^F$ of W_j^F and $G_{t,j}^R$ of W_j^R. Therefore, a total of 12 parametric distribution functions are modeled in this sub-section. Finally, the set of distribution functions $G_{t,j}^F$ is combined into one distribution function G^F. Likewise, the set of distribution functions $G_{t,j}^R$ is combined into one distribution function G^R.

The sample space is assumed here to be a bounded interval. Due to the physical nature of the problem, it is impossible to consider negative values of yield, as well as unrealistic high values. Although certain conditions may lead to a very low yield values in extreme cases, the lower and upper bounds of the sample space were selected based on a historical analysis of the yield in the location of each farmer.

The same methodology was applied to construct the parametric models for $G_{t,j}^i$, for $i = F, R$, and $j = 1, ..., 6$. The procedures to construct a parametric model described as follows were based on the methodology developed by Krzysztofowicz (2014, 2016).

Step 1. Construct an empirical distribution of W_j^i using the assessed quantiles $w_j^i(p)$.

Step 2. Hypothesize parametric models of $G^i_{t,j}$. In this step, a catalogue produced by Krzysztofowicz (2014) was used to select parametric models.

Step 3. Estimate the parameters of the models selected.

Step 4. Choose the parametric model and parameter values that minimize the Maximum Absolute Difference between the empirical distribution function (step 1) and the estimated parametric distribution function (steps 2 and 3).

Step 5. Analyze the goodness-of-fit of the parametric model $G^i_{t,j}$.

Four different parametric models for functions G^i_j were estimated using this methodology. First, the Log-Reciprocal Type I–Weibull (LC1–WB) was estimated for $G^F_{t,6}$ and $G^R_{t,3}$. The LC1–WB distribution function can be written as

$$G(w) = 1 - \exp\left[-\left(\frac{y}{\alpha}\right)^{\beta}\right], \tag{4.1}$$

and the LC1–WB density function

$$g(w) = \frac{1}{\eta_U - w}\frac{\beta}{\alpha}\left(\frac{y}{\alpha}\right)^{\beta-1}\exp\left[-\left(\frac{y}{\alpha}\right)^{\beta}\right], \tag{4.2}$$

where

$$y = ln\frac{\eta_U - \eta_L}{\eta_U - w}. \tag{4.3}$$

Second, the Log-Reciprocal Type I–Inverted Weibull (LC1–IW) was estimated for $G^R_{t,5}$. The LC1–IW distribution function can be written as

$$G(w) = \exp\left[-\left(\frac{\alpha}{y}\right)^{\beta}\right], \tag{4.4}$$

and the LC1–IW density function

$$g(w) = \frac{1}{\eta_U - w}\frac{\beta}{\alpha}\left(\frac{\alpha}{y}\right)^{\beta+1}\exp\left[-\left(\frac{\alpha}{y}\right)^{\beta}\right], \tag{4.5}$$

where

$$y = ln\frac{\eta_U - \eta_L}{\eta_U - w}. \tag{4.6}$$

Third, the Log-Reciprocal Type II–Inverted Weibull (LC2–IW) was estimated for $G^F_{t,1}$, $G^F_{t,2}$, $G^F_{t,4}$, $G^R_{t,1}$, $G^R_{t,2}$, and $G^R_{t,6}$. The LC2–IW distribution function can be written as

$$G(w) = 1 - \exp\left[-\left(\frac{\alpha}{y}\right)^{\beta}\right], \tag{4.7}$$

and the LC2–IW density function

$$g(w) = \frac{1}{w - \eta_L}\frac{\beta}{\alpha}\left(\frac{\alpha}{y}\right)^{\beta+1}\exp\left[-\left(\frac{\alpha}{y}\right)^{\beta}\right], \tag{4.8}$$

where

$$y = ln\frac{\eta_U - \eta_L}{w - \eta_L}. \tag{4.9}$$

Finally, the Log Ratio Type 1–Laplace (LR1–LP) was estimated for $G^F_{t,3}$, $G^F_{t,5}$, and $G^R_{t,4}$. The LR1–LP distribution function can be written as

$$G(w) = \begin{cases} \frac{1}{2}\exp\left(\frac{y-\beta}{\alpha}\right), & \text{if } y \le \beta, \\ 1 - \frac{1}{2}\exp\left(-\frac{y-\beta}{\alpha}\right), & \text{if } \beta \le y, \end{cases} \tag{4.10}$$

and the LR1–LP density function

$$g(w) = \frac{\eta_U - \eta_L}{(w - \eta_L)(\eta_U - w)}\frac{1}{2\alpha}\exp\left(-\frac{|y-\beta|}{\alpha}\right), \tag{4.11}$$

where

$$y = ln\frac{w - \eta_L}{\eta_U - w}. \tag{4.12}$$

The next sub-section lays out the parameters for each distribution function and the analysis of goodness-of-fit.

4.4 Parameters and Graphs

Table 4.2 summarizes the parameter values of distribution functions $G^i_{t,j}$ of W^i for $j = 1, ..., 6$. The equations for all distributions previously stated along with their parameters provide a systematical approach to quantify the uncertainty of the farmers about the yield of their field and the yield of their region. The Maximum Absolute Difference (MAD) was calculated for each hypothesized distribution, and the one with the lowest MAD was selected. According to Krzysztofowicz (2014), the distribution functions with MAD below 0.05 indicate excellent to good fit, and the MAD between 0.05 and 0.10 indicate good to adequate fit. The distribution in table 4.2 fall into these two categories.

Figure 4.2 shows the judgmental quantiles assessed from the farmers in Mato Grosso and the estimated distribution functions $G^F_{t,j}$ of W^F. Similarly, figure 4.3 shows the estimated distribution functions $G^R_{t,j}$ of W^R. They display the farmers' assessments in terms of probabilities, and the shape of the distribution function can provide insightful information about how confident a farmer is about his predictions.

The shape of some of the distributions in figures 4.2 and 4.3, such as $G^F_{t,1}$, shows a shorter right tail. This shape could represent the farmers' confidence about their yield limitations, possibly from historical observations of yield in that same field or region, and biological restrictions of the seed. However, the farmers seem to be less confident about their estimates closer to the lower bounds. In this case, many external variables that are hard to predict can influence the yield, such as weather, diseases, or insect infestations.

Table 4.2 The parameter values and goodness-of-fit measures of the prior distribution functions $G^i_{t,j}$ of W^i, judgmentally assessed by the farmers in Mato Grosso, Brazil.

j	i	Distribution	α^i_j	β^i_j	$\eta^i_{L,j}$	$\eta^i_{U,j}$	MAD
1	F	LC2–IW	0.36	2.42	30	90	0.049
2	F	LC2–IW	0.50	3.80	30	90	0.061
3	F	LR1–LP	0.55	-0.30	30	110	0.042
4	F	LC2–IW	0.28	1.19	20	90	0.054
5	F	LR1–LP	0.74	-0.16	20	90	0.020
6	F	LC1–WB	0.81	2.39	30	90	0.026
1	R	LC2–IW	0.62	2.70	30	90	0.054
2	R	LC2–IW	0.83	5.90	30	90	0.056
3	R	LC1–WB	0.53	5.29	30	90	0.016
4	R	LR1–LP	0.36	-0.11	30	90	0.029
5	R	LC1–IW	0.36	2.11	30	90	0.030
6	R	LC2–IW	0.86	4.67	30	90	0.047

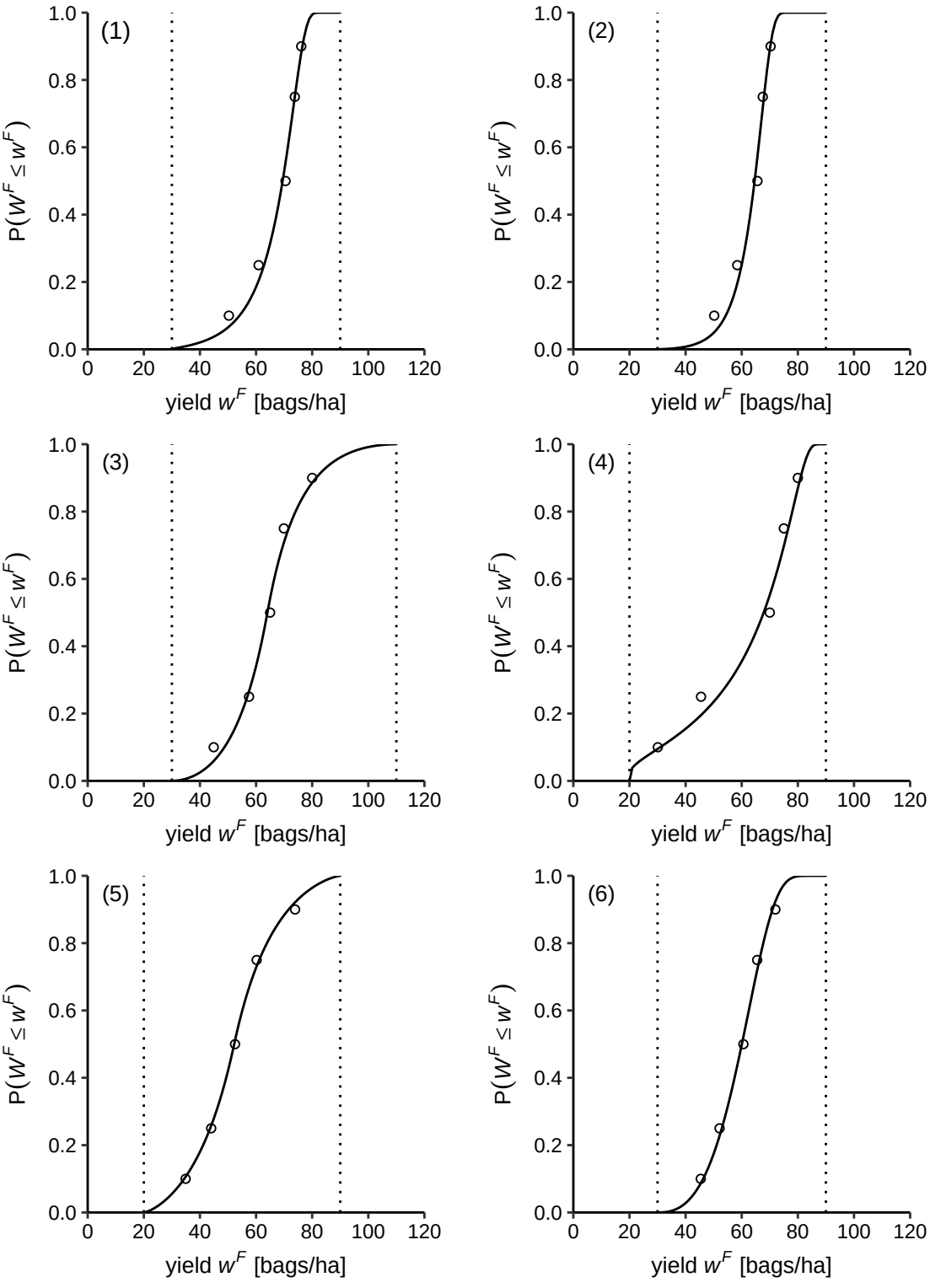


Figure 4.2 Distribution functions G_j^F of W_j^F, for $j = 1, ..., 6$, in terms of the five quantiles judgmentally assessed.

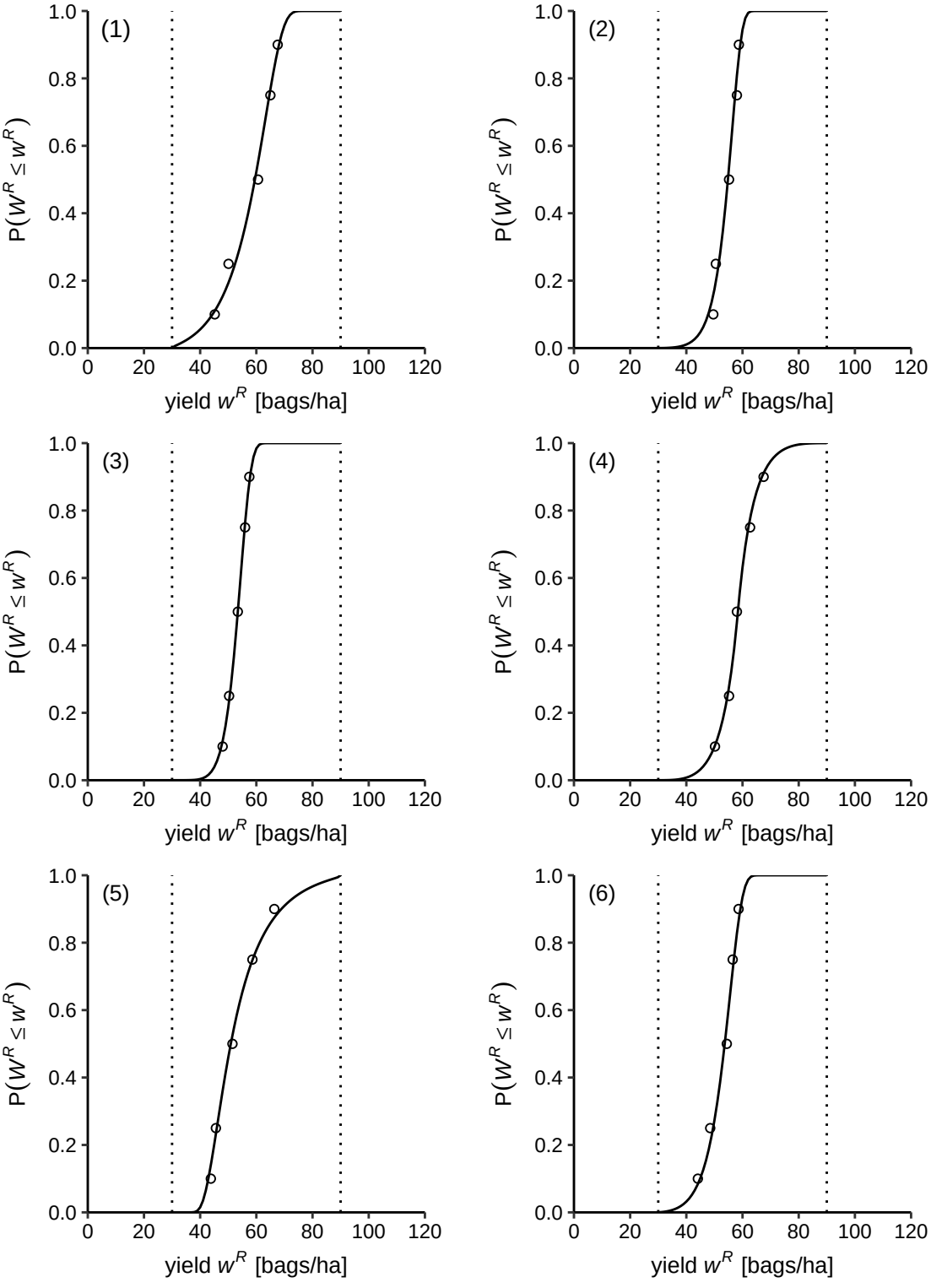


Figure 4.3 Distribution functions G_j^R of W_j^R, for $j = 1, ..., 6$, in terms of the five quantiles judgmentally assessed.

4.5 Combining Judgmentally Assessed Distribution Functions

In forecasting problems, groups of experts are often called upon to provide forecasts instead of one single expert. Many authors have devoted their research to the problem of combining judgmentally assessed distributions into a single aggregated distribution function. Winkler (1968) refers to this specific problem as the "consensus problem".

Krzysztofowicz (2014) describes the problem of group forecasting and proposes a solution through reconciling and combining the assessments. This is a behavioral aggregation approach according to Clemen and Winkler (1999). Another approach to aggregate multiple assessments is the mathematical combination of distributions.

Clemen and Winkler (1999) describe two categories of mathematical combinations: an axiomatic, and a Bayesian. The axiomatic approach consists of deriving the form of the aggregated distribution function based on pre-determined assumptions. An example of this method is the linear combination of individual distribution functions into a single distribution function. The Bayesian method may be more challenging to apply. It is usually described in studies through the decision maker's perspective, i.e., how the decision maker can update his prior distribution function using the experts' distribution functions. This study considers a weighted-average method with equal weights first, and then expands the theory of Bayesian Model Averaging (BMA) to construct a weighting system.

Consider a Bayesian forecaster of a continuous predictand W, with a continuous predictor X, when J experts are requested to provide their prior information about W. Each expert provides a different prior density function of W, and the task is to obtain a single posterior density function of predictand W, conditional on the realization of the predictor $X = x$.

4.5.1 Generic Model

Each farmer j will provide a different prior density function, $g_j^i(w^i)$, for $j = 1, ..., J$. The individual density functions will be combined in order to derive one single posterior density function. The mixture of the J component prior density functions is:

$$g^i(w^i) = \sum_{j=1}^{J} \lambda_j g_j^i(w^i). \tag{4.13}$$

The expected density function of *predictor* X is:

$$\begin{aligned} \kappa^i(x) &= \int_{\mathcal{W}^i} f^i(x|w^i) g^i(w^i) dw^i \\ &= \int_{\mathcal{W}^i} f^i(x|w^i) \sum_{j=1}^{J} \lambda_j g_j^i(w^i) dw^i \\ &= \sum_{j=1}^{J} \lambda_j \int_{\mathcal{W}^i} f^i(x|w^i) g_j^i(w^i) dw^i \\ &= \sum_{j=1}^{J} \lambda_j \kappa_j^i(x). \end{aligned} \tag{4.14}$$

The posterior density function of *predictand* W is:

$$\begin{aligned} \phi^i(w^i|x) &= \frac{f^i(x|w^i) g^i(w^i)}{\kappa^i(x)} \\ &= \frac{f^i(x|w^i) \sum_{j=1}^{J} \lambda_j g_j^i(w^i)}{\kappa^i(x)} \\ &= f^i(x|w^i) \frac{\sum_{j=1}^{J} \lambda_j g_j^i(w^i)}{\sum_{k=1}^{J} \lambda_k \kappa_k^i(x)} \\ &= \frac{\sum_{j=1}^{J} \lambda_j g_j^i(w^i) f^i(x|w^i) \kappa_j^i(x) / \kappa_j^i(x)}{\sum_{k=1}^{J} \lambda_k \kappa_k^i(x)} \\ &= \sum_{j=1}^{J} \Lambda_j(x) \phi_j^i(w^i|x). \end{aligned} \tag{4.15}$$

The weight system $\Lambda_j(x)$ is:

$$\Lambda_j(x) = \frac{\lambda_j \kappa_j^i(x)}{\sum_{k=1}^{J} \lambda_k \kappa_k^i(x)} = \frac{\lambda_j \kappa_j^i(x)}{\kappa^i(x)}, \tag{4.16}$$

and

$$\phi_j^i(w^i|x) = \frac{f^i(x|w^i) g_j^i(w)}{\kappa_j^i(x)}. \tag{4.17}$$

Functions (4.13) and (4.15) provide a comparison between the mixture of prior density functions and the mixture of posterior density functions. Clemen and Winkler (1999) summarize some of the findings about comparisons between simple weighting systems, such as averages, and more complex systems when combining probabilities or probability distributions. In terms of probabilities, many studies conclude that simpler combination methods can perform better than more complex methods. However, in the case of probability distributions, the Bayesian approach may have advantages over the axiomatic approach.

Intuitively, a farmer whose forecasts are more informative should be assigned a higher weight in the mixture model than a farmer whose forecasts are less informative, assuming both farmers are well calibrated. For this reason, the weight λ_j could be interpreted as the probability of farmer j being well calibrated and most informative. According to Krzysztofowicz (2016), the analysis of calibration of a forecaster must be done over time by recording the assessed quantiles and the realizations of *predictand* W.

This interpretation could be used in conjunction with the Bayesian Model Averaging (BMA) theory in order to estimate the weights λ_j, once these data are available. The BMA was created as an alternative to address the uncertainty regarding model selection using Bayes theorem (Hoeting et al., 1999), but this research will apply BMA to the prior mixture weighting system within the Bayesian forecaster.

4.5.2 Uniform Combination

Winkler (1968) refers to the combination of subjective distribution functions with equal weights as the case when the decision maker is unable to evaluate the performance of the experts in forecasting a variate. It is reasonable to assume initially that the models are equally likely to be the true model. Hoeting et al. (1999) refer to this case as a *neutral* choice.

Consider a mixture model in which the aggregated prior density function g^i, defined in (4.13), is a simple average of the J density functions, i.e., it is a linear combination with equal weights $\lambda_j = 1/J$ for $j = 1, ..., J$. For the case of g^F with J = 6,

$$g^F(w^F) = \sum_{j=1}^{6} \frac{1}{6} g^F_{t,j}(w^F). \tag{4.18}$$

Since the estimated parametric models for $G^F_{t,j}$ take several forms, function G^F would be the combination of the distribution functions detailed in table 4.2. Figures 4.4 and 4.5 show the individual distribution functions, $G^i_{t,j}$, and the combined distribution functions G^i, for $i = F, R$.

In order to obtain a conventional distribution function from the mixture, additional steps are taken. First, a data set is generated in the form of a sequence of values from η_L to η_U by 1 unit. Second, the function G^F is utilized to calculate the corresponding sequence of probabilities. Third, a new distribution function is optimally fitted to the generated sequence of values and the corresponding probabilities.

4.5.3 Bayesian Model Averaging (BMA)

The BMA is a method to combine forecasts from different models using Bayes' Theorem, given the same data set. In this study, the BMA approach can be used to obtain λ_j, for $j = 1, ..., J$, given a historical record of the farmers' forecasts. Since this data is

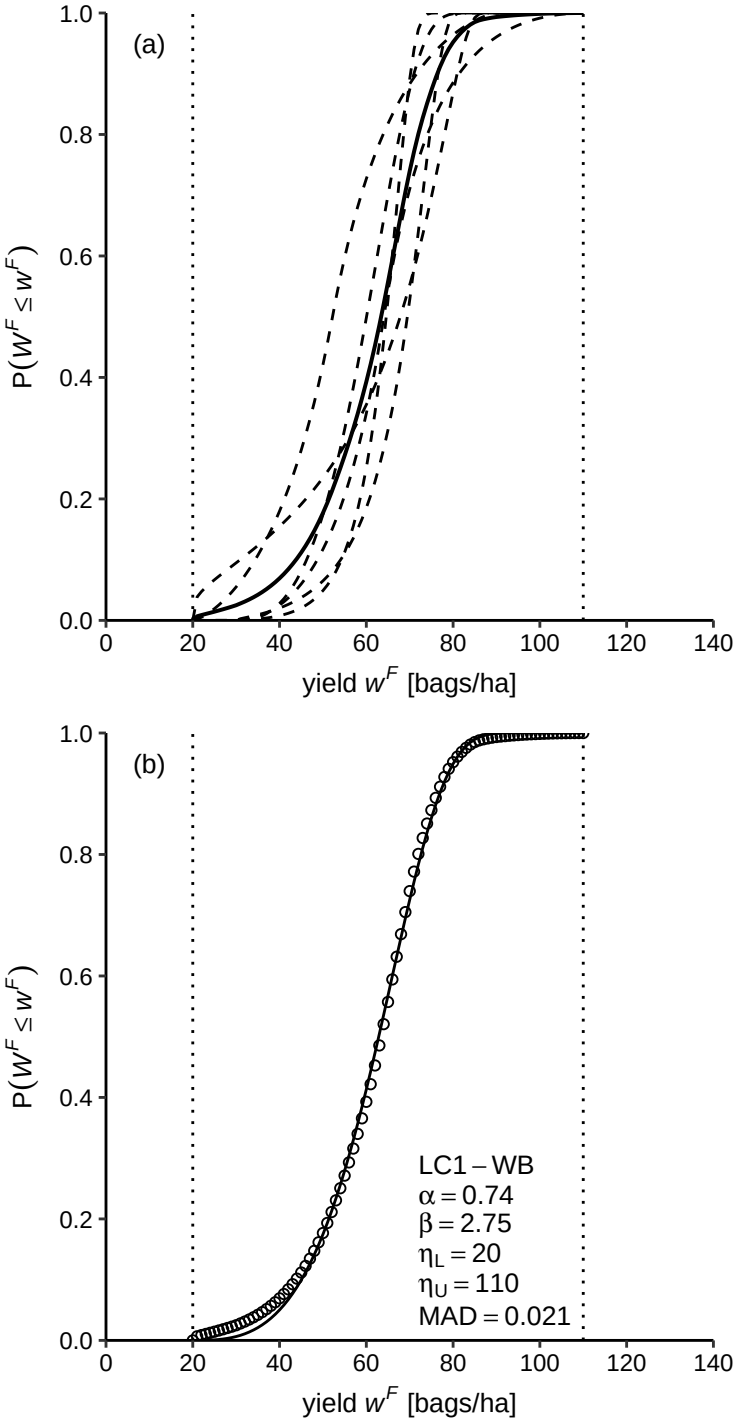


Figure 4.4 Combined prior distribution function G^F of W^F superimposed on individual prior distribution functions G_j^F , for $j = 1, ..., 6$, respectively.

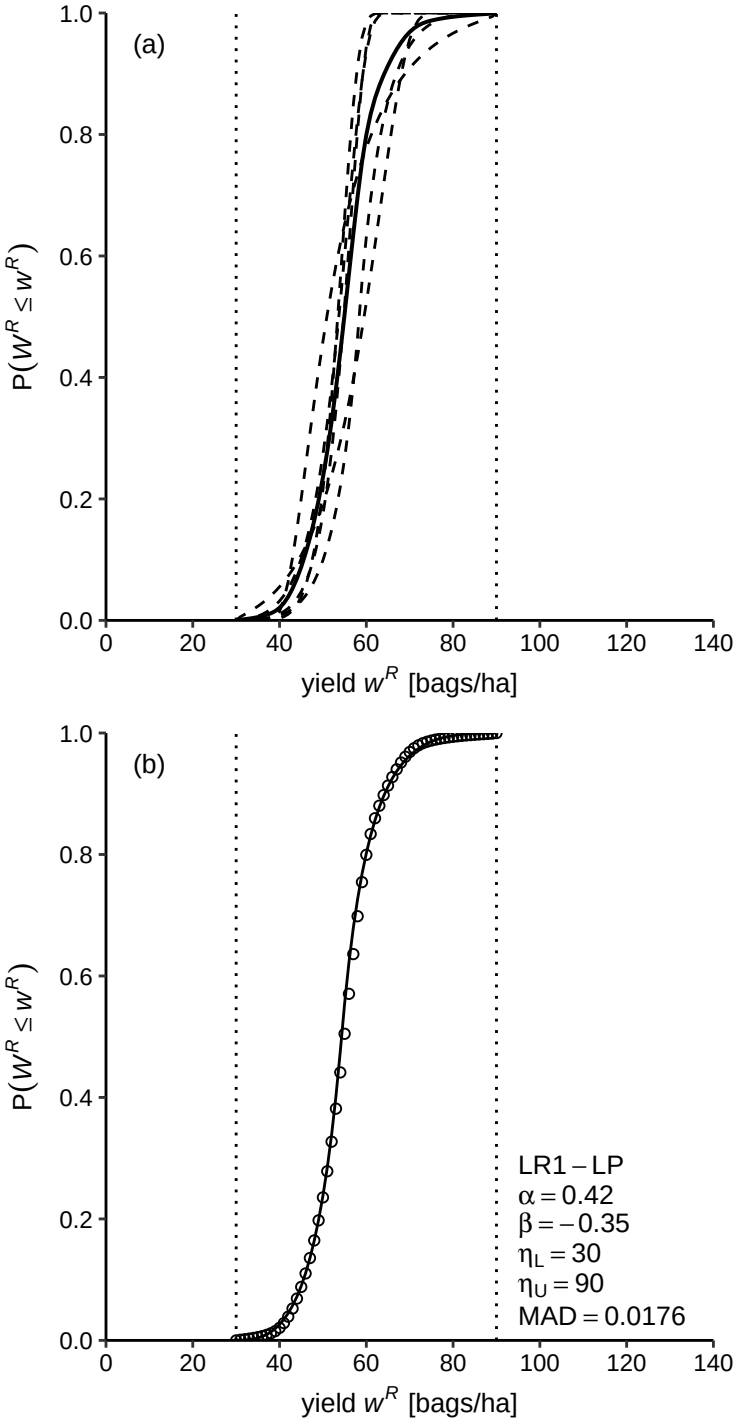


Figure 4.5 Combined prior distribution function G^R of W^R superimposed on individual prior distribution functions G_j^R , for $j = 1, ..., 6$, respectively.

unavailable at this moment, a simulation was conducted in Appendix A to illustrate this approach.

Consider that each farmer provided judgmental quantiles that were used to estimate a parametric distribution function. Each judgmental distribution function $G^i_{t,j}$ of W^i is a different model. Briefly, the BMA is an average of the models, each weighted by its posterior model probability, i.e., the probability of the model being the true one, given the data available. The BMA approach developed in this sub-section is applicable to both $i = For R$. Therefore, the index i is dropped in order to have a clearer notation.

Expanding this concept, an expert's judgmental distribution functions can be evaluated by retrospectively comparing the actual realizations with his assessments. Consider a group of experts S_j, for $j = 1, ..., J$. Instead of examining the probability of a model being the true one, this study examines the posterior probability of expert S_j being well-calibrated given the realizations of W^i, and compare it to the other experts. Therefore, the goal is to estimate posterior probability of expert S_j being well-calibrated.

Consider a situation where the historical data of W is available until the previous forecast time $T-1$, i.e., $D = \{(g_{t,j}, w(t)) : t = 0, ..., T-1\}$. Using the BMA, adapted from Hoeting et al. (1999), the combined prior density function of $W(T)$, given the data D from T years is

$$g_T(w(T)|D) = \sum_{j=1}^{J} g_{T,j}(w(T))P(S_j|D), \tag{4.19}$$

where D is the historical data sample of *predictand* W, and $P(S_j|D)$ is the posterior probability of S_j being well-calibrated given data D, such that

$$\sum_{j=1}^{J} P(S_j|D) = 1. \tag{4.20}$$

In sub-section 4.5.1, $P(S_j|D)$ is represented by λ_j. The first task is to model $P(S_j|D)$.

Using Bayes' theorem,

$$P(S_j|D) = \frac{p(D|S_j)P(S_j)}{p(D)}, \tag{4.21}$$

where $p(D)$ is the expected density function of the observed data D and can be estimated, according to Liu (2018), by

$$p(D) = \sum_{j=1}^{J} \left[\prod_{t=0}^{T-1} g_{t,j}(w(t)) \right] P(S_j), \tag{4.22}$$

and $p(D|S_j)$ is the likelihood of farmer S_k being well-calibrated given the historical data D

$$p(D|S_j) = \prod_{t=0}^{T-1} g_{t,j}(w(t)). \tag{4.23}$$

Once the prior probability $p(S_j)$ of farmer S_j being well-calibrated is adequately modeled, the expected density function of the observed data D, $p(D)$, can be derived. Then, the posterior probability of S_j being well-calibrated given data D, $P(S_j|D)$, can be derived and utilized in equation (4.19) to obtain the posterior density function $g_T(w(T)|D)$.

However, since there is no data available to model $P(S_j)$ at this moment, let us consider once again the same prior probability $P(S_j) = 1/6$, for each farmer j. Inserting equations (4.22) and (4.23) into equation (4.21) allows us to derive the posterior probability of farmer j being well-calibrated given the data. For farmer 1, we have the following equation:

$$\begin{aligned} P(S_1|D) &= \frac{\prod_{t=0}^{T-1} g_{t,1}(w(t))P(S_1)}{\sum_{j=1}^{J} \left[\prod_{t=0}^{T-1} g_{t,j}(w(t)) \right] P(S_j)} \\ &= \frac{\prod_{t=0}^{T-1} g_{1,t}(w(t))}{\sum_{j=1}^{J} \left[\prod_{t=0}^{T-1} g_{t,j}(w(t)) \right]} \end{aligned} \tag{4.24}$$

The same rationale can be applied to the other farmers. Using the distribution functions previously modeled and listed in table 4.2, it is possible to obtain equation (4.24).

Finally, equation (4.24) can be inserted in (4.19) to obtain the combined prior density function,

$$\begin{aligned} g_T(w(T)|D) &= \sum_{j=1}^{J} g_{T,j}(w) P(S_j|D) \\ &= g_{1,T}(w) \frac{\prod_{t=0}^{T-1} g_{1,t}(w(t))}{\sum_{j=1}^{J} \left[\prod_{t=0}^{T-1} g_{t,j}(w(t)) \right]} + ... + g_{T,J}(w) \frac{\prod_{t=0}^{T-1} g_{t,J}(w(t))}{\sum_{j=1}^{J} \left[\prod_{t=0}^{T-1} g_{t,j}(w(t)) \right]} \\ &= \frac{g_{T,1}(w) \prod_{t=0}^{T-1} g_{t,1}(w(t)) + ... + g_{T,J}(w) \prod_{t=0}^{T-1} g_{t,J}(w(t))}{\sum_{j=1}^{J} \left[\prod_{t=0}^{T-1} g_{t,j}(w(t)) \right]}. \end{aligned} \tag{4.25}$$

4.6 Summary

This section provided a guideline to assess judgmental distribution functions from farmers and an initial discussion on how to combine them. However, both topics are still the focus of research in various fields. One of the shortcomings of assessing probabilities instead of point forecasts is the limited knowledge of the general public about statistical concepts, such as probability and quantiles. Therefore, the researcher must circumvent this fact using methodologies such as the one discussed in this chapter.

Once the proper methodology is applied, it s possible to construct a distribution function of a variate of interest instead of having a simple point forecast. This framework allows the researcher to analyze the performance of the forecasters in terms of calibration, informativeness, and confidence, given a historical record of forecasts. It also allows the researcher to take action based on these metrics, for instance, by providing training to improve calibration or combining forecasts using different weights using the BMA approach.

5. Field-Region Stochastic Transformation

In the previous section, farmers were asked to provide their assessments regarding the yield of their respective fields and the region. This section analyzes the stochastic relationship between the yield of a particular field and its corresponding region in order to obtain the input data for the forecast model of regional yield using local prior. Modeling this relationship allows us to use the farmers' assessments of the yield of a field to forecast the yield of their region. First, the locations of the interviewed farmers are described. The relationship between their field yield and regional yield is modeled, and the field quantiles are transformed into regional quantiles. Finally, the prior distribution functions are modeled for each farmer and combined into one prior distribution function to be used in the Bayesian Processor of Forecasts.

5.1 Overview

The agricultural production of an entire region is hardly homogeneous. Several factors can vary according to the location, such as soil type, altitude, weather conditions, and even community types. A smaller area, such as a specific field in a farm, presents more consistent conditions that a farmer intimately understands; this potentially reduces uncertainty. For that reason, farmers may have a higher "degree of belief" about the assessments of their field yield than the assessments of their regional yield.

The IBGE divides the Brazilian territory into states, mesoregions (Portuguese: mesorregiões), and microregions (Portuguese: microrregiões). Mesoregions are defined by a combination of similarity of communities, landscape, and organizational structure. Microregions are subsets of the mesoregions. They are defined by more specific attributes, such as the organizational structure of agricultural, industrial, and mining activities (IBGE, 2017a).

Figure 5.1 Map of Mato Grosso divided into microregions with the locations of the farmers interviewed.

Figure 5.1 shows the map of the state of Mato Grosso, and the location of each interviewed farmer. The gray lines divide the state into microregions. The center-west region of Brazil, composed of the states of Mato Grosso, Mato Grosso do Sul, and Goiás, produces more soybean than anywhere else in Brazil. Mato Grosso itself produced approximately 26% of the total Brazillian soybean production in 2017 with a total area of approximately 7.8 times the state of Ohio.

In the same year, Mato Grosso exported approximately 22% (14.8 million tons) of the total soybean grain exported by Brazil . Out of this amount, 72% was destined to Asia, 22% to Europe, and 4% to the Middle-East (MDIC, 2018). This high demand presents huge logistical challenges to transport or store the crop after harvest. Historically, the soybean exports from Mato Grosso have been heavily directed to Ports of Santos and Paranaguá, in the states of São Paulo and Paraná, respectively.

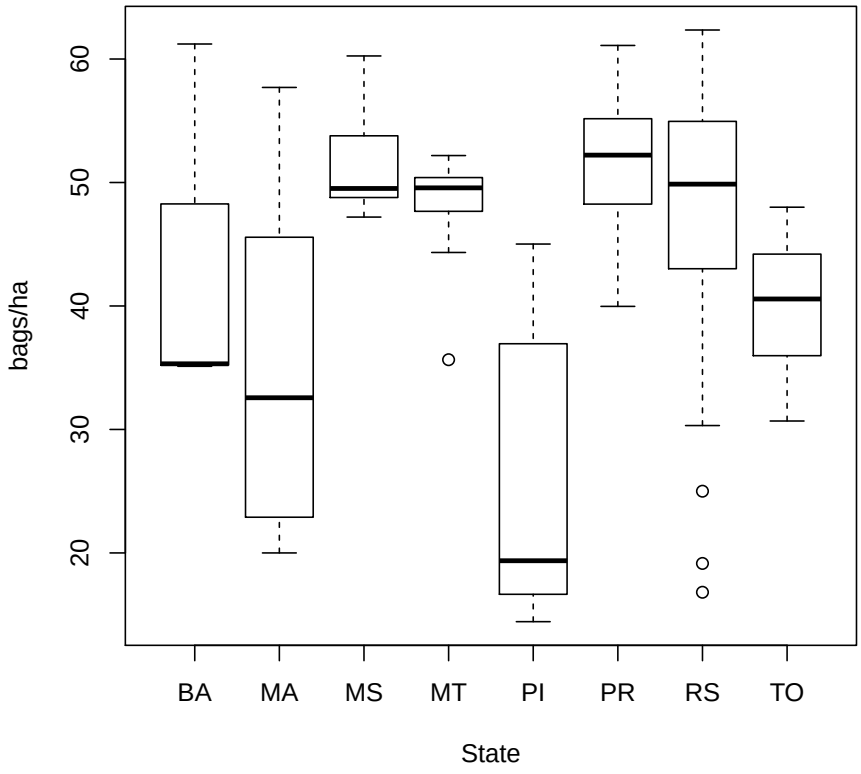


Figure 5.2 Boxplots of the microregion soybean yield in the 2016 season for selected states in Brazil.

The distance between the city of Sorriso (MT), a major logistic hub for soybean, and the Port of Santos (SP), for instance, is approximately 2,000 km (1242.7 miles) by truck. Nearly 47% of the soybean crop exported in 2017 passed through the Port of Santos, but a concerted effort has been made to make better use of the ports in the north, such as Barcarena (PA) with 19% of the soybean exportation in 2017 in order to relieve the demand in the south.

Figure 5.2 summarizes the soybean yield in microregions of several states in Brazil for the 2016 season. Some states such as Maranhão (MA) and Piauí (PI) present a considerable range of yield in their microregions. The state of Mato Grosso (MT) has less variable yields when compared to the other states. This aspect could represent more homoge-

Table 5.1 Location of farmers and corresponding microregions.

Farmer j	City	Microregion
1	Sorriso	Alto Teles Pires
2	Sorriso	Alto Teles Pires
3	Paranatinga	Paranatinga
4	Sapezal	Parecis
5	Diamantino	Parecis
6	Canarana	Canarana

neous production conditions in the state of Mato Grosso. The stochastic transformation presented in this section quantifies this uncertainty in a convenient way to be used along with the Bayesian Forecasting methodology.

In order to produce yield estimates, IBGE collects data from different sources, including their own collection system, technicians from other departments, farmers, and other individuals or institutions that are involved in preparing statistics for agriculture (IBGE, 2002). The regional harvested yield of a certain crop can be estimated by computing the total production and dividing by the total harvested area of a region. Therefore, the correlation between the yield of Mato Grosso and the yield of its microregions is expected.

Table 5.1 identifies the city and the microregion where each farmer interviewed in this study is located. The farmers are located in four different microregions within Mato Grosso. Therefore, for the purposes of this research, the yield of each microregion is mapped into the yield of the State of Mato Grosso. This mapping is used to transform the judgmentally assessed quantiles $w^F(p)$ into $w^R(p)$.

The yield of a particular microregion is used here as an approximation of the yield of a farmer's field. Ideally, the historical records of the yield of each farmer's land would be used to transform the field yield to the regional yield. However, the availability of such records depends upon the reliability of record keeping by farmers and their willingness to make private data public. However, the methodology developed in this study can be applied to individual data sets by farmers or analysts.

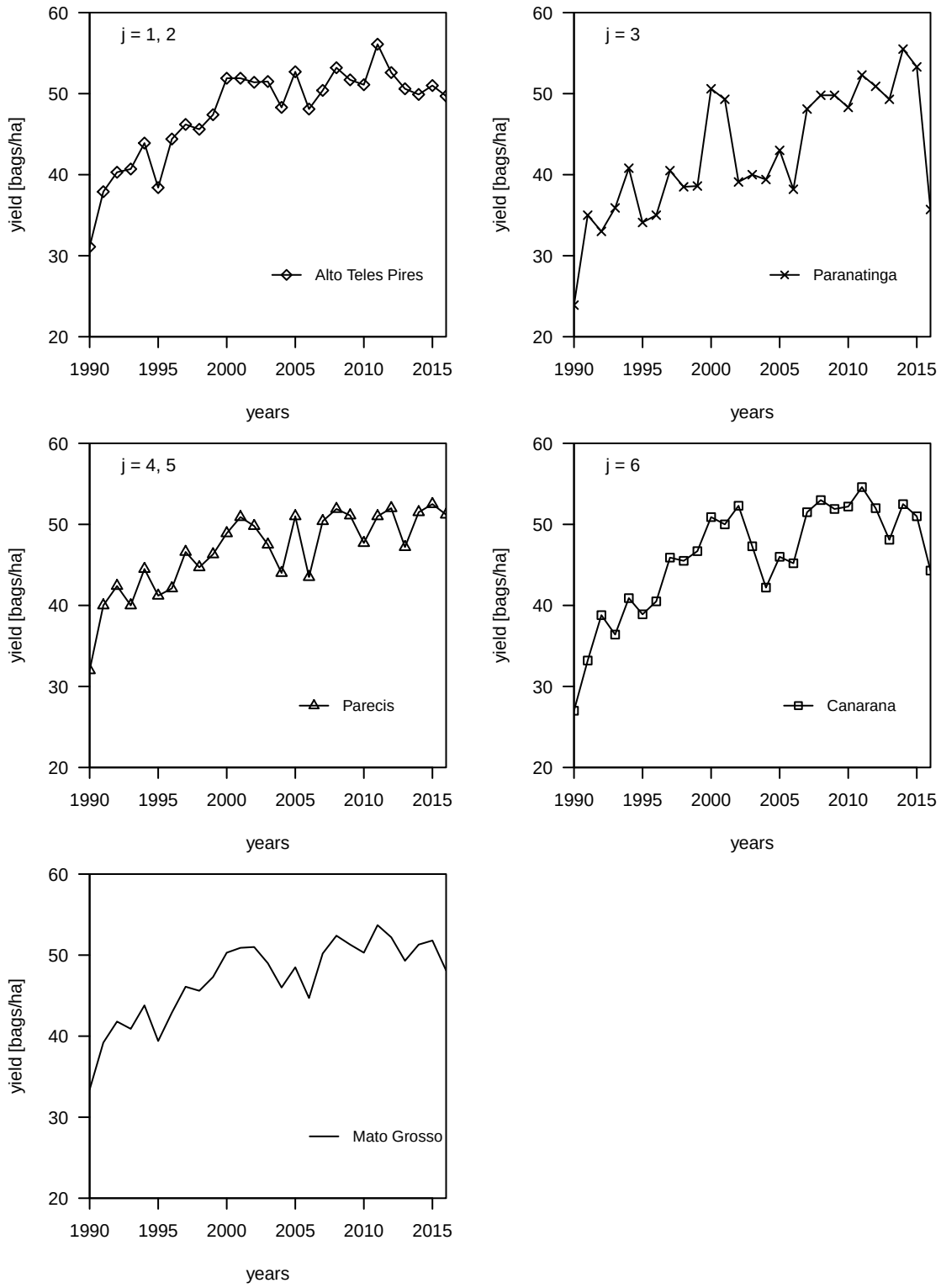


Figure 5.3 Soybean crop yield in Mato Grosso and 4 microregions from 1990 to 2016.

To overcome the insufficient availability of data, a stochastic transformation will be constructed for the microregion Alto Teles Pires and it will be utilized for farmers 1 and 2. Similarly, a Paranatinga transformation will be utilized for farmer 3, a Parecis transformation for farmers 4 and 5, and a Canarana transformation for farmer 6.

Figure 5.3 shows the soybean crop yield in Mato Grosso and in the 4 microregions previously mentioned. This is the data utilized to model the stochastic transformation in this chapter. The sample size for all sets are $N = 27$.

The apparent long-term upward trend in the yield of Mato Grosso could possibly represent advancements in technologies, such as in genetics, machinery, and productions methods. However, advancements in technology have a lower impact in the short-term. Variations from year to year are more likely to be caused by shifts in the weather or production conditions.

In general, the data sets in figure 5.3 do not seem to present any outliers, except for a large negative variation in the yield of Paranatinga from 2015 to 2016. Upon further investigation, farmers in Mato Grosso reported abnormal weather conditions during the planting stage of the 2015/2016 season. The insufficient precipitation affected the germination of the soybean seed, and many farmers had to replant. The entire state was affected, but some regions suffered more damage (CONAB, 2016). In addition, there is no clear seasonal patterns in these data sets that could be successfully modeled at this time.

The following section will model the normal-linear stochastic transformation between the yield of the microregions and the yield of Mato Grosso. The next step is to transform the assessed quantiles of W^F into quantiles of W^R, and model the judgmental prior distribution functions to be used in the Bayesian forecasters.

5.2 Normal-Linear Stochastic Transformation

Transforming values of the yield of a particular field, W_j^F, into the yield of a region, W^R, adds uncertainty to the forecast of W^R. Therefore, in addition to modeling the relationship between these two variates, this study quantifies the uncertainty related to this problem by deriving the stochastic transformation $\Psi_j(w^R|W_j^F = w_j^F)$, for $j = 1, ..., J$.

5.2.1 Framework

According to DeGroot (1970), a stochastic transformation from W_j^F to W^R is a nonnnegative function Ψ_j on the product space $\mathcal{W}^R \times \mathcal{W}_j^F$ that satisfies the condition

$$\int_{\mathcal{W}^R} \Psi_j(w^R|w_j^F)dw^R = 1, \tag{5.1}$$

for all $w_j^F \in \mathcal{W}_j^F$. The normal-linear stochastic transformation is a convenient model, starting with the form

$$w^R = c_j w_j^F + d_j + T_j, \tag{5.2}$$

where c_j and d_j are coefficients, and T_j is a residual, for $j = 1, ..., J$. Assuming that $T_j \sim N(m = 0, \tau_j^2)$, it follows that the stochastic transformation $\Psi_j(w^R|W_j^F = w_j^F)$ is a normal distribution with mean and variance

$$\begin{aligned} E(W^R|W_j^F = w_j^F) &= c_j w_j^F + d_j, \\ Var(W^R|W_j^F = w_j^F) &= \tau_j^2. \end{aligned} \tag{5.3}$$

The parameters c_j, d_j and τ_j can be estimated using linear regression. The normality and homoscedasticity of the random variable T_j must also be confirmed in order to proceed with this method. However, once these assumptions are verified, it is possible to rewrite

the linear model as $(W^R|W_j^F = w_j^F) \sim\ N(c_j w_j^F + d_j, \tau_j^2)$. In other words, the conditional distribution of W^R, given $W_j^F = w_j$ is normal-linear.

Based on these results, it is possible to define the normal-linear stochastic transformation

$$\Psi_j(w^R|w_j^F) = Q\left(\frac{w^R - c_j w_j^F - d_j}{\tau_j}\right), \tag{5.4}$$

where Q is the standard normal distribution function.

Each farmer j provided a set of points $\{(w_j^F(p), p)\}$, where $w_j^F(p)$ are the p-probability quantiles of W_j^F, and p is a fixed set of probabilities. Equation (5.4) can be used to transform this data set into $\{(w_j^R(p), p)\}$. The p-probability quantile $w_j^R(p)$ of W^R, given any quantile $w_j^F(p)$ of W^F, is

$$p = Q\left(\frac{w_j^R(p) - c_j w_j^F(p) - d_j}{\tau_j}\right), \quad 0 < p < 1. \tag{5.5}$$

The inverse of equation (5.5) is

$$Q^{-1}(p) = \frac{w_j^R(p) - c_j w_j^F(p) - d_j}{\tau_j}, \tag{5.6}$$

and it can be rearranged into

$$w_j^R(p) = c_j w_j^F(p) + d_j + \tau_j Q^{-1}(p). \tag{5.7}$$

Equation (5.7) transforms the judgmentally assessed quantiles of W_j^F into quantiles of W_j^R while accounting for the uncertainty associated with the stochastic transformation between these two variates. Deterministic models, while sometimes easier to process, might fail to carry this information forward. The vulnerability of this proposed model lies in the validation of the assumptions regarding the residuals of the linear regression.

This normal-linear stochastic transformation is the first step to addressing the hypoth-

esis that farmers may understand the dynamics of their field better than the dynamics of their region. The same formulation could also be explored when modeling other situations involving judgmental assessments where the decomposition of complex system is available.

The normal-linear stochastic transformation discussed here can be modeled according to the following procedure:

Step 1. Retrieve the joint sample of W^R and W_j^F, $\{(w^R, w_j^F)\}$, for each farmer $j = 1, ..., J$.

Step 2. Perform a linear regression of W^R on W_j^F to obtain the coefficients c_j, d_j and the sample of residuals $\{t_j\}$.

Step 3. Test the residuals $\{t_j\}$ for normality and homoscedasticity. If these two conditions are met, move on to the next step.

Step 4. Transform the quantiles $w_j^F(p)$ into $w_j^R(p)$ using equation (5.7).

5.2.2 Application

Let W^R be the yield of the state where the farmers are located, as estimated by IBGE. In this case, W^R is the soybean yield of Mato Grosso. Let an approximation of W_j^F be the soybean yield of the microregion where farmer j is located, also issued by IBGE in the annual publication Produção Agrícola Municipal (PAM). Since there are four microregions (table 5.1), a different normal-linear stochastic transformation is constructed for each microregion.

Figure 5.4 shows the linear regressions of W^R on W_j^F and the corresponding coefficients for each model. The same stochastic transformation can be utilized for groups of farmers in the same region. Table 5.1 summarizes the location of the farmers. Figure 5.5 shows the residuals from the regressions.

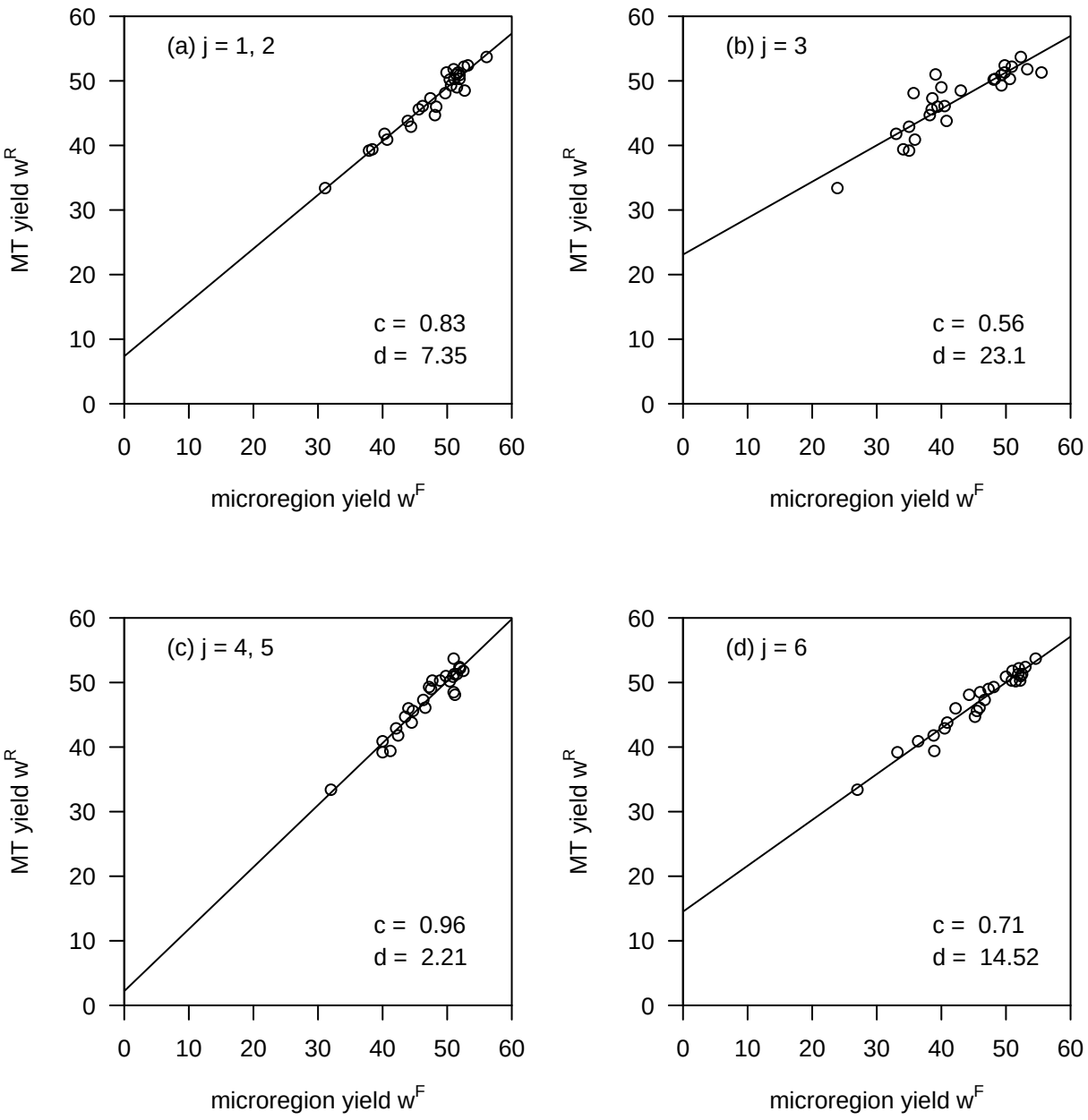


Figure 5.4 Linear regression of W^R on W_j^F [bags per hectare] in the microregions: (a) Alto Teles Pires, (b) Paranatinga, (c) Parecis, and (d) Canarana.

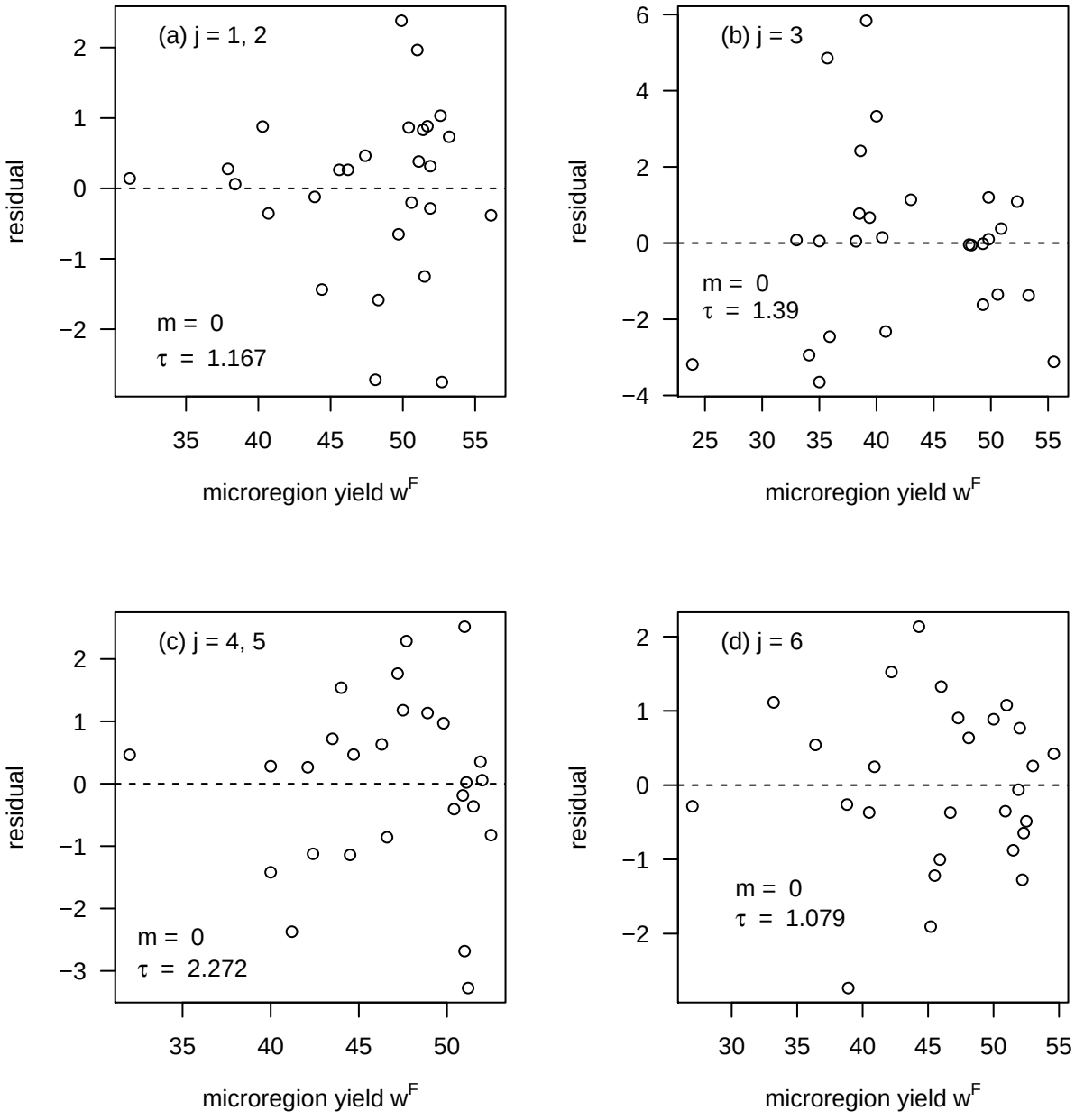


Figure 5.5 Residuals obtained from the linear regression of the regional yield on the local yield in the microregions: (a) Alto Teles Pires, (b) Paranatinga, (c) Parecis, and (d) Canarana.

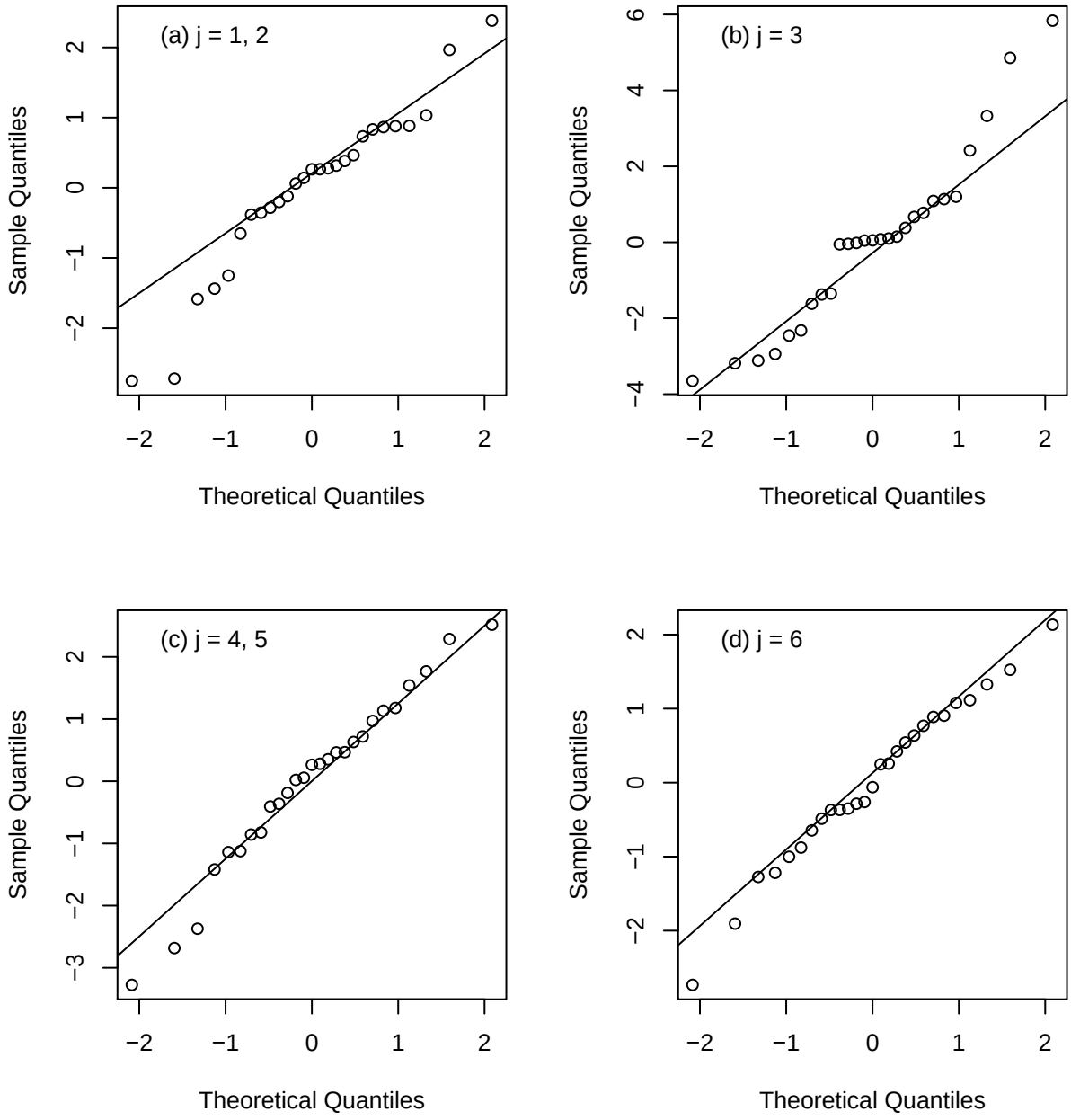


Figure 5.6 Normal probability plots for the residuals of the linear regression for the regional yield on the local yield in the microregions: (a) Alto Teles Pires, (b) Paranatinga, (c) Parecis, and (d) Canarana.

Table 5.2 Parameters of normal distribution functions fitted to regression residuals.

Regression	τ_j	MAD	K-S statistic	CV	Significance Level
Alto Teles Pires	1.167	0.151	0.118	0.201	≥ 0.20
Paranatinga	1.390	0.152	0.154	0.201	≥ 0.20
Parecis	2.272	0.088	0.064	0.201	≥ 0.20
Canarana	1.079	0.069	0.075	0.201	≥ 0.20

The coefficients of the linear regression were estimated using the method of least squares. This approach is particularly convenient for the stochastic transformation proposed here, since its solution seeks to minimize the sum of the squares of the residuals. The orthogonality condition between the residuals and the regressor in the least squares method forces the sum of the residuals to be 0. Although Hoaglin et al. (2000) indicate that the least squares method offers no resistance, it is suitable in this problem since the data set does not contain any obvious outliers. The assumption $T_j \sim N(m = 0, \tau_j^2)$ must be validated using the results of these regression models.

The residuals in Figure 5.5 do not appear to have a trend, i.e., they seem to be randomly placed around the average line. This implies a constant variance, or in other words, homoscedasticity of the residuals. Therefore, assuming τ_j^2 to be the variance of their distribution is suitable. The values of τ_j^2 were calculated using the maximum likelihood estimator.

Figure 5.6 shows the Gaussian probability plots of each set of residuals. The plots were constructed by adapting the methodology described by Hoaglin et al. (2000) with the meta-Gaussian plotting positions as described by Krzysztofowicz (2014). While the QQ plots for the residuals of the Parecis and Canarana regression indicate a normal distribution, the QQ plots of Alto Teles Pires and Paranatinga suggest a light-tailed distribution. However, this analysis must take into consideration the relatively small sample size N of 27 observations.

Table 5.2 shows the goodness-of-fit statistics for the normal distribution functions considered in each residual set. The MAD imply a good fit for the residual sets of Parecis

Table 5.3 Judgmental quantiles $w_j^R(p)$ of W^R transformed from $w_j^F(p)$ of W^F, in [bags/ha].

Farmer	City	$w_j^F(0.1)$	$w_j^F(0.25)$	$w_j^F(0.5)$	$w_j^F(0.75)$	$w_j^F(0.9)$
1	Sorriso	50.3	60.9	70.5	73.8	76.1
2	Sorriso	50.2	58.4	65.6	67.5	70.3
3	Paranatinga	44.9	57.5	65.0	69.9	80.0
4	Sapezal	30.1	45.5	70.0	75.0	80.0
5	Diamantino	34.9	44.0	52.5	60.2	73.9
6	Canarana	45.4	52.1	60.6	65.5	72.0

Farmer	City	$w_j^R(0.1)$	$w_j^R(0.25)$	$w_j^R(0.5)$	$w_j^R(0.75)$	$w_j^R(0.9)$
1	Sorriso	47.6	57.1	65.9	69.4	72.0
2	Sorriso	47.5	55.0	61.8	64.2	67.2
3	Paranatinga	46.5	54.4	59.5	63.2	69.7
4	Sapezal	28.2	44.4	69.4	75.7	81.9
5	Diamantino	32.8	42.9	52.6	61.5	76.1
6	Canarana	45.4	50.8	57.5	61.8	67.0

and Canarana, as expected from analyzing their QQ plots. Although the MAD for the residuals of the Alto Teles Pires and Paranatinga regressions suggest adequate fit, the Kolmogorov-Smirnov test did not reject the null hypothesis in both cases. Therefore, the assumption of $T_j \sim N(0, \tau_j^2)$ is validated.

The quantiles $w_j^F(p)$ are transformed into $w_j^R(p)$ using equation (5.7) and the coefficients in Figure 5.4. Table 5.3 summarizes the final results of the stochastic transformation. The next chapter will use these quantiles to model prior distribution functions of W^R, combine these functions, and construct Bayesian forecasters for agricultural yield.

5.3 Historical vs Judgmental Prior

Three types of prior distribution functions were modeled in this study: (1) the *historical* prior distribution G, (2) the *judgmental* prior distribution G_t^R of W^R constructed from the judgmentally assessed quantiles of W^R , and (3) the *transformed* prior distribution G_t^S of W^R obtained from the stochastic transformation of the judgmentally assessed quantiles of W^F.

The historical prior distribution function G was modeled using the records of yield issued by the LSPA/IBGE in September of each year, from 1993 to 2017 (sample size $N = 25$). The parametric model for G of W^R is LC2–IW

$$G(w^R|\alpha,\beta) = \exp\left[-\left(\frac{\alpha}{y}\right)^{\beta}\right], \tag{5.8}$$

where

$$y = ln\frac{\eta_U - \eta_L}{w^R - \eta_L}. \tag{5.9}$$

The prior distribution function G_t^R , constructed using the judgmentally assessed quantiles of W^R, and the prior distribution function G_t^S, constructed using the stochastic transformation of the judgmentally assessed quantiles of W^F into quantiles of W^R, follow the same model for the year $t = 2018$. The parametric model for G_t^R and G_t^S is LR1–LP

$$G_t^R(w^R|\alpha,\beta) = G_t^S(w^R|\alpha,\beta) = \begin{cases} \frac{1}{2}\exp\left(\frac{y-\beta}{\alpha}\right) & \text{if} \quad y \leq \beta \\ 1 - \frac{1}{2}\exp\left(-\frac{y-\beta}{\alpha}\right) & \text{if} \quad \beta \leq y \end{cases} \tag{5.10}$$

where

$$y = ln\frac{w^R - \eta_L}{\eta_U - w^R}. \tag{5.11}$$

These prior distributions are used at different moments of the construction of the Bayesian Forecaster. Table 5.4 summarizes the parameter values of the prior distribution functions and figure 5.7 shows the empirical and parametric prior distribution functions.

Table 5.4 Prior distribution functions of the net harvested yield in Mato Grosso, Brazil.

Prior	Distribution	α	β	η_L	η_U	MAD	K-S stat	Critical Value
G	LC2–IW	1.0777	6.7246	30	90	0.0712	0.110	0.208
G_t^S	LR1–LP	0.7369	-0.0329	30	90	0.0296	-	-
G_t^R	LR1–LP	0.4154	-0.3489	30	90	0.176	-	-

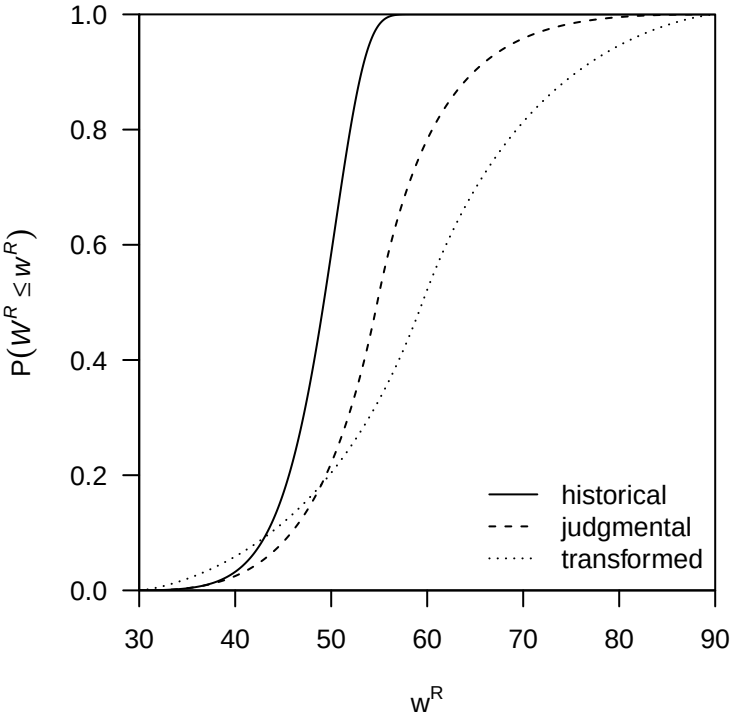


Figure 5.7 Parametric distribution functions G, G_t^S, and G_t^R of W^R.

The prior distribution functions G_t^R and G_t^S in figure 5.7 indicate a higher variance of W^R than the function G does. Intuitively, when considering the same variate, farmers with an expert understanding of the dynamics of the regional production would be able to indicate a lower variance of W^R. In other words, the scenario would be the opposite. In addition, the tails of the prior distribution functions G_t^R and G_t^S suggest a higher probability for values of yield considered extreme by the historical records.

This leads to conclude that the judgmental and transformed prior distribution functions modeled in this study are not completely accurate. Many factors can have influenced the shape of these distribution functions, such as incomplete understanding of the variate of interest by the experts or limitations in the methodology to assess quantiles. Nevertheless, these functions were useful to analyze the sensitivity of the Bayesian forecaster to judgmental prior distributions in chapter 7.

These shortcomings represent opportunities to improve the forecasting models devel-

oped in this chapter. Ideally, improvements in this type of functions can be made by collecting the subjective quantiles from the same experts over time, since the judgmental prior distribution functions in figure 5.7 only represent the year $t = 2018$. However, the user can construct new forecasters by updating the judgmental prior distribution functions using the framework developed in this research.

5.4 Summary

This chapter developed a stochastic transformation to map quantiles of W^F into quantiles of W^R. This model addresses the challenge of assessing information from experts about a certain variate, while the variate of interest for the forecasting problem is another. In this case, the methodology allows the growers to think about a variate that may be more familiar to them, the yield of a field, instead of the yield of the state.

6. Probabilistic Forecasting of Agricultural Yield

After observing the deterministic forecasts of production and yield, and modeling the judgmental prior distributions from farmers, decision makers can produce probabilistic forecasts using this information. This chapter uses the framework described in chapter 3 to address the problem of probabilistic forecasting of agricultural yield. Different Bayesian processors of forecasts (BPF) are constructed using a historical prior distribution function for various lead times.

6.1 Overview

The BPF models constructed in this chapter are in their conventional form, i.e., the prior information used is historical as opposed to being judgmentally assessed. In other words, the inputs to each model come exclusively from observed data. Analysts can use these models to quantify uncertainty about the yield of Mato Grosso and the deterministic forecasts of CONAB and IBGE, and merge them to produce probabilistic forecasts.

This chapter starts by analyzing the prior distribution function followed by the predictors before developing the specific structure of each forecasters. This layout allows an initial comparison between the prior distribution function and the marginal distribution functions of each predictor. Finally, each forecaster is constructed and the results are displayed. The forecasters differ mainly according to the lead time and the number of predictors used.

Table 6.1 aids to convert the yield data into the same unit, bags per hectare, which is used throughout this research. Although this is a trivial task, using units familiar to the farmers is essential to help them to understand the probabilistic forecasts.

Table 6.1 Agricultural unit conversions.

Unit	Equivalent
1 acre	0.404 hectare
1 bushel (soybean)	0.027 metric ton
1 bushel (soybean)	0.454 bags (60 kilograms)
1 bushel (soybean) per acre	0.06725 metric ton per hectare
1 bushel (soybean) per acre	1.122 bags (60 kilograms) per hectare
1 ton per hectare	16.667 bags (60 kilograms) per hectare

6.2 Variates and Samples

Let the continuous variate W^R with sample space $\mathcal{W}^R$ be the *predictand* — the net harvested yield of the soybean crop in the State of Mato Grosso, Brazil, in bags per hectare . Its realization is denoted $w^R \in \mathcal{W}^R$, where $\mathcal{W}^R = \{w^R : 0 < w^R < \infty\}$. In Brazil, IBGE and CONAB estimate the soybean crop yield of several states, including Mato Grosso. The actual value of W^R is considered to be the estimate issued in September by the LSPA/IBGE. A sample of $N = 25$ realizations was extracted from the LSPA/IBGE reports from 1993 to 2017.

Let the continuous variate W^F with sample space $\mathcal{W}^F$ represent the net harvested yield of a particular field, in bags per hectare . Its realization is denoted $w^F \in \mathcal{W}^F$, where $\mathcal{W}^F = \{w^F : 0 < w^F < \infty\}$. The farmer observes W^F every season after the harvest is complete. The observations of the yield come from recorded data in the farm. Ideally, a farmer with a well-structured historical database of his farm can use his own records to construct this model. However, this study uses the yield of the microregion within which the farmer resides as an approximation of w^F. A sample of $N = 25$ realizations was extracted from the LSPA/IBGE reports from 1993 to 2017 (see table 6.2).

The soybean yield can be affected, like many other crops, by weather conditions, nutrition, diseases, and many other factors. Growers can form their yield expectations based on current production conditions and using forecasts from other sources. Likewise, other agents in the food supply chain, such as traders, consultants, and the government

Table 6.2 Deterministic forecasts and actual estimates of the soybean crop yield in Mato Grosso, in bags per hectare.

Year	n	x_1	x_2	x_3	x_4	x_5	w^R
1993	1	na	na	39.68	na	40.60	40.90
1994	2	na	na	42.00	na	43.65	43.75
1995	3	na	na	41.80	na	41.55	41.55
1996	4	na	na	40.47	na	40.42	40.42
1997	5	na	na	42.77	na	43.53	43.53
1998	6	na	na	46.30	na	45.13	45.03
1999	7	na	na	45.65	na	45.65	47.30
2000	8	na	na	47.30	na	46.95	50.37
2001	9	na	na	50.50	na	49.98	50.92
2002	10	na	na	51.10	na	50.27	51.00
2003	11	na	na	52.08	na	51.48	48.07
2004	12	50.00	na	48.80	na	48.32	45.97
2005	13	48.83	na	50.33	na	48.42	48.42
2006	14	48.33	46.73	49.52	na	45.53	44.70
2007	15	48.67	50.83	48.80	49.67	50.55	50.32
2008	16	49.67	50.50	52.58	52.27	52.27	52.42
2009	17	50.00	51.67	51.75	51.65	51.38	51.33
2010	18	50.45	51.30	51.30	50.60	50.23	50.28
2011	19	51.00	52.25	52.25	53.17	53.72	53.72
2012	20	51.67	53.17	52.95	52.00	52.12	52.15
2013	21	51.67	52.12	51.60	50.17	49.38	49.32
2014	22	51.83	51.83	52.13	51.72	51.15	51.27
2015	23	52.13	51.92	51.80	52.75	51.65	51.83
2016	24	52.98	50.07	50.68	49.27	49.73	48.28
2017	25	52.18	54.62	53.87	54.55	54.93	55.03

Source: LSPA/IBGE, CONAB.
*na = not available

Table 6.3 Description of *predictors* X_l according to issued month, sample size, and source.

Variate	Forecast	Interval	Sample Size	Source
X_1	October	2004 to 2017	14	CONAB
X_2	February	2006 to 2017	12	CONAB
X_3	February	1993 to 2017	25	IBGE
X_4	May	2007 to 2017	11	CONAB
X_5	May	1993 to 2017	25	IBGE

can use farmers' assessments to form their forecasts as well as forecasts from other sources. For this reason, the *predictors* considered in this model are deterministic forecasts issued by IBGE and CONAB — two important government sources of agricultural information in Brazil.

Let the continuous variate X_l with sample space $\mathcal{X}_l$ be the *predictor* for $l = 1, ..., 5$. It represents the deterministic forecasts issued by IBGE and CONAB at different lead times. Its realization is $x_l \in \mathcal{X}_l$, where $\mathcal{X}_l = \{x_l : 0 < x_l < \infty\}$. Table 6.3 summarizes the attributes of each predictor considered in this problem. The soybean crop harvest usually happens from March through May in Mato Grosso. Therefore, the May forecast is expected to be the closest to the realization. Table 6.2 reports the realizations.

This sample is used to estimate the distribution functions of each predictor and predictand. The distribution functions are used to construct the BPF models. These functions can be updated over time once more observations are recorded. However, due to the nature of the problem, only one observation of each variate is recorded per year.

On the other hand, the models developed in this study can be applied to different types of crop and allow the use of multiple predictors. The readers can customize their BPF models according to the information available and with potential new predictors, increasing the sample size.

6.3 Prior Information

The prior information in this study is modeled through historical data of W^R . The historical prior distribution function G was modeled using the records of yield issued by the LSPA/IBGE in September of each year, from 1993 to 2017 (sample size $N = 25$). The parametric model for G of W^R is Log-reciprocal type II inverted Weibull (LC2–IW)

$$G(w^R|\alpha,\beta) = 1 - \exp\left[-\left(\frac{\alpha}{y}\right)^{\beta}\right], \tag{6.1}$$

where

$$y = ln\frac{\eta_U - \eta_L}{w^R - \eta_L}, \tag{6.2}$$

and η_L and η_U are the lower and upper bounds, respectively. These bounds are selected based on physical constraints associated with the variates, for instance, yields cannot be lower than 0, or subjectively according to a rigorous analysis of the sample space of the variate.

Figure 6.1 shows the empirical and parametric distribution functions of W^R and table 6.4 shows the parameter values of G. The MAD suggests good to adequate fit and the K-S statistic does not reject the model at a 20% significance level. Therefore, the LC2–IW model is good for the prior distribution function G. This function will be used in the construction of the BPF models in this chapter.

Table 6.4 Prior distribution function of the net harvested yield in Mato Grosso, Brazil.

Predictand	Distribution	α	β	η_L	η_U	MAD	K-S stat.	Critical value*
W^R	LC2–IW	1.0777	6.7246	30	90	0.0712	0.110	0.208

*At the significance level $\alpha = 0.20$

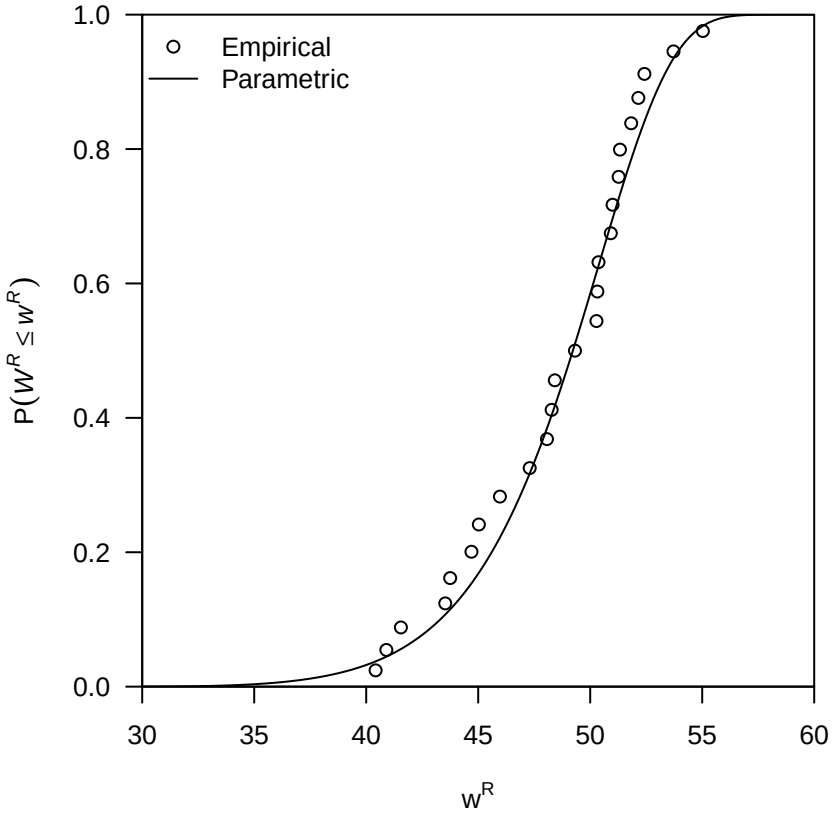


Figure 6.1 Empirical and parametric distribution functions of the W^R.

6.4 Predictors

The predictors considered in this study are the deterministic forecasts issued by CONAB and IBGE at different lead times before the harvest is complete. This section is devoted to model the initial estimate of the marginal distribution function $\bar{K}_i$ of X_i, for $i = 1, ..., 5$. The distribution functions are modeled according to the methodology in appendix B.

The parametric model for $\bar{K}_1, \bar{K}_3, \bar{K}_4, \bar{K}_5$ is Log-reciprocal type II inverted Weibull (LC2–IW)

$$\bar{K}_i(x_i|\alpha, \beta) = 1 - \exp\left[-\left(\frac{\alpha}{y}\right)^{\beta}\right], \tag{6.3}$$

where

$$y = ln\frac{\eta_U - \eta_L}{x_i - \eta_L}. \tag{6.4}$$

The parametric model for $\bar{K}_2$ is Log-reciprocal type I log-logistic (LC1–LL)

$$\bar{K}_i(x_i|\alpha, \beta) = \left[1 + \left(\frac{y}{\alpha}\right)^{-\beta}\right]^{-1}, \tag{6.5}$$

where

$$y = ln\frac{\eta_U - \eta_L}{\eta_U - x_i}. \tag{6.6}$$

Table 6.5 summarizes the marginal distribution functions estimated for each predictor. The models estimated for $X_1, X_2, X_4, \text{and} X_5$ are suitable, with MADs implying good fit, and K-S test not rejecting the null hypothesis (variate X_i has the distribution function $\bar{K}_i$) at a 20% significance level.

With a MAD indicating poor fit, the model $\bar{K}_3$ is still suitable with the K-S test not rejecting the null hypothesis at a 20% significance level. The modeling of such variates

Table 6.5 Marginal distribution functions of the *predictors* of the net harvested yield in Mato Grosso, Brazil.

Predictor	Distribution	α	β	η_L	η_U	MAD	K-S stat.	Critical value*
X_1	LC2-IW	1.03	16.07	30	90	0.071	0.122	0.267
X_2	LC1-LL	0.45	30.46	30	90	0.070	0.139	0.297
X_3	LC2-IW	1.05	7.45	30	90	0.106	0.139	0.208
X_4	LC2-IW	0.99	18.59	30	90	0.066	0.117	0.309
X_5	LC2-IW	1.08	6.76	30	90	0.074	0.111	0.208

*At the significance level $\alpha = 0.20$

can be improved with time, given a larger sample. Figure 6.2 shows the empirical and parametric distribution of each predictor.

The distribution functions $\bar{K}_1$, $\bar{K}_2$, and $\bar{K}_4$ modeled after the data issued by CONAB are quite steep. The variance of these deterministic forecasts is smaller than the variance of the IBGE forecasts. For instance, the probability of the yield of Mato Grosso being greater than 50 bags per hectare according to the estimated parametric distribution function $\bar{K}_2$ is approximately 0.96. The same event has probability of 0.49 according to $\bar{K}_3$.

Consider now the probability of the yield of Mato Grosso being greater than 53 bags per hectare. According to the CONAB February forecast, this probability is 0.10, while the IBGE February forecast suggests 0.14. This is the kind of disparity that can influence the reliability of these sources in decision making. The models developed in this research are able to take advantage of both predictors and merge this uncertainty in a systematical approach.

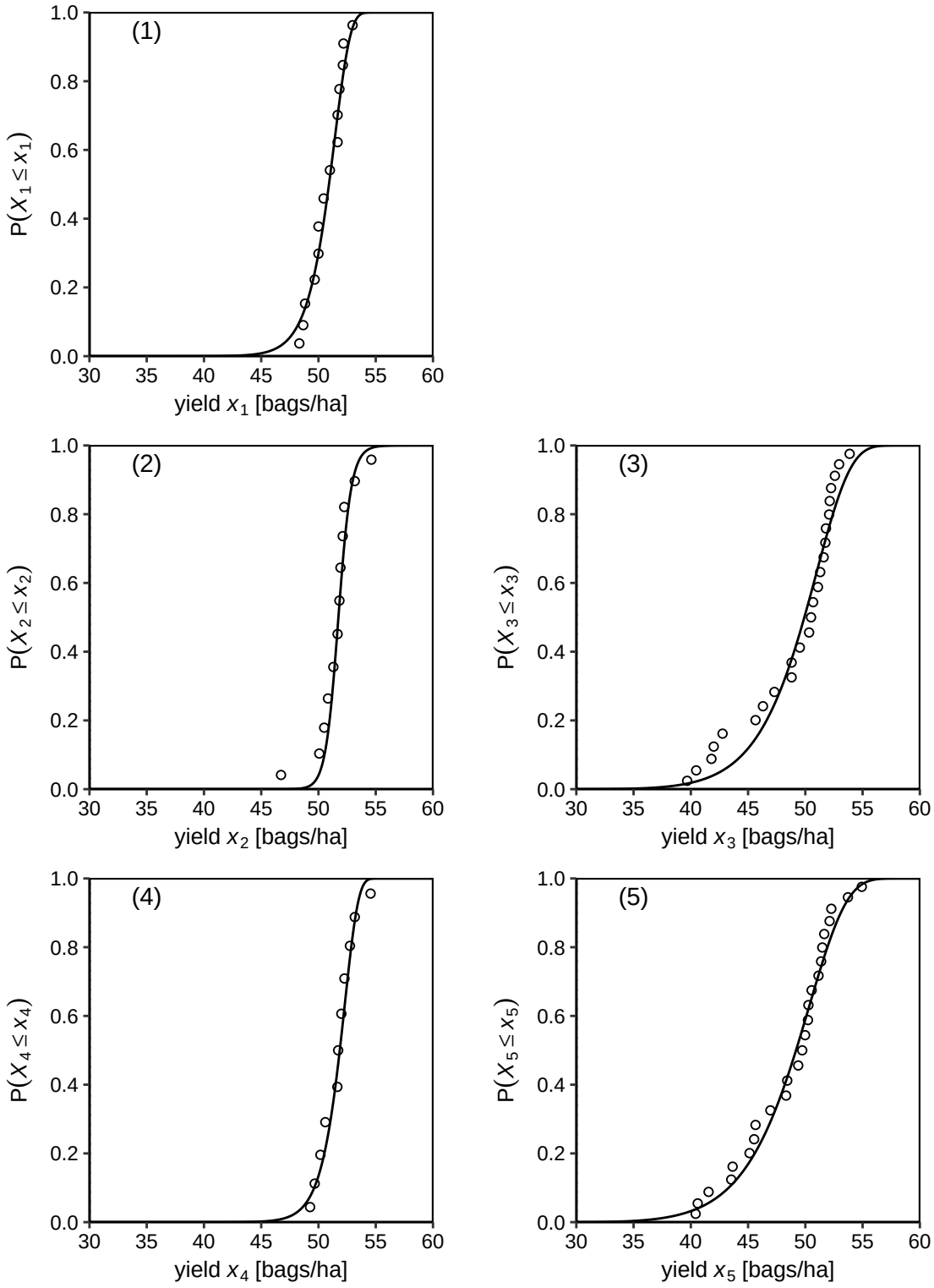


Figure 6.2 Empirical and parametric distribution functions of the: (1) CONAB October forecasts X_1, (2) CONAB February forecasts X_2, (3) IBGE February forecasts X_3, (4) CONAB May forecasts X_4, and (5) the IBGE May forecasts X_5.

6.5 Forecasting Regional Yield

Consider the case of a trading company looking to export a certain amount of soybean from Brazil to China. The exportation contract is closed before the farmers have planted the soybean. In order to fulfill this contract, the company purchases grain from farmers located in different states. This company is likely to be interested in yield forecasts during the development of the crop season.

Without much effort, an analyst can collect deterministic monthly yield forecasts for a state issued by IBGE and CONAB. These forecasts are free and available to the public. In addition, this analyst can purchase deterministic yield forecasts from consulting companies. The analyst's task is to determine which source is more reliable and how to use the deterministic forecast in decision making.

In this section, Bayesian processor of forecasts (BPF) models are constructed to address similar problems. Suppose an analyst has a data set containing records of the yield of Mato Grosso, and deterministic yield forecasts issued by IBGE and CONAB. The BPF produces probabilistic yield forecasts using both sources of deterministic forecasts. The framework utilized in this section is described in chapter 3.

The BPF models expose the uncertainty related to the deterministic yield forecasts, and provide a method to obtain probabilistic forecasts of regional yield. Three BPFs will be constructed in this section. Table 6.6 displays the predictors used in each BPF. The main difference between each BPF is the set of predictors. Additional information from newer forecasts, as well as changes in the farmers' judgments, can be incorporated using this framework.

Two frameworks must be applied to construct the BPF models in this section. First, a Bayesian meta-Gaussian model using one predictor is constructed for the BPF_{Oct}. Second, two Bayesian meta-Gaussian models using multiple predictors are constructed for BPF_{Feb} and BPF_{May}.

Table 6.6 *Predictors* used in each *Bayesian Processor of Forecasts*.

BPF	Predictors	Lead time (Months)
BPF_{Oct}	X_1	11
BPF_{Feb}	X_2, X_3	7
BPF_{May}	X_4, X_5	4

6.5.1 BPF - October

The report issued in October by CONAB is the first in the season; therefore, many decision makers usually expect its release. The BPF_{Oct} quantifies the uncertainty related to the yield forecasts for Mato Grosso and allows the user to produce probabilistic forecasts. The BPF_{Oct} considers the deterministic forecast issued in October by CONAB as the predictor X_1.

According to Maranzano (2006), modeling each Bayesian meta-Gaussian forecaster using one predictor involves the following steps:

Step 1. *Estimate the marginal distribution functions.* This step has been discussed in sections 6.3 and 6.4. Each BPF uses a different combination of predictors as shown in table 6.6. Therefore, the parametric models of the marginal distribution function are summoned as needed in the following sections.

Step 2. *Transform the variates using the NQT.* The Normal Quantile Transformation is a composition of the inverse of the standard normal distribution function and the distribution function of the variate.

Step 3. *Model the family of likelihood functions.* The family of likelihood functions is modeled in the transformed space. The dependence structure between the predictor and the predictand is assumed to be Normal-linear.

Step 4. *Validate dependence structure.* This step involves analyses whether the assumptions of the meta-Gaussian model hold.

Step 5. *Compute posterior parameters.* The posterior parameters along with the marginal distribution functions previously modeled are used to obtain the posterior distribution and density functions.

These steps are executed below in order to estimate and validate the BPF_{Oct}.

Step 1. Estimate marginal distribution functions

The parametric model of the marginal distribution function G for the yield of Mato Grosso is LC2–IW, as described in equations (6.1) and (6.2). The parametric model of the marginal distribution function $\bar{K}_1$ of X_1 is LC2–IW, as described in equations (6.3) and (6.4). The parameter values for G and $\bar{K}_1$ are shown in tables 5.4 and 6.5, respectively.

Step 2. Transform the variates using the NQT

Variate W^R is transformed using the NQT through the historical prior distribution G^H, and the predictor X_1 is transformed using the marginal distribution $\bar{K}_1$. The transformed variates using the NQT are

$$V = Q^{-1}(G(W^R)), \tag{6.7}$$

$$Z_1 = Q^{-1}(\bar{K}_1(X_1)). \tag{6.8}$$

Equations (6.7) and (6.8) are utilized to obtain the joint sample $\{(z_1, v)\}$. Each realization of variate V is computed as $v = Q^{-1}(G(w^R))$; that is, first evaluating the historical prior distribution function at the observation w^R, then evaluating the inverse of the standard normal distribution function at that value. Using this transformation for all the recorded values of W^R results in a transformed sample with same size as the original sample. The same process is made for $z_1 = Q^{-1}(\bar{K}_1(x_1))$. The family of likelihood functions is modeled in the transformed space. Figure 6.3-a shows a scatterplot of the transformed joint sample. This joint sample is shown in table 6.7.

Table 6.7 Transformed data obtained by the NQT.

Year	n	z_1	z_2	z_3	z_4	z_5	v
1993	1	na	na	-2.14	na	-1.75	-1.70
1994	2	na	na	-1.74	na	-1.22	-1.20
1995	3	na	na	-1.78	na	-1.59	-1.59
1996	4	na	na	-2.01	na	-1.78	-1.78
1997	5	na	na	-1.61	na	-1.24	-1.24
1998	6	na	na	-0.91	na	-0.93	-0.96
1999	7	na	na	-1.05	na	-0.83	-0.48
2000	8	na	na	-0.68	na	-0.55	0.32
2001	9	na	na	0.18	na	0.22	0.49
2002	10	na	na	0.37	na	0.31	0.51
2003	11	na	na	0.72	na	0.69	-0.29
2004	12	-0.53	na	-0.31	na	-0.22	-0.77
2005	13	-1.08	na	0.13	na	-0.20	-0.21
2006	14	-1.29	-3.85	-0.11	na	-0.85	-1.02
2007	15	-1.15	-0.98	-0.31	-1.29	0.39	0.31
2008	16	-0.70	-1.31	0.91	0.32	0.96	1.00
2009	17	-0.53	-0.08	0.60	-0.16	0.65	0.62
2010	18	-0.28	-0.49	0.44	-0.82	0.29	0.30
2011	19	0.05	0.55	0.78	1.18	1.54	1.51
2012	20	0.52	1.42	1.06	0.10	0.91	0.90
2013	21	0.52	0.41	0.54	-1.04	0.05	0.03
2014	22	0.64	0.10	0.74	-0.11	0.58	0.60
2015	23	0.89	0.20	0.61	0.75	0.75	0.79
2016	24	1.70	-1.69	0.23	-1.47	0.15	-0.24
2017	25	0.93	2.43	1.47	3.12	2.11	2.12
Sample mean		-0.02	-0.27	-0.16	0.05	-0.06	-0.08
Sample sd		0.88	1.53	1.03	1.26	1.00	1.00

*na = not available

Step 3. Model the family of likelihood functions

According to the framework discussed in chapter 3, the Bayesian meta-Gaussian model assumes that the stochastic dependence between Z and V is normal-linear. Therefore, a least squares regression is fitted using the joint sample $\{(z_1, v)\}$. The linear regression follows the model described in chapter 3:

$$Z_1 = aV + b + \Xi. \tag{6.9}$$

Figure 6.3-a shows this transformed joint sample, the least squares regression of Z_1 on V, and the 90% central credible interval. Figure 6.3-b shows the meta-Gaussian median regression of X_1 on W^R obtained by mapping the regression from the transformed space into the original space.

Step 4. Validate dependence structure

The residuals of the regression of Z_1 on V with slope $a = 0.408$ and intercept $b = -0.194$ are shown in figure 6.4-a. Initially, a visual analysis of the residuals does not indicate clear heteroscedasticity. The residuals seem to be randomly distributed around the null horizontal line. Figure 6.4-b shows the QQ plot of the residuals obtained using the meta-Gaussian plotting positions. The QQ plot indicates a possible discrepancy in the left and right tail from the Gaussian distribution. Therefore, an additional analysis is required in order to strengthen the validation process.

The goodness-of-fit of a hypothesized Gaussian distribution can be analyzed according to the methodology in Appendix B. The MAD of 0.0517 indicates good fit. The K-S statistic of 0.110 does not reject the null hypothesis that the residuals follow the Gaussian distribution with critical value of 0.276 at a significance level of 0.20. Therefore, there is no strong evidence to invalidate the assumptions of homoscedasticity, null mean, and normality of the residuals at this moment. The estimated parameters of the regression can be used to estimate the posterior parameters of the Bayesian forecaster.

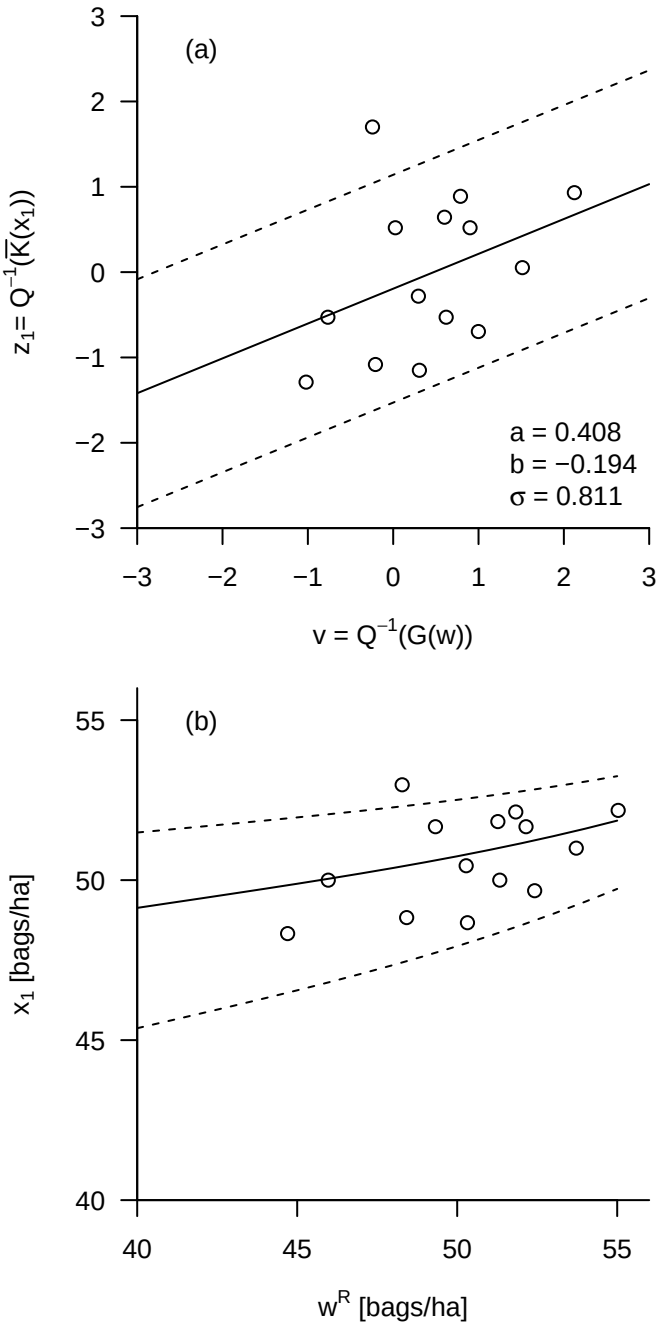


Figure 6.3 (a) Linear regression of Z_1 on V, and 90% central credible interval; (b) meta-Gaussian median regression of the deterministic forecast X_1 issued by CONAB in October on the regional yield W^R.

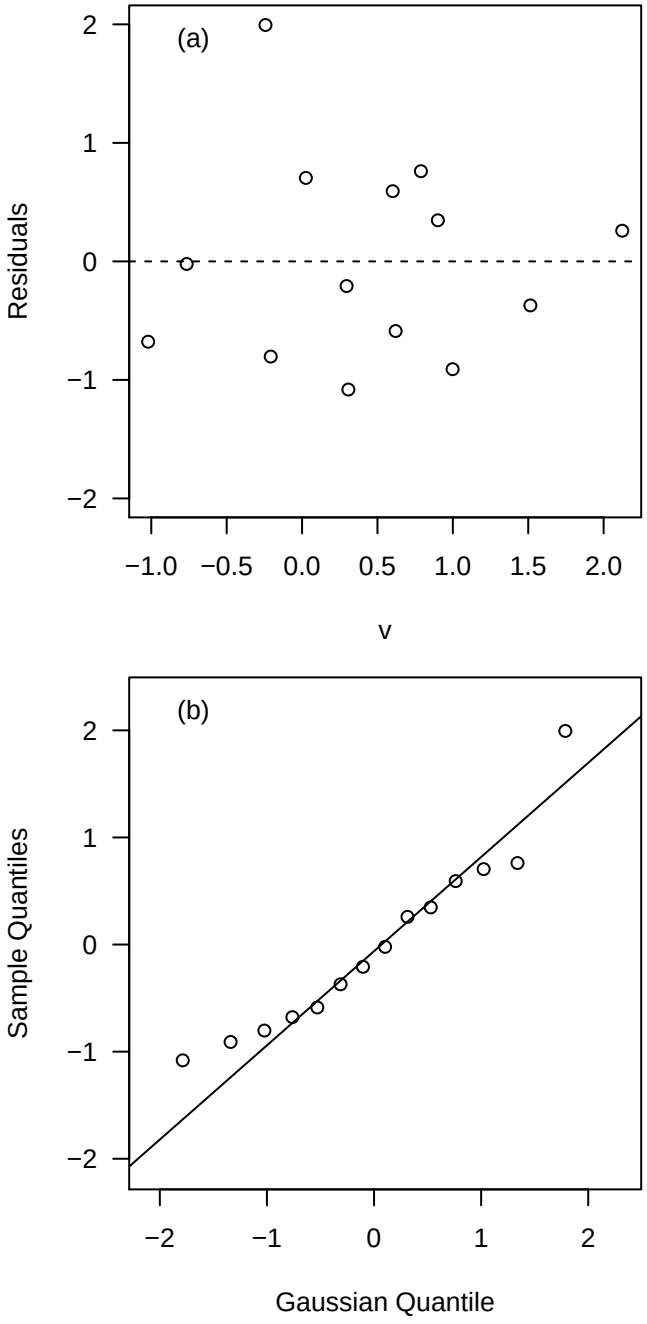


Figure 6.4 (a) Residuals from the linear regression of Z_1 on V; (b) QQ plot of the residuals constructed using the meta-Gaussian plotting positions.

Table 6.8 Quantiles $w^R(p)$ of the yield of Mato Grosso estimated from the historical prior distribution and posterior quantiles $w^R(p|x_1)$, given $x_1 = 47, 49, 52$ and $p = 0.01, 0.25, 0.5, 0.75, 0.9$.

Function	Forecast	0.01	0.25	0.5	0.75	0.9
Historical Prior	$w^R(p)$	37.1	46.4	49.2	51.5	53.2
Posterior Quantile	$w^R(p\|x_1 = 47)$	34.1	42.7	45.9	48.5	50.5
Posterior Quantile	$w^R(p\|x_1 = 49)$	36.2	44.8	47.6	50.0	51.7
Posterior Quantile	$w^R(p\|x_1 = 52)$	41.5	48.8	50.9	52.7	54.0

Step 5. Compute posterior parameters

Finally, the posterior parameters are computed. The estimates of posterior parameters of the Bayesian meta-Gaussian model are

$$c_1 = 0.495, \quad c_0 = 0.096, \quad \text{and } T = 0.893.$$

The posterior parameters along with the parameters of G and $\bar{K}_1$ are enough to specify the posterior quantile , distribution, and density functions of W^R. These functions produce probabilistic forecasts of W^R. The posterior quantile function of W^R, conditional on a realization of $X_1 = x_1$, is given by equation (3.13). The posterior density and distribution functions are given by equations (3.10) and (3.11), respectively.

The posterior quantile function produces forecasts of W^R given any values of $p \in (0, 1)$ and x_1. For instance, suppose CONAB released a deterministic forecast of the yield of Mato Grosso in October of $x_1 = 47$ bags per hectare and a decision maker needs to forecast the yield of Mato Grosso. Given equation (3.13) and the previously estimated parameters, it is possible to calculate the quantiles in table 6.8. Then, the decision maker can conclude, for instance, that it is equally likely that the yield of the region will be above or below 45.9 bags per hectare, or that the probability of the yield being greater than 48.5 bags per hectare is 0.25. Another conclusion could be that the probability of the yield being between 42.7 and 48.5 bags per hectare is 0.5.

Figure 6.5 shows posterior density and distribution functions, conditional on different

values of X_1, along with the historical prior density and distribution functions. These plots are useful to analyze particular cases of the family of posterior functions and compare them with the prior functions. The different shapes of the density functions in figure 6.5-b shows the impact of the deterministic forecast issued by CONAB in October on the uncertainty about the yield of Mato Grosso.

Intuitively, the deterministic forecasts with smaller lead time should converge to the actual realization of W^R. Therefore, it is expected that the posterior density functions follow a narrower shape than the prior density function according to the lead time of the deterministic forecast. Some of the posterior density functions in figure 6.5-b are not necessarily narrower than the prior density, but they represent a shift according to the observation x_1, such as for $x_1 = 47$ and $x_1 = 49$.

However, some posterior functions can represent a greater reduction in the uncertainty about the yield of Mato Grosso, given the October forecast from CONAB, such as the case of $\phi(\cdot|x_1 = 52)$. Suppose a trading company is concerned about low yields in Mato Grosso. Historically, the company knows that the probability of an yield lower than 46 bags per hectares is approximately 0.22, but after CONAB reported a forecast of 52 bags per hectare, the company can use the BPF_{Oct} model to calculate the posterior probability. The posterior probability of observing an yield lower than 46 bags per hectare is 0.08, given $x_1 = 52$.

The BPF_{Oct} is the first model to produce probabilistic forecasts of agricultural yield in this research. It takes a yield estimate issued by CONAB in October along with prior knowledge about the yield of Mato Grosso and outputs forecasts and their related probabilities. This represents a contribution to the current models used for forecasting agricultural yield.

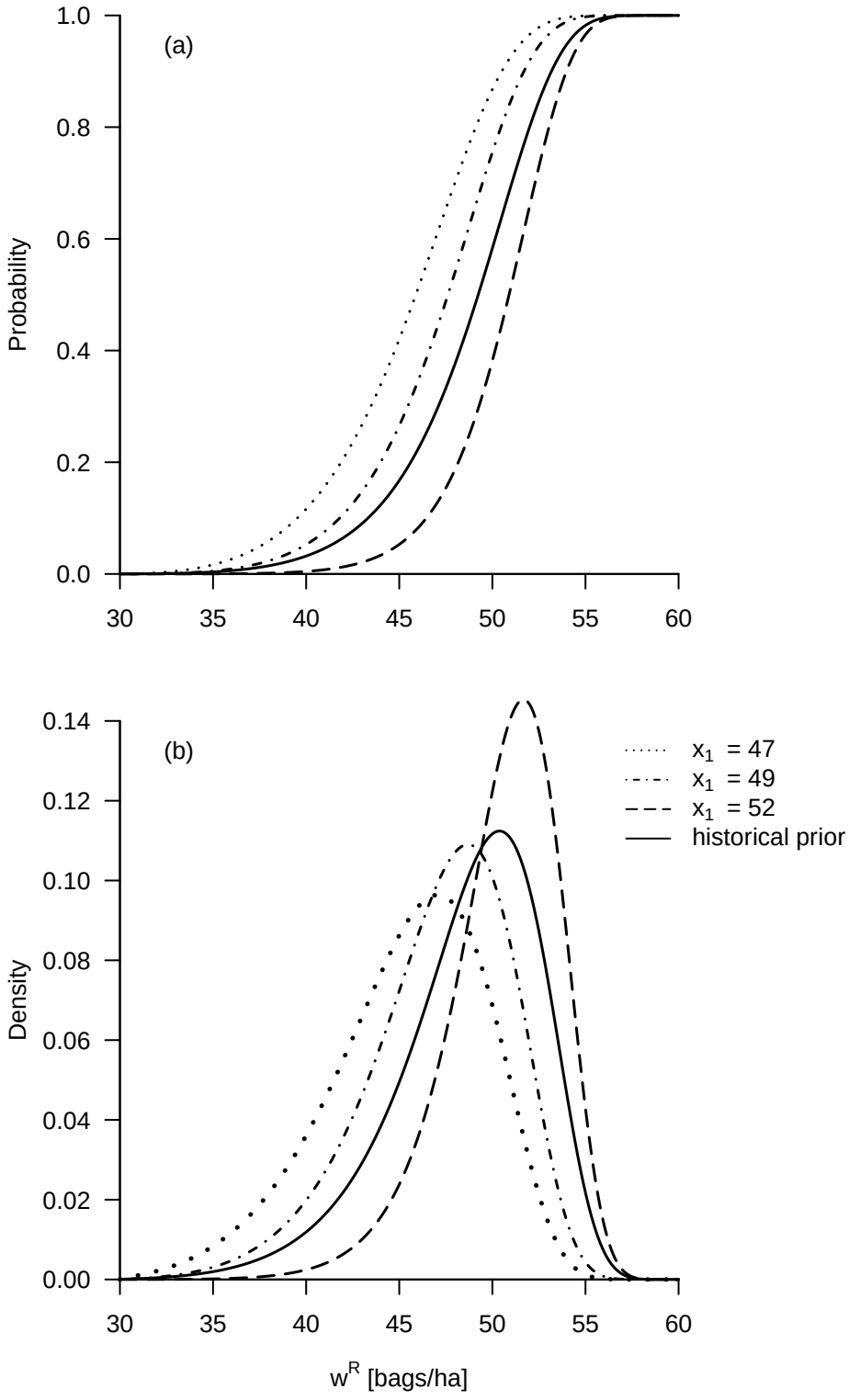


Figure 6.5 (a) Historical prior distribution function G, and posterior distribution function $\Phi(\cdot|x_1)$, given the deterministic forecast $x_1 = 47$, $x_1 = 49$, and $x_1 = 52$; (b) corresponding density functions $\phi(\cdot|x_1)$.

6.5.2 BPF - February

The BPF_{Feb} uses the deterministic forecasts issued by CONAB and IBGE in February as predictors. Usually the soybean crop fields are in an advanced stage of their cycle at this moment. The farmers and the forecasters have a better idea about the final yield and the deterministic forecasts are usually adjusted to significant changes in the production conditions, such as abnormal weather or diseases incidence.

Modeling the BPF_{Feb} follows the framework described in sub-section 3.2.4. The prior distribution function remains the same as the BPF_{Oct}, but the model in this section utilizes multiple predictors as mentioned. According to Maranzano (2006), modeling each Bayesian meta-Gaussian forecaster using multiple predictors involves the following steps:

Step 1. *Estimate the marginal distribution functions.* This step has been discussed in sections 6.3 and 6.4. Each BPF uses a different combination of predictors as shown in table 6.6. Therefore, the parametric models of the marginal distribution function are summoned as needed in the following sections.

Step 2. *Transform the variates using the NQT.* The procedure to apply the NQT is similar to the model developed for October.

Step 3. *Estimate the moments of variates in the transformed space.* The moments $\boldsymbol{\mu}_Q$ and $\boldsymbol{\Sigma}_Q$ are composed here. This step also includes computing the informativeness score (IS) and performing a predictor screening.

Step 4. *Validate meta-Gaussian dependence structure.* This step involves analyses whether the assumptions of the meta-Gaussian model hold.

Step 5. *Compute posterior parameters.* The posterior parameters along with the marginal distribution functions previously modeled are used to obtain the posterior distribution and density functions.

These steps are executed below in order to estimate and validate the BPF_{Feb}. The end results are the equations and parameter values of a posterior density, distribution, and quantile functions that can be used to produce probabilistic forecasts of the yield of Mato Grosso, given the deterministic forecasts of CONAB and IBGE in February.

Step 1. Estimate marginal distribution functions

The parametric model of the prior distribution function G is the same as in section 6.5.1, i.e., LC2–IW, as described in equations (6.1) and (6.2). This is the historical prior distribution function modeled using observed data of the yield of Mato Grosso.

The parametric model of the marginal distribution function $\bar{K}_2$ of X_2 is LC1–LL, as described in equations (6.5) and (6.6), and the model of function $\bar{K}_3$ of X_3 is LC2–IW, as described in equations (6.3) and (6.4). The parameter values for G, $\bar{K}_2$ and $\bar{K}_3$ are shown in tables 5.4 and 6.5, respectively.

Step 2. Transform the variates using the NQT

The parametric models described in the previous step were utilized to obtain the transformed joint samples $\{(z_2, v)\}$, $\{(z_3, v)\}$, and $\{(z_2, z_3)\}$ as follows:

$$V = Q^{-1}(G(W^R)), \tag{6.10}$$

$$Z_2 = Q^{-1}(\bar{K}_2(X_2)), \tag{6.11}$$

$$Z_3 = Q^{-1}(\bar{K}_3(X_3)). \tag{6.12}$$

Figures 6.6-a and 6.8-a show the transformed joint samples $\{(z_2, v)\}$ and $\{(z_3, v)\}$, respectively. Figure 6.10-a shows the joint sample $\{(z_2, z_3)\}$. These transformed samples will be used to estimate the likelihood parameters and validate the meta-Gaussian assumptions. These joint samples are shown in table 6.7.

Step 3. Estimate the moments of variates in the transformed space

The transformed samples obtained in the previous step were used to compute $\boldsymbol{\mu}_Q$ and $\boldsymbol{\Sigma}_Q$. The estimates calculated using the joint sample (X_2, X_3, V) of size $N = 12$ are

$$\boldsymbol{\mu}_Q = \begin{bmatrix} -0.274 \\ 0.580 \\ 0.576 \end{bmatrix}, \quad \boldsymbol{\Sigma}_Q = \begin{bmatrix} 2.345 & 0.560 & 0.994 \\ 0.560 & 0.217 & 0.296 \\ 0.994 & 0.296 & 0.607 \end{bmatrix}. \tag{6.13}$$

The IS for predictors X_2 and X_3 are

$$IS_2 = 0.889, \quad IS_3 = 0.876.$$

The informativeness score for X_2 is larger than the score of X_3. Therefore, the deterministic forecast issued by CONAB in February is more informative than the deterministic forecast issued by IBGE in the same month for forecasting the actual yield of Mato Grosso. This is an important conclusion considering the resources allocated in producing these estimates.

Step 4. Validate meta-Gaussian dependence structure

According to Maranzano (2006), validating the meta-Gaussian dependence structure involves checking the assumption that vector (Z_2, Z_3, V) is multivariate Gaussian. One of the approaches to validate this assumption is based on the necessary condition that if the vector (Z_2, Z_3, V) is multivariate Gaussian, all the pairs of variates in this vectors are bivariate Gaussian.

Figures 6.7 and 6.9 show the scatterplots and QQ plots of the residuals of the regression of Z_2 on V and Z_3 on V, respectively. The analysis of the residuals of the linear regression of Z_2 on V do not indicate clear signs of heteroscedasticity. However, the QQ

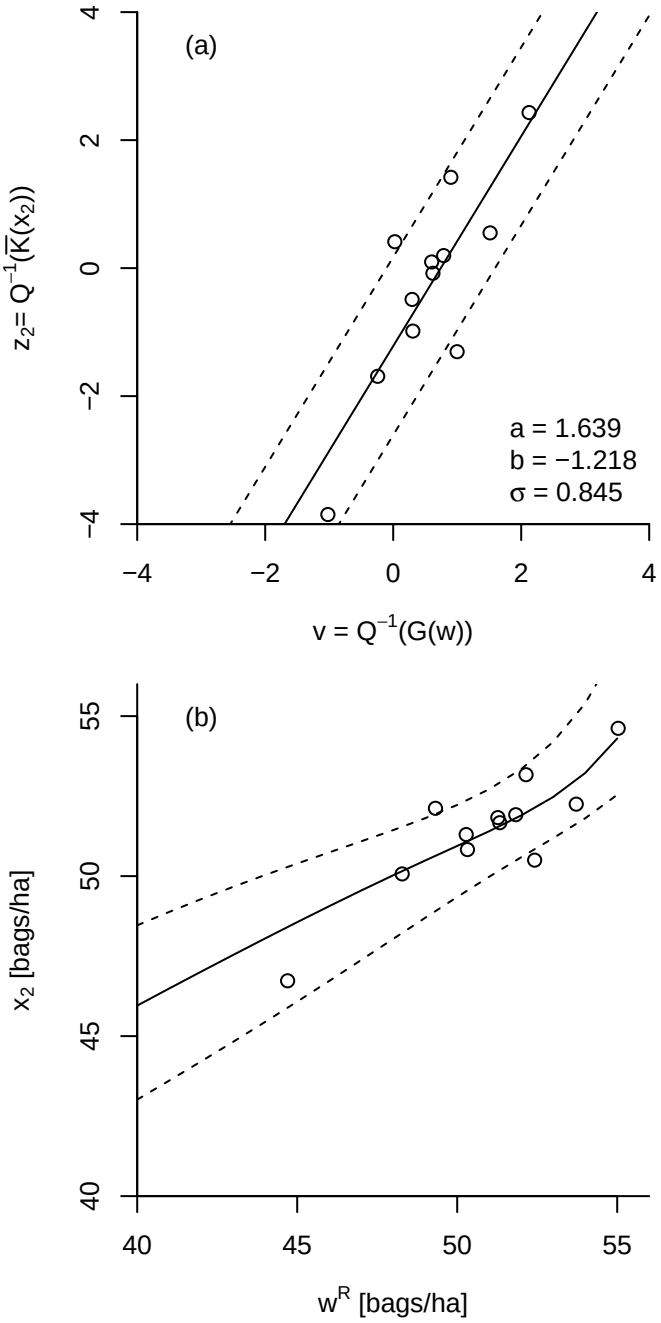


Figure 6.6 (a) Linear regression of Z_2 on V, and 90% central credible interval; (b) Bayesian meta-Gaussian median regression of the deterministic forecast X_2 issued by CONAB in February on the regional yield W^R.

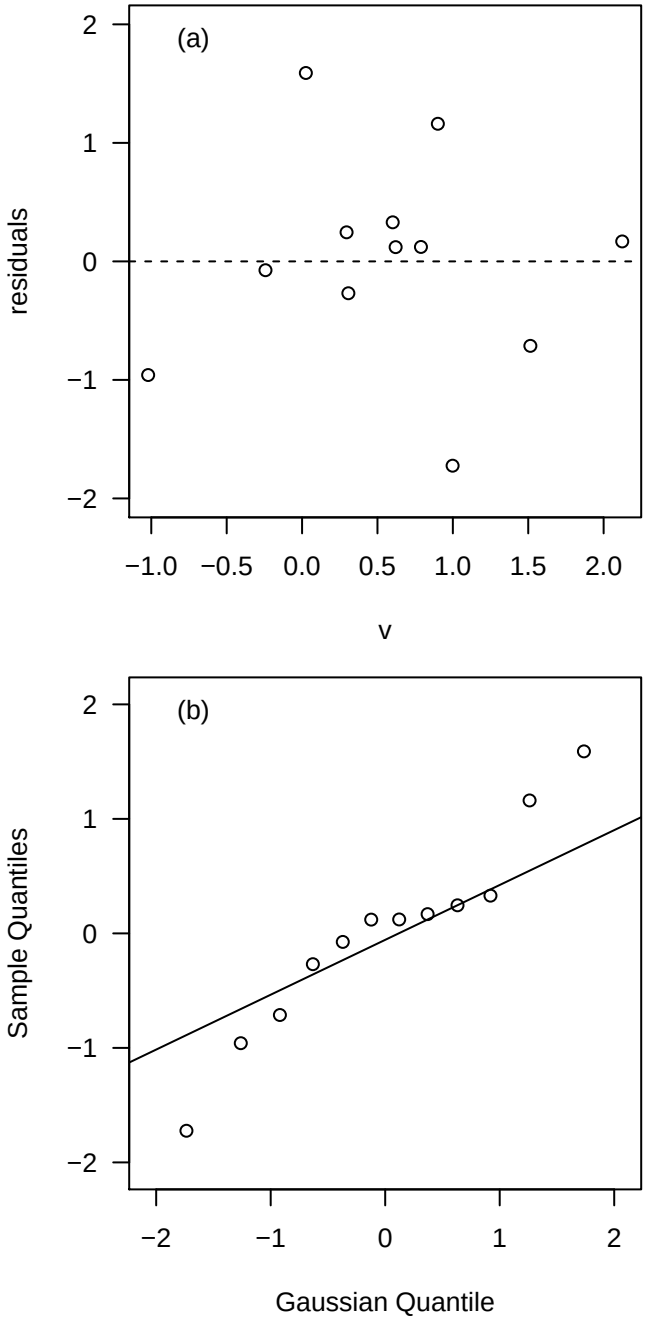


Figure 6.7 (a) Residuals from the linear regression of Z_2 on V; (b) QQ plot of the residuals constructed using the meta-Gaussian plotting positions.

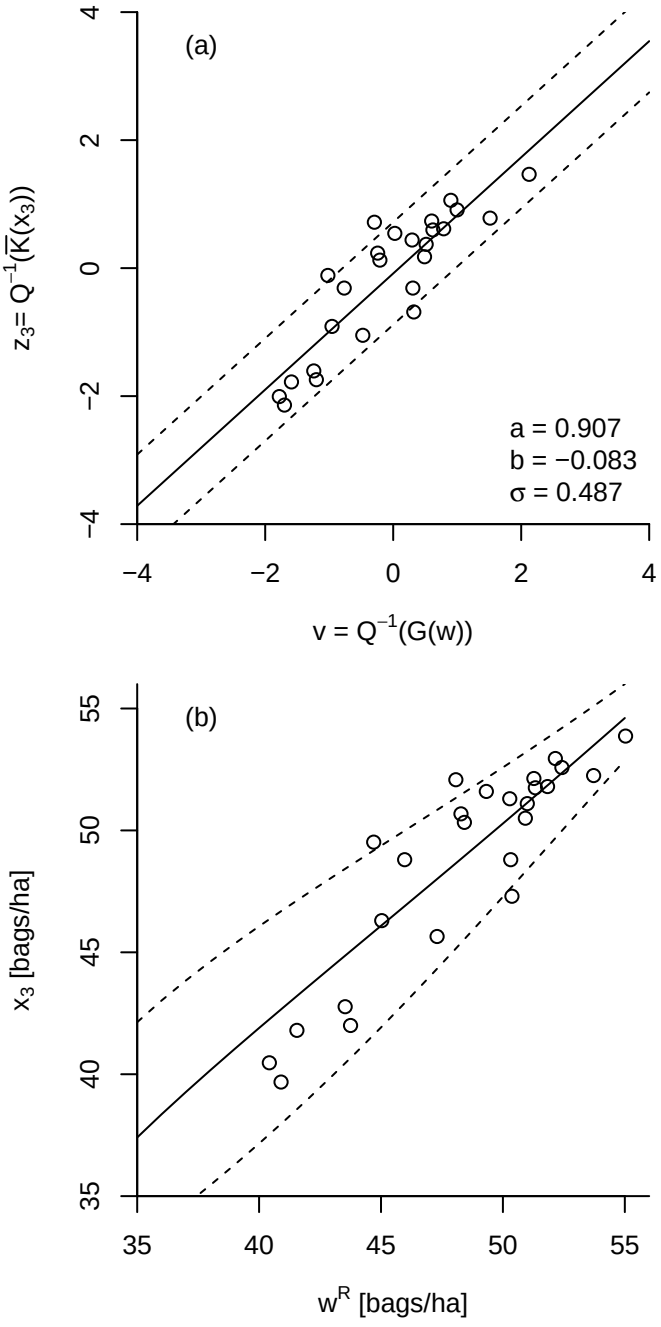


Figure 6.8 (a) Linear regression of Z_3 on V, and 90% central credible interval; (b) Bayesian meta-Gaussian median regression of the deterministic forecast X_3 issued by IBGE in February on the regional yield W^R.

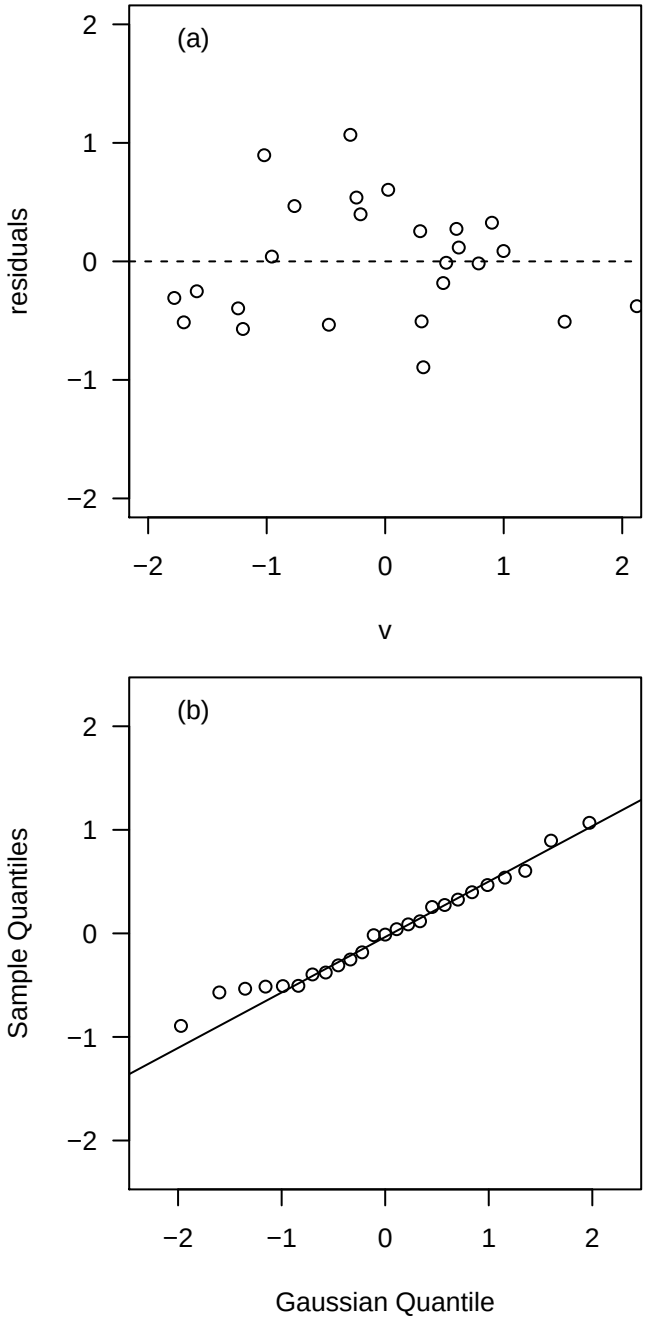


Figure 6.9 (a) Residuals from the linear regression of Z_3 on V; (b) QQ plot of the residuals constructed using the meta-Gaussian plotting positions.

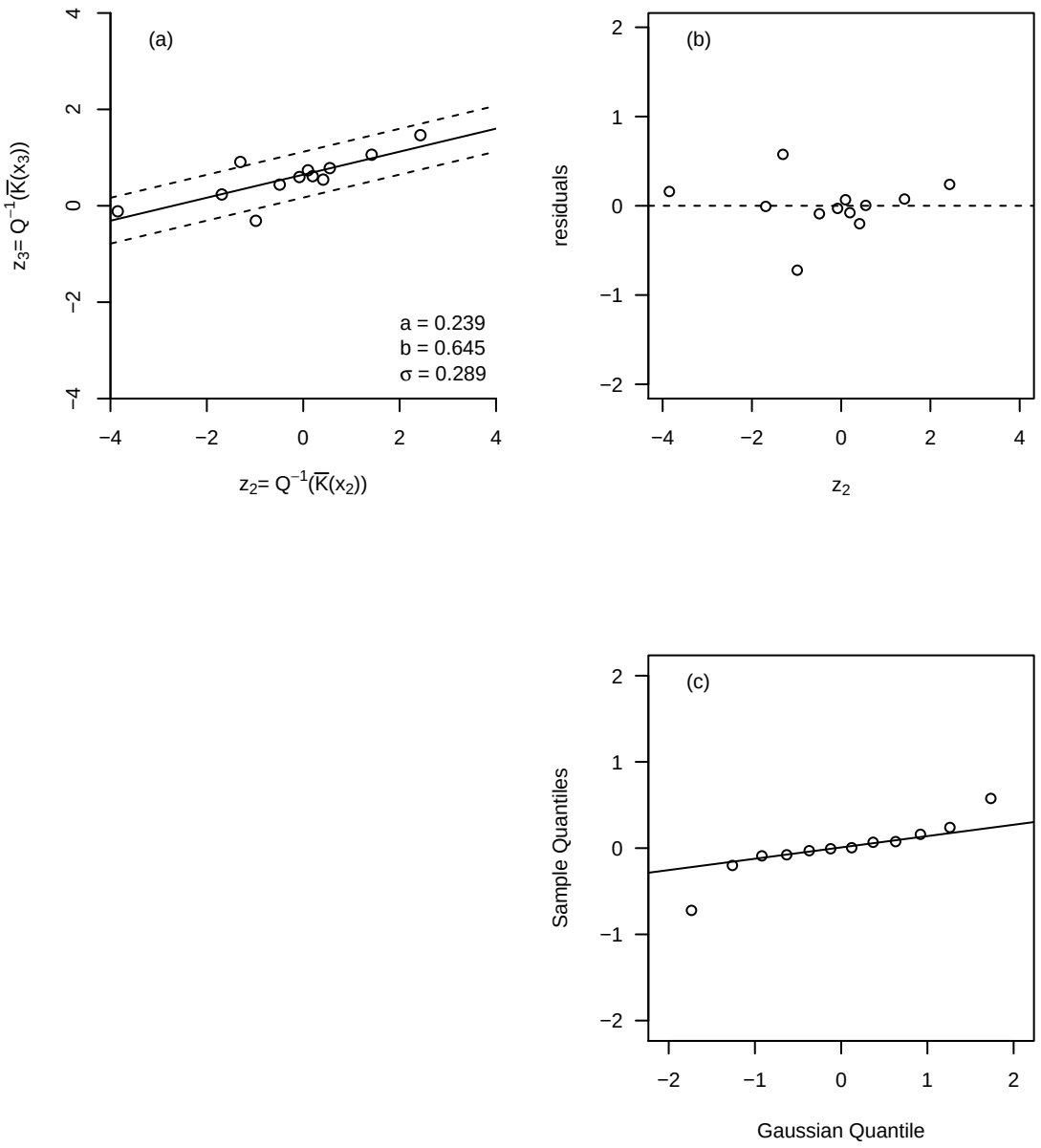


Figure 6.10 (a) Linear regression of Z_3 on Z_2, and 90% central credible interval; (b) residuals from the linear regression; (c) QQ plot of the residuals constructed using the meta-Gaussian plotting positions.

plot indicates the possibility of a heavy-tailed distribution. The plots in figure 6.9 suggest homoscedasticity and Gaussianity for the residuals of the linear regression of Z_3 on V.

Figure 6.10 shows the linear regression of Z_3 on Z_2 as well as the scatterplot and QQ plot of the residuals to validate the regression. Visually, there is no clear evidence that the residuals from this linear regression are heteroscedastic. The QQ plot indicates the possibility of a heavy-tailed distribution, but close to Gaussian in the central region. This analysis can become more explicit once the sample size of $\{(z_2, z_3)\}$ increases. At this moment, this research considers that the assumptions related to the meta-Gaussian dependence structure hold, but this analysis can be improved by collecting data over time.

Step 5. Compute posterior parameters

The estimates of the posterior parameters are

$$c_2 = 0.282, \quad c_3 = 0.778,$$
$$c_0 = 0.111, \quad T = 0.398.$$

These estimates were obtained using equations 3.21 and they were applied to equations 3.23, 3.24, and 3.25 to obtain the posterior distribution, density, and quantile functions, respectively. The posterior parameters along with the equations for the prior distribution function and the marginal distribution functions of X_2 and X_3 are sufficient to construct the BPF_{Feb}.

Figure 6.11 shows examples of posterior density functions $\phi(\cdot|x_2, x_3)$. For instance, Figure 6.11-a shows the case with $x_2 = 50$ and $x_3 = 45$ and the corresponding posterior density function $\phi(\cdot|x_2 = 50, x_3 = 45)$. In other words, it represents the situation when CONAB forecasts an yield of 50 bags per hectare and IBGE forecasts an yield of 45 bags per hectare in February for the state of Mato Grosso. The decision maker can use the posterior density function associated with those realizations and produce probabilistic forecasts of the yield of Mato Grosso.

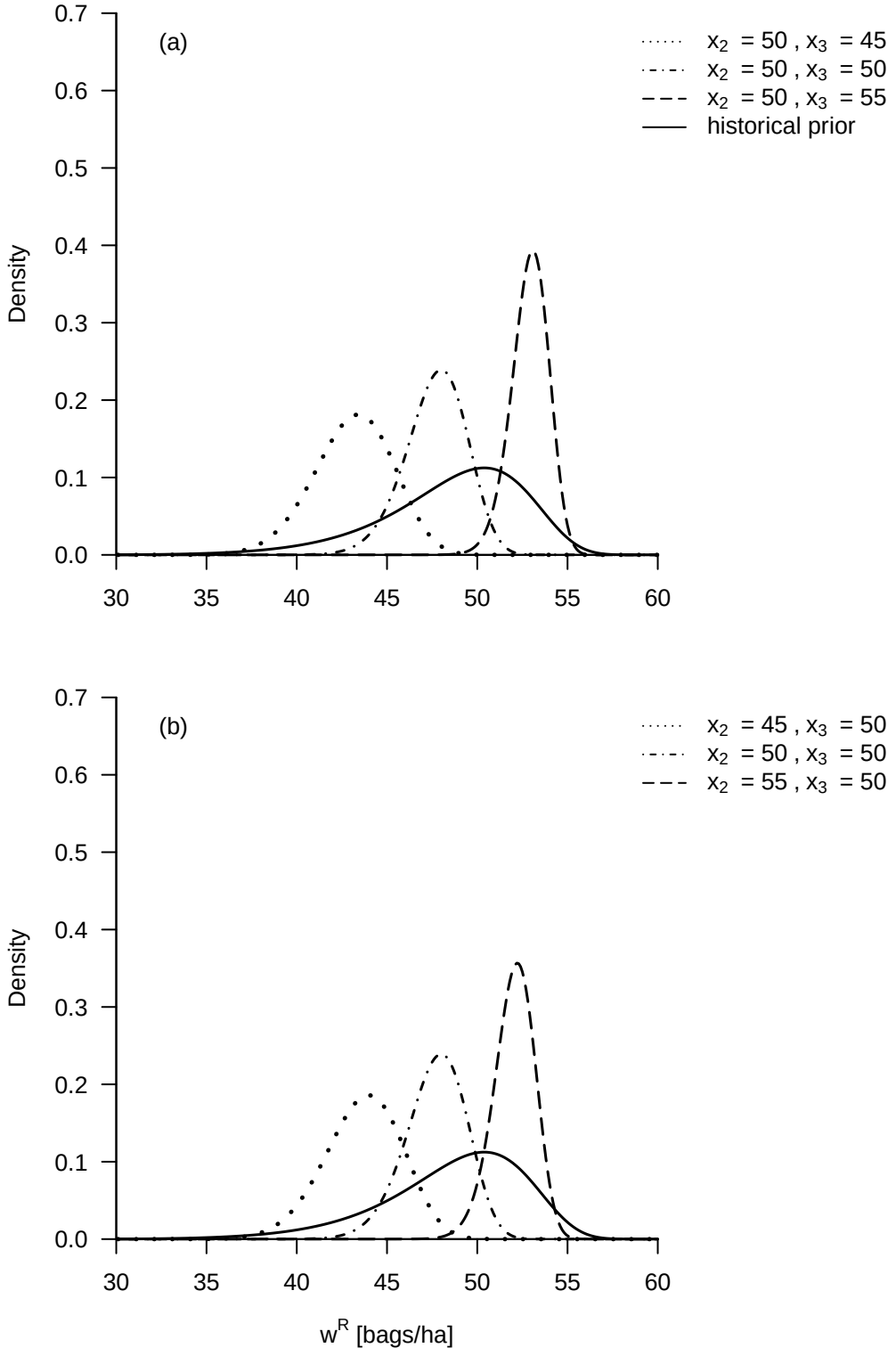


Figure 6.11 (a) Historical prior density function g and posterior density functions $\phi(\cdot|x_2 = 50, x_3)$ for $x_3 = 45, 50,$ and 55; (b) posterior density functions $\phi(\cdot|x_2, x_3 = 50)$ for $x_2 = 45, 50,$ and 55.

6.5.3 BPF - May

The BPF_{May} is constructed using the same framework applied in the BPF_{Feb}, a meta-Gaussian model using multiple predictors. In this case, the predictors are the deterministic forecasts issued by CONAB (X_4) and IBGE (X_5) in May for the yield of the state of Mato Grosso. Intuitively, these deterministic forecasts should carry less uncertainty about the actual realization of yield , since they are released with a shorter lead time. Yet having probabilistic forecast at this time can bring an economic value to the decision maker, since many marketing activities are still being done throughout the year.

The following steps are executed to model the BPF_{May}:

Step 1. Estimate marginal distribution functions

Once again, the parametric model of the prior distribution function G is the same as in section 6.5.1. The parametric models of the marginal distribution functions $\bar{K}_4$ of X_4 and $\bar{K}_5$ of X_5 are LC2–IW, as described in equations (6.3) and (6.4). The parameter values for G, $\bar{K}_4$ and $\bar{K}_5$ are shown in tables 5.4 and 6.5, respectively.

Step 2. Transform the variates using the NQT

The transformed joint samples $\{(z_4, v)\}$, $\{(z_5, v)\}$, and $\{(z_4, z_5)\}$ are obtained by applying the NQT as follows:

$$V = Q^{-1}(G(W^R)), \tag{6.14}$$

$$Z_4 = Q^{-1}(\bar{K}_4(X_4)), \tag{6.15}$$

$$Z_5 = Q^{-1}(\bar{K}_5(X_5)). \tag{6.16}$$

Figures 6.12-a, 6.14-a, 6.16-a and shows the transformed joint samples $\{(z_4, v)\}$, $\{(z_5, v)\}$, $\{(z_4, z_5)\}$, respectively.

Step 3. Estimate the moments of variates in the transformed space

The estimates of $\boldsymbol{\mu}_Q$ and $\boldsymbol{\Sigma}_Q$ using the joint sample (X_4, X_5, V) of size $N = 11$ are

$$\boldsymbol{\mu}_Q = \begin{bmatrix} 0.052 \\ 0.763 \\ 0.721 \end{bmatrix}, \quad \boldsymbol{\Sigma}_Q = \begin{bmatrix} 1.576 & 0.701 & 0.774 \\ 0.701 & 0.342 & 0.369 \\ 0.774 & 0.369 & 0.409 \end{bmatrix}. \tag{6.17}$$

The IS for predictors X_4 and X_5 are

$$IS_4 = 0.985, \quad IS_5 = 0.994.$$

Predictor X_5 is slightly more informative than X_4. In fact, the IS of both predictor are quite close to 1, which would mean they are close to being perfect. Intuitively, this shall come as no surprise, since IBGE and CONAB update their estimates monthly according to variations in the production conditions up to the end of the season. Therefore, forecasts with shorter lead time will likely be closer to the actual realization of yield.

Step 4. Validate meta-Gaussian dependence structure

Figures 6.13, 6.15, and 6.16 show the scatterplots and QQ plots of the linear regressions of Z_4 on V, Z_5 on V, and Z_5 on Z_4, respectively. The analysis of figures 6.13 and 6.16 indicates that the vectors (Z_4, V) and (Z_4, Z_5) are bivariate Gaussian. However, The QQ plot in figure 6.15-b suggest a heavy-tailed distribution for the residuals of the regression of Z_5 on V. Once again, the small sample size in this case provides little conclusive arguments to invalidate this assumption, but this analysis should be revised once more data are available. Therefore, this research concludes that there is no clear evidence that the vector (Z_4, Z_5, V) is not multivariate Gaussian.

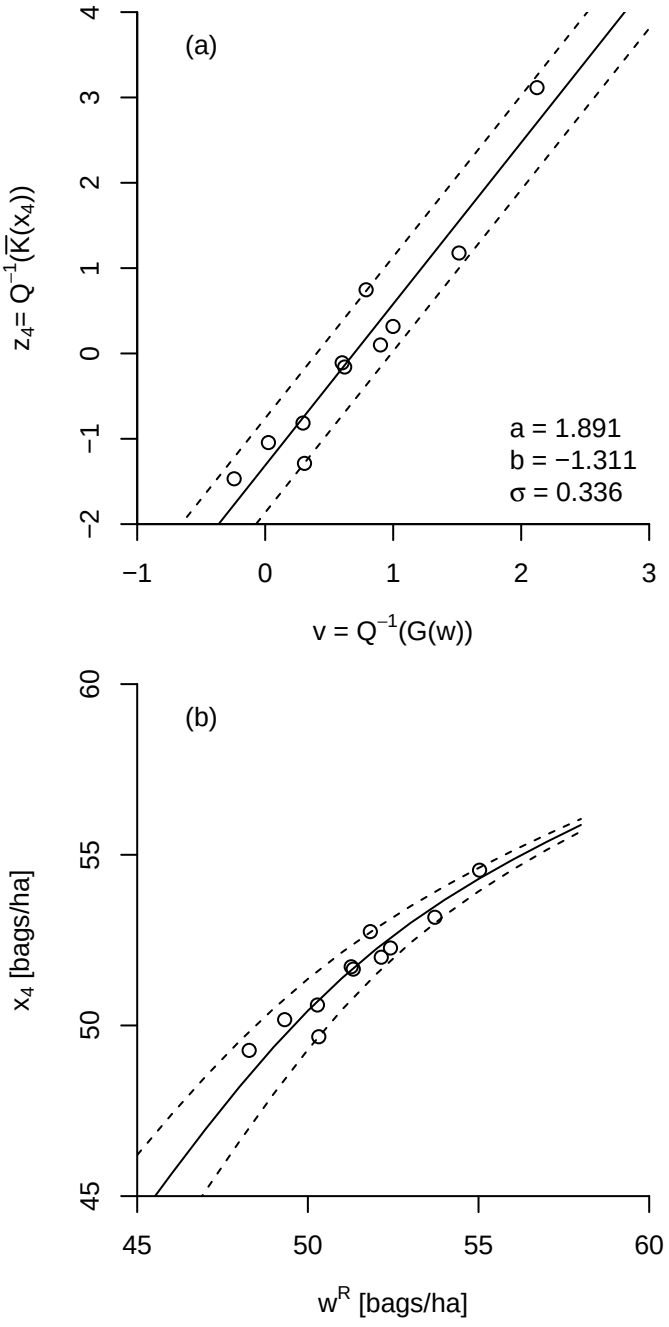


Figure 6.12 (a) Linear regression of Z_4 on V, and 90% central credible interval; (b) Bayesian meta-Gaussian median regression of the deterministic forecast X_4 issued by CONAB in May on the regional yield W^R.

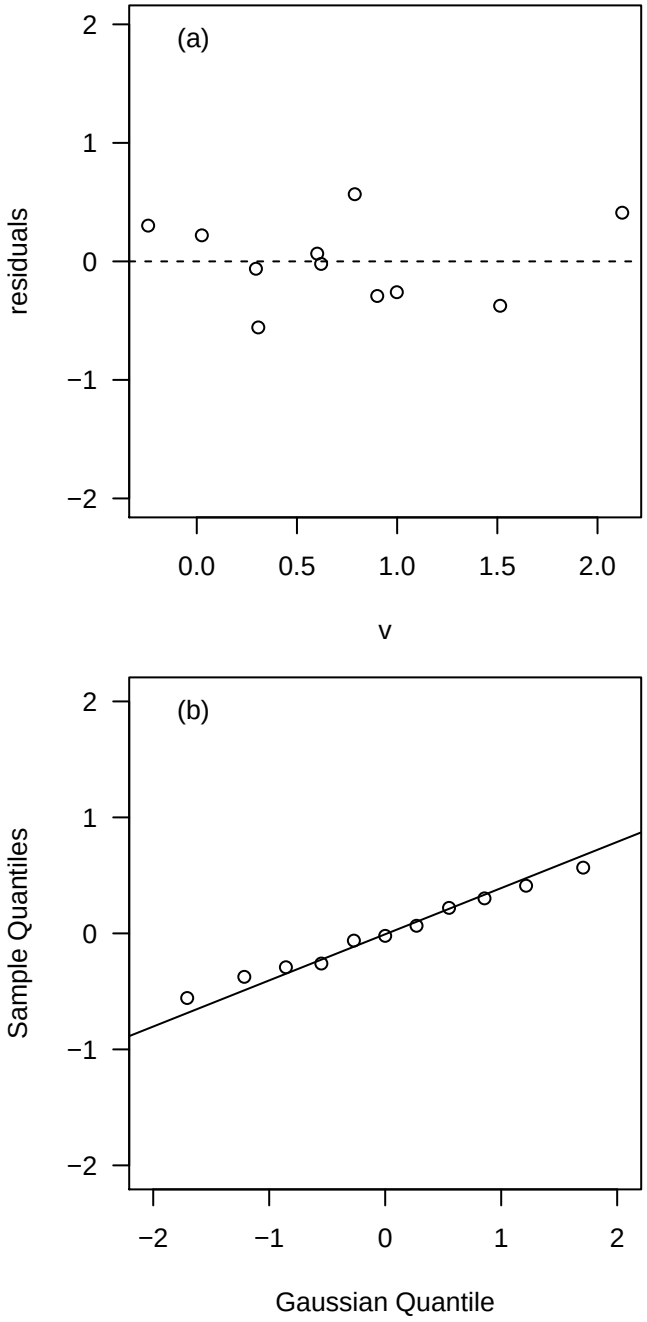


Figure 6.13 (a) Residuals from the linear regression of Z_4 on V; (b) QQ plot of the residuals constructed using the meta-Gaussian plotting positions.

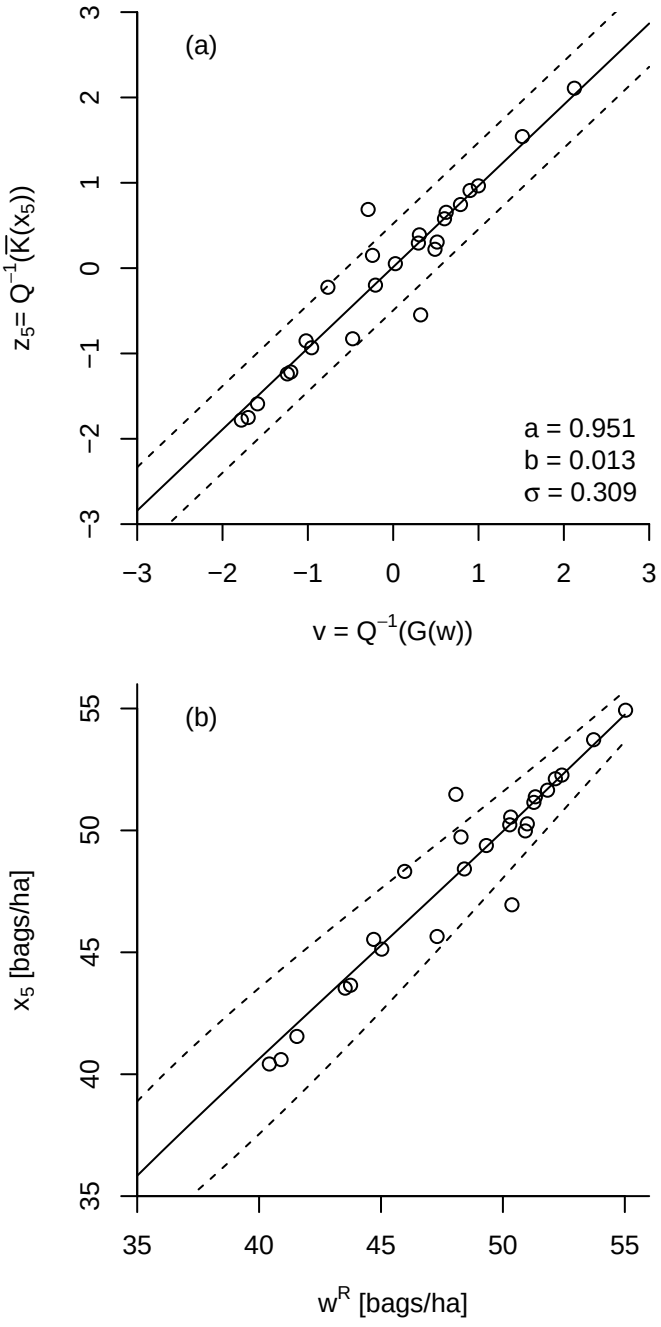


Figure 6.14 (a) Linear regression of Z_5 on V, and 90% central credible interval; (b) Bayesian meta-Gaussian median regression of the deterministic forecast X_5 issued by IBGE in May on the regional yield W^R.

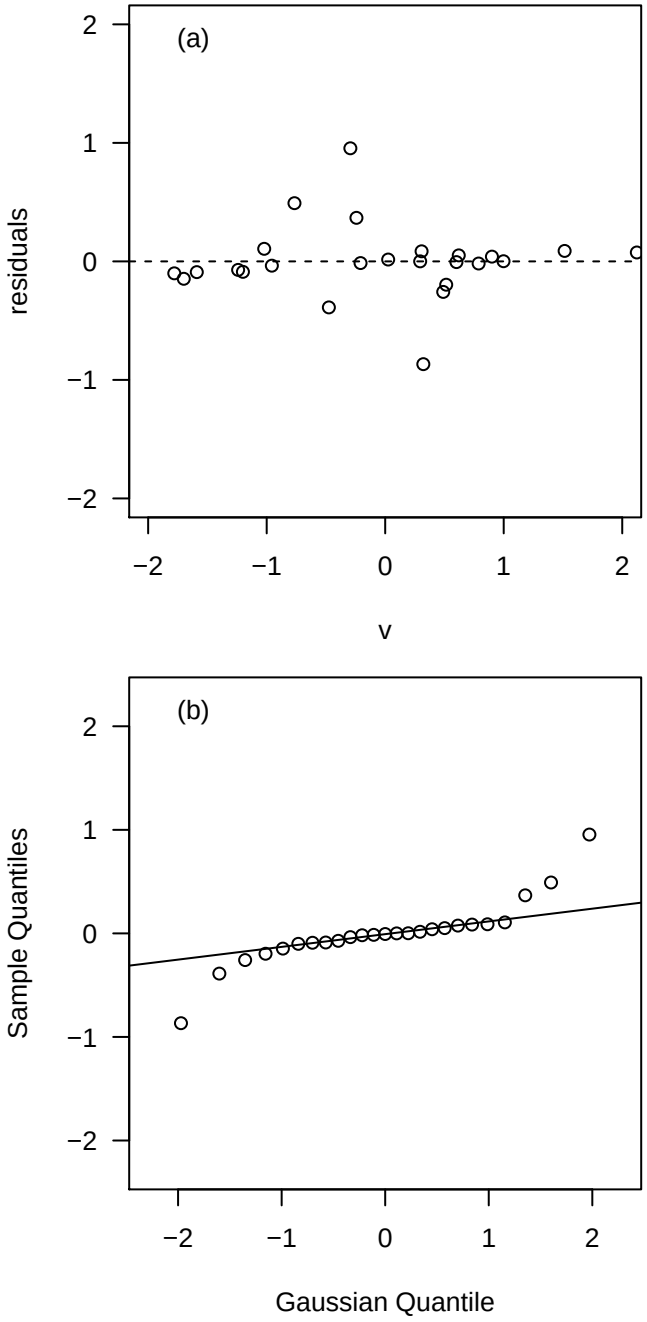


Figure 6.15 (a) Residuals from the linear regression of Z_5 on V; (b) QQ plot of the residuals constructed using the meta-Gaussian plotting positions.

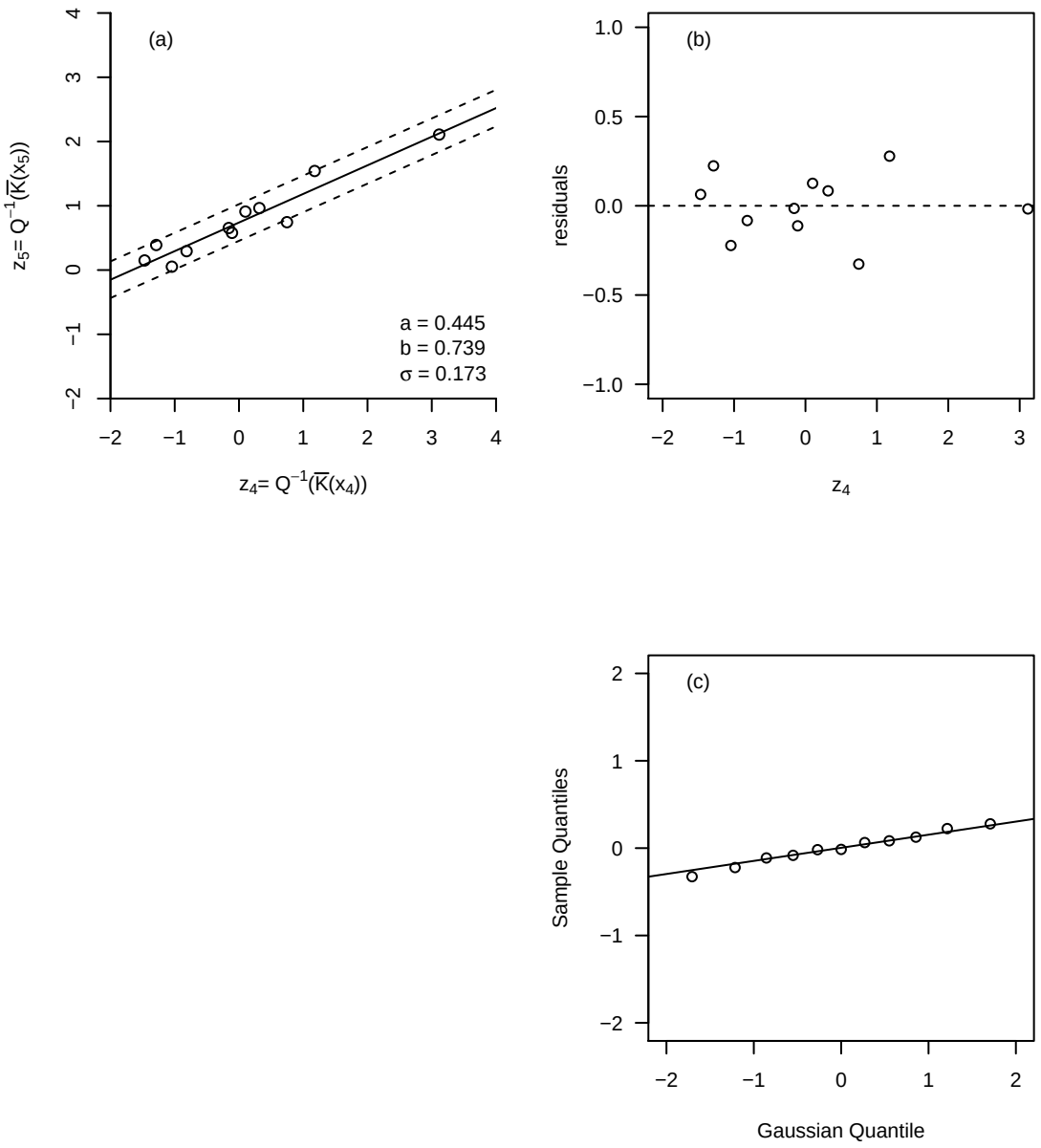


Figure 6.16 (a) Linear regression of Z_5 on Z_4, and 90% central credible interval; (b) residuals from the linear regression; (c) QQ plot of the residuals constructed using the meta-Gaussian plotting positions.

Step 5. Compute posterior parameters

The estimates of the posterior parameters are

$$c_4 = 0.128, \quad c_5 = 0.829,$$
$$c_0 = 0.075, \quad T = 0.096.$$

The posterior parameters, the parametric models of the prior distribution function G and of the marginal distribution functions $\bar{K}_4$ and $\bar{K}_5$ are sufficient to construct the posterior density, distribution, and quantile functions of W^R, conditional on the realization of the vector of predictors (X_4, X_5). Figure 6.17 shows examples of posterior density function, given different realizations of X_4 and X_5.

The posterior density functions are quite narrower than the prior density function, indicating a considerable decrease in uncertainty from the prior to the posterior functions. Both predictors are very informative, but changes in x_5 result in a slightly greater change in the mode and spread of the posterior density functions, when compared to changes in x_4.

6.5.4 Summary

The BPFs constructed so far use the historical prior information. These models decrease the uncertainty associated with the prior information at different times in the soybean season. Each model has a different economic value due to different lead times and the types of decisions that the grower must make at various stages of the season.

These models address the problem of having different deterministic forecasts from different sources regarding the same variate, in this case, the yield of Mato Grosso. The same framework can be applied to different regions, agricultural crops, predictors, and lead times. For instance, an analyst could merge additional deterministic forecasts purchased from private companies in this model and analyze whether these forecasts decrease uncertainty about the actual yields.

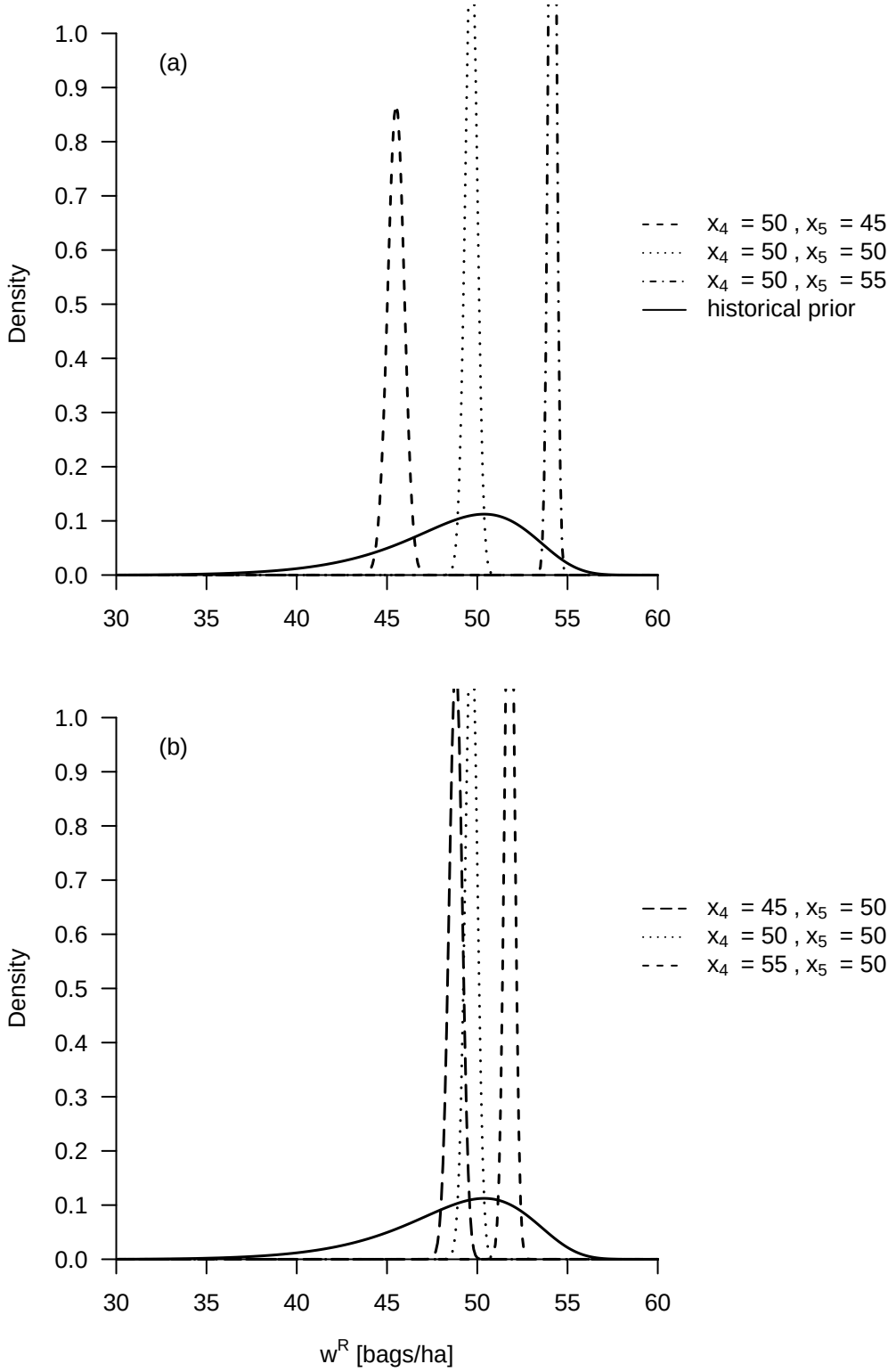


Figure 6.17 (a) Historical prior density function g and posterior density functions $\phi(\cdot|x_4 = 50, x_5)$ for $x_5 = 45, 50,$ and 55; (b) posterior density functions $\phi(\cdot|x_4, x_5 = 50)$ for $x_4 = 45, 50,$ and 55.

6.6 Example of Real Forecast

The models developed so far are able to produce probabilistic forecasts of the yield of Mato Grosso using the deterministic forecasts of IBGE and CONAB. The observed data used was recorded up to 2017. Let us analyze the same variate during the year of 2018. Table 6.9 shows the deterministic forecasts issued by CONAB and IBGE at different times in the crop season. The first forecast was issued in October by CONAB and the observed yield was 55.8 bags per hectare.

The organizations clearly updated their forecasts over time and got closer to the actual observation in September. However, many agents had to make decisions based on the October, February, and May forecasts. The difference between the CONAB October forecast and the IBGE September estimate was of 4.7 bags per hectare. This is a significant difference considering the size of certain farms. Figure 6.18 shows the posterior distribution and density functions of each BPF, conditional on the realizations of the predictors in table 6.9.

Suppose a grower wants to know the probability of the actual yield being above the 51.1 bags per hectare forecasted by CONAB in October. According to the historical prior, the probability of this event is 0.29, while the posterior distribution function of the BPF_{Oct}, conditional on the realization $X_1 = 51.1$, tell us that the probability of the same event is 0.33. The BPF_{Oct} slightly decreased the uncertainty about W^R.

The BPF_{Feb} and BPF_{May}, considerably decreased the uncertainty about W^R. Suppose

Table 6.9 Deterministic forecasts and actual realization of the soybean crop yield, in bags per hectare, in Mato Grosso in the 2017/2018 crop season issued by IBGE and CONAB .

	2017	2018	2018	2018
	Oct	Feb	May	Sep
CONAB	51.1	53.6	55.8	-
IBGE	-	54.5	55.9	55.8

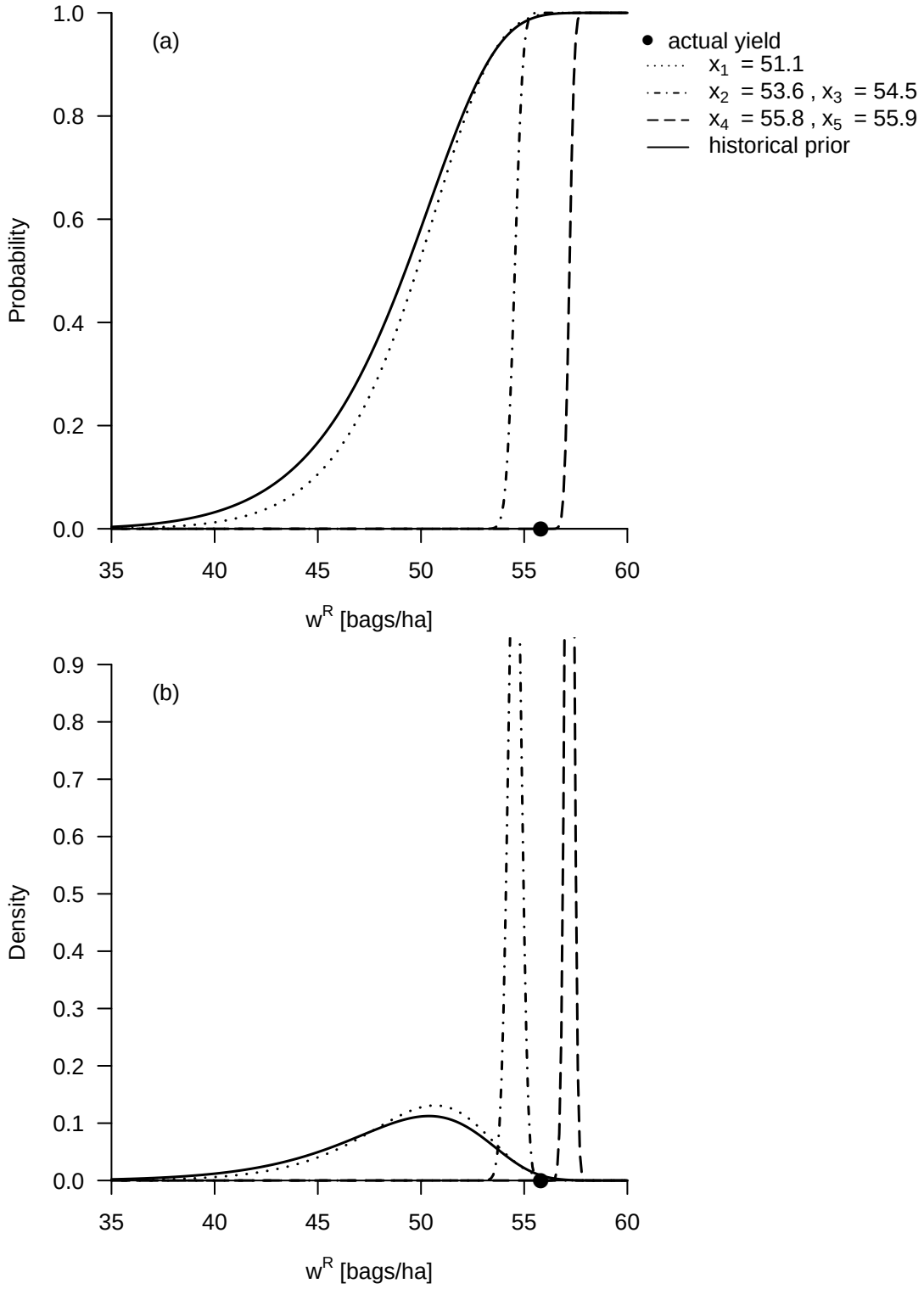


Figure 6.18 (a) Historical prior distribution function G and posterior distribution functions $\Phi(\cdot|x_1 = 51.1)$, $\Phi(\cdot|x_2 = 53.6, x_3 = 54.5)$, and $\Phi(\cdot|x_4 = 55.8, x_5 = 55.9)$; (b) corresponding posterior density functions.

Table 6.10 Quantiles $w^R(p)$ of the yield of Mato Grosso from the historical prior distribution and posterior quantiles $w^R(p|x_1 = 51.1)$, $w^R(p|x_2 = 53.6, x_3 = 54.5)$, and $w^R(p|x_4 = 55.8, x_5 = 55.9)$, for $p = 0.01, 0.25, 0.5, 0.75, 0.9$.

Function	Forecast	0.01	0.25	0.5	0.75	0.9
Historical Prior	$w^R(p)$	37.1	46.4	49.2	51.5	53.2
Posterior Quantile	$w^R(p\|x_1 = 51.1)$	39.5	47.4	49.8	51.7	53.2
Posterior Quantile	$w^R(p\|x_2 = 53.6, x_3 = 54.5)$	53.8	54.3	54.5	54.8	54.9
Posterior Quantile	$w^R(p\|x_4 = 55.8, x_5 = 55.9)$	56.7	57.1	57.2	57.3	57.3

a grower wants to make probabilistic forecasts after receiving the reports from CONAB and IBGE in February. At this moment, the development of the crop seasons is almost at the end and the forecasters at these organizations have a better idea regarding the final estimations of yield.

The IBGE forecasted 54.5 and CONAB forecasted 53.6 bags per hectare in February. Growers can produce probabilistic forecasts using the posterior quantile function in order to have a more illustrative set of probabilities. Table 6.10 shows values of five quantiles. Given the deterministic forecasts in February, for instance, farmers can use table 6.10 to have a comprehensive set of forecasts and their respective probabilities.

According to CONAB (2018), the soybean yield in Mato Grosso for the 2017/2018 broke the historical record due to large increase in the planted area, optimum planting timing, favorable weather conditions, and improvements in seed technology. In fact, it was considered to be an extreme event above the 0.9-quantile of the February forecast. However, the BPF_{May} adjusted these quantiles after the May forecasts were released.

While improvements in technology are persistent, optimal weather conditions do not repeat year after year. Many factors can drive the yield down, such as an increase in disease incidence, drought, and so on. The BPF model updates the probabilistic forecasts based on the deterministic forecasts issued by these organizations. The same analysis can be made in the following years, using new deterministic forecasts.

6.7 Summary

This chapter developed models to produce probabilistic forecasts of agricultural yield. The main difference between the models was the lead time. The deterministic forecasts used as predictors were issued in October, February, and May. During the month of October, the growers in Brazil are starting to prepare for planting. In February, the season is at an advanced stage. The growers have observed the weather conditions and other risk factors that can significantly influence the yield. In May, the harvest season is almost over.

The Bayesian forecasters developed in this chapter quantify the uncertainty related to the actual soybean yield of Mato Grosso and the deterministic forecasts issued by CONAB and IBGE at different lead times. The model merges information from these two sources in order to produce the probabilistic forecast. This addresses the current problem of credibility associated with these two sources. Instead of choosing a deterministic forecast to use among these sources, growers can use the BPF models to take advantage of both estimates.

Moreover, the probabilistic forecasts provide not only a forecast, conditional on the realization of a deterministic forecast, but also a quantification of the uncertainty associated with that forecast. Growers can incorporate this important information in their decision making throughout the development of the crop season. The real example in subsection 6.6 shows an application of the BPF models in a peculiar crop season, the 2017/2018. Although the yield of this season was a historical record, the posterior functions were able to decrease uncertainty about the yield of Mato Grosso.

Table 6.7 reports the transformed observations obtained from the NQT for reference. The calculations can be reproduced and the results can be validated using this table. The next chapter explores the sensitivity of the models developed here to type of information, the judgmental prior distributions.

7. Sensitivity to Judgmental Priors

This chapter continues to explore the problem of probabilistic forecasting of agricultural yield, but using judgmental prior distribution functions. These judgmental functions were obtained from growers and they are described in chapters 4 and 5. First, a model to forecast the yield of a field is developed, followed by models to forecast the yield of a region using different types of prior distributions.

7.1 Overview

Yield forecasts are often released to the public as point estimates. The Bayesian processor of forecasts (BPF) previously implemented in chapter 3 provides a convenient solution to quantify the existing uncertainty of deterministic yield forecasts. Moreover, the expertise of farmers regarding their own fields and region is integrated into modeling in order to provide additional information for the forecast.

The construction of the BPF models follows the framework exposed in chapter 3. The BPF models differ according to the type of prior distribution function used. The judgmental prior distribution functions are modeled in chapter 4, and the stochastic transformation utilized to map the field quantiles into region quantiles is discussed in chapter 5. The next section provides a guideline to produce probabilistic forecasts of the yield of a field and it analyzes the impact of judgmental prior information. The third section analyzes the sensitivity of the models developed in chapter 6 to judgmental prior information.

7.2 Forecasting Local Yield Using Judgmentally Assessed Prior

Suppose a grower keeps records of his production, planted area, and yield for a number of fields within his property. He has a good understanding about the dynamics of the

production in his various fields and he wishes to forecast the yield of a specific field. Many growers can produce point estimates of future yield of a field based on the inputs applied, weather conditions, disease incidence, and so on. This section provides a guideline to take a step further and produce probabilistic forecasts using historical records of the yield and their judgmental prior distribution functions.

Let the continuous variate W^F with sample space $\mathcal{W}^F$ be the *predictand* — the net harvested yield of the soybean crop in a field, in bags per hectare . Its realization is denoted $w^F \in \mathcal{W}^F$, where $\mathcal{W}^F = \{w^F : 0 < w^F < \infty\}$. Let the continuous variate X_1 with sample space $\mathcal{X}_1$ be the *predictor* — a deterministic forecast of the yield of the region where the field is located, in bags per hectare . Its realization is denoted $x_1 \in \mathcal{X}_\infty$, where $\mathcal{X}_1 = \{x_1 : 0 < x_1 < \infty\}$.

In order to illustrate this model, a hypothetical example was designed with real data as proxies for the yield of a field. The yield of the city of Sorriso was considered as a proxy of the yield of a certain field. These yield values were estimated by PAM/IBGE and the data set contains records from 1987 to 2017. The deterministic forecasts are also issued by the PAM/IBGE for the microregion of Alto Teles Pires, where the city of Sorriso is located. These forecasts are from 2000 to 2017. Assume that these estimates represent forecasts issued in February. Table 7.1 shows the observations of both variates.

In practice, growers can replace these observations with their own records of yield and desired predictor. The sample sizes need not be the same. The example created in this section considers a time series of W^F longer than X_1. The Bayesian forecaster is able to use all data as opposed to using only the joint sample. The methodology applied in this section is described in chapter 3.

It is worth mentioning that the sample of W^F in table 7.1 seems to be nonstationary. The values of w^F appear to be nonstationary in the mean, with a lower mean for the first part of the time series, perhaps until 1995, and a higher mean after that. Figure 7.1 shows the time series graph of w^F and moving averages calculated with different orders.

Table 7.1 Joint samples $\{(x_1, w^F)\}$ and $\{(z_1, v)\}$.

year	x_1	w^F	z_1	v
1987	na	38.0	na	-1.53
1988	na	33.8	na	-2.00
1989	na	39.2	na	-1.40
1990	na	31.0	na	-2.46
1991	na	39.1	na	-1.41
1992	na	43.0	na	-0.97
1993	na	40.6	na	-1.25
1994	na	43.4	na	-0.92
1995	na	37.0	na	-1.63
1996	na	45.0	na	-0.72
1997	na	46.3	na	-0.54
1998	na	46.0	na	-0.58
1999	na	48.3	na	-0.22
2000	51.9	55.0	0.41	1.55
2001	51.9	54.3	0.41	1.27
2002	51.4	52.0	-0.01	0.55
2003	51.5	51.0	0.07	0.31
2004	48.3	52.0	-1.48	0.55
2005	52.7	52.0	0.83	0.55
2006	48.1	50.0	-1.55	0.09
2007	50.4	51.0	-0.66	0.31
2008	53.2	52.0	1.08	0.55
2009	51.7	52.0	0.26	0.55
2010	51.1	49.7	-0.23	0.03
2011	56.1	58.0	2.18	3.39
2012	52.6	54.0	0.80	1.16
2013	50.6	52.1	-0.54	0.58
2014	49.9	52.4	-0.90	0.66
2015	51.0	52.5	-0.27	0.69
2016	49.7	48.0	-0.99	-0.27
2017	56.4	58.0	2.27	3.39
sample mean	51.6	47.6	0.1	0.0
sample sd	2.12	6.88	1.05	1.33

Source: PAM/IBGE.
*na = not available

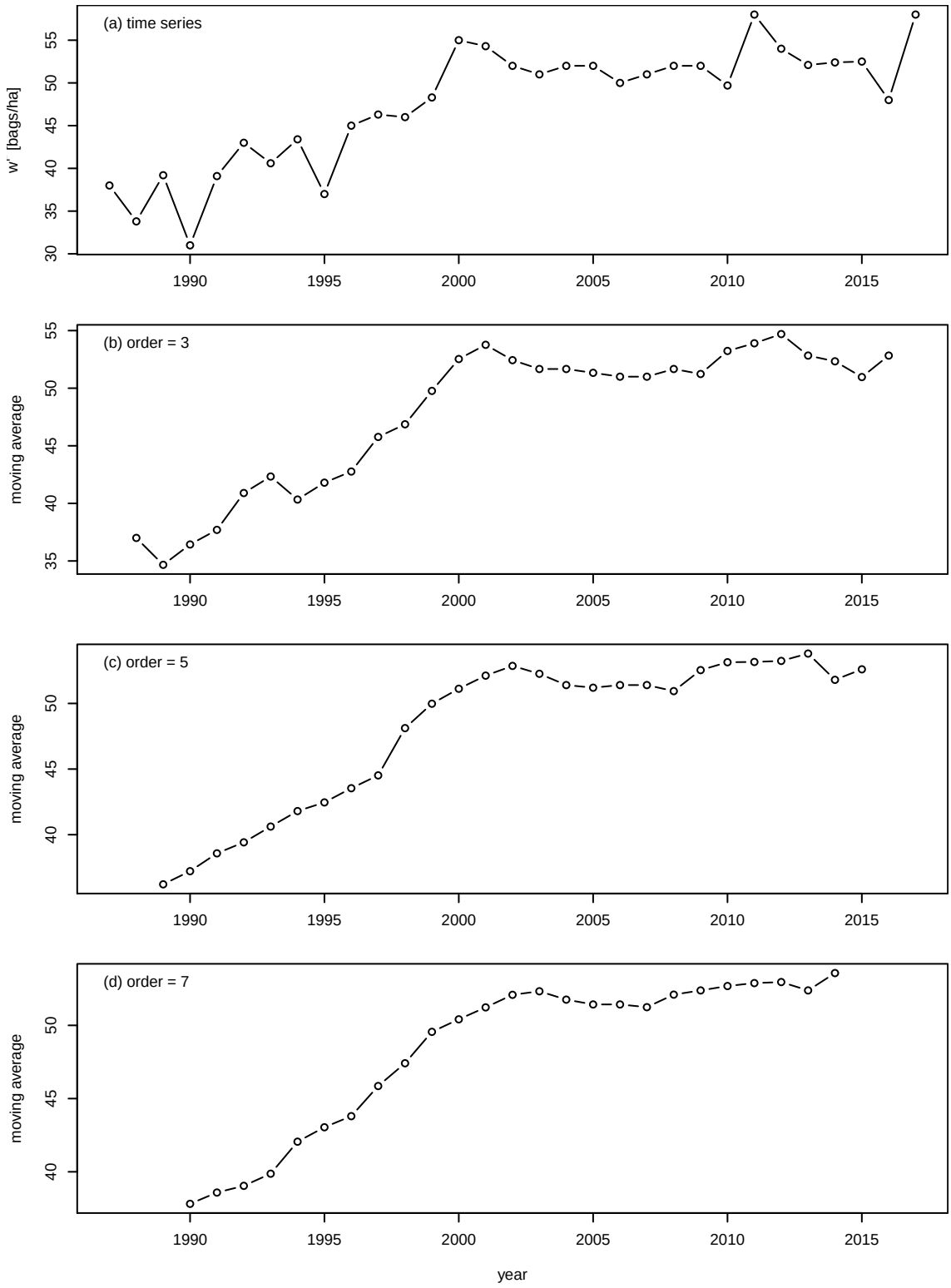


Figure 7.1 (a) Time series plot of w^F, (b) moving average plot of order 3, (c) moving average plot of order 5, and (d) moving average plot of order 7.

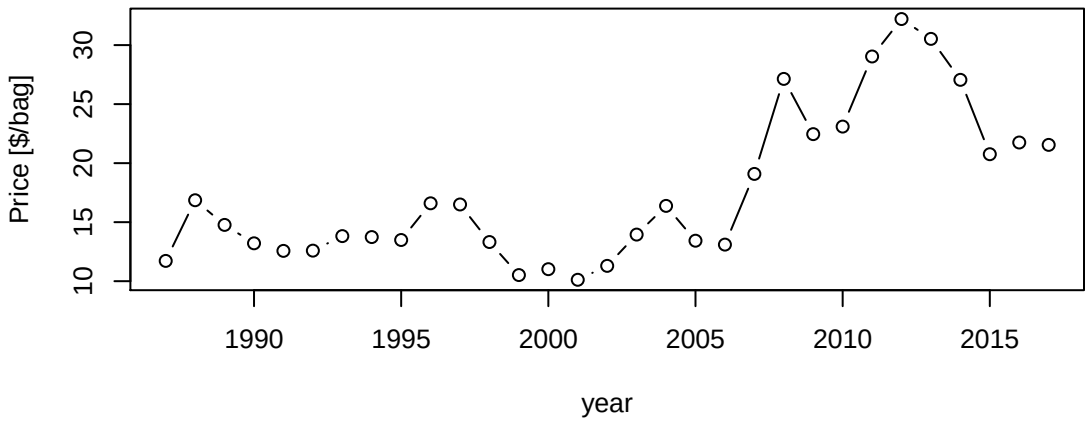


Figure 7.2 Historical annual average international soybean prices from 1987 to 2017.
Source: macrotrends.net

The graphs in figure 7.1 indicate that the average yield increases until approximately the year of 2000 and then it stabilizes for the remaining years. In real applications, this nonstationarity may pose an obstacle to modeling the prior distribution function. It could be argued that the values of w^F follow different distributions before and after 2000.

Upon further investigation, the 90s were reported to be a time of expansion of the soybean crop in Brazil. The ascending profitability in the first half of that decade motivated the expansion of the cultivated area. However, in order to understand the increase in the yields, one must analyze the international soybean prices in figure 7.2.

The good international prices in 1996 and 1997 had a positive impact on the growers' profit. This fact allowed them to take action in the following years when the international price fell considerably. According to the LSPA reports, the growers invested heavily in new technologies to reduce risk associated with their exposure to international prices. This scenario may explain the jump in the yields in the 90s.

The entire sample is taken for the hypothetical example to show how samples with different sizes can be utilized in this approach. In a more thorough analysis, the nonstationarity of this time series must be addressed possibly by transforming the data set into a stationary time series.

Table 7.2 Parameter values of the historical prior distribution function of W^F and the marginal distribution function of X_1.

Function	Distribution	α	β	η_L	η_U	MAD	K-S stat	Critical Value
G	LC2–IW	0.5052	2.5402	30	65	0.0867	0.129	0.188
$\bar{K}$	LR1–LP	0.1783	0.4485	30	65	0.0576	0.096	0.244

The historical prior distribution function G of W^F modeled using the data in table 7.1 is Log-reciprocal type II inverted Weibull (LC2–IW)

$$G(w^F|\alpha,\beta) = \exp\left[-\left(\frac{\alpha}{y}\right)^{\beta}\right], \tag{7.1}$$

where

$$y = ln\frac{\eta_U - \eta_L}{w^F - \eta_L}, \tag{7.2}$$

and η_L and η_U are the lower and upper bounds, respectively. The parameter values are in table 7.2. The MAD suggests adequate to good fit and the K-S statistic does not reject this distribution function at a 20% significance level. Therefore, this function is good for G of W^F.

The marginal distribution $\bar{K}$ of X_1 is LR1–LP

$$\bar{K}(x_1) = \begin{cases} \frac{1}{2}\exp\left(\frac{y-\beta}{\alpha}\right), & \text{if } y \le \beta, \\ 1 - \frac{1}{2}\exp\left(-\frac{y-\beta}{\alpha}\right), & \text{if } \beta \le y, \end{cases} \tag{7.3}$$

where

$$y = ln\frac{x_1 - \eta_L}{\eta_U - x_1}. \tag{7.4}$$

The parameter values for this distribution function are in table 7.2. Figures 7.3 and 7.4 show the empirical and parametric distributions functions of W^F and X_1.

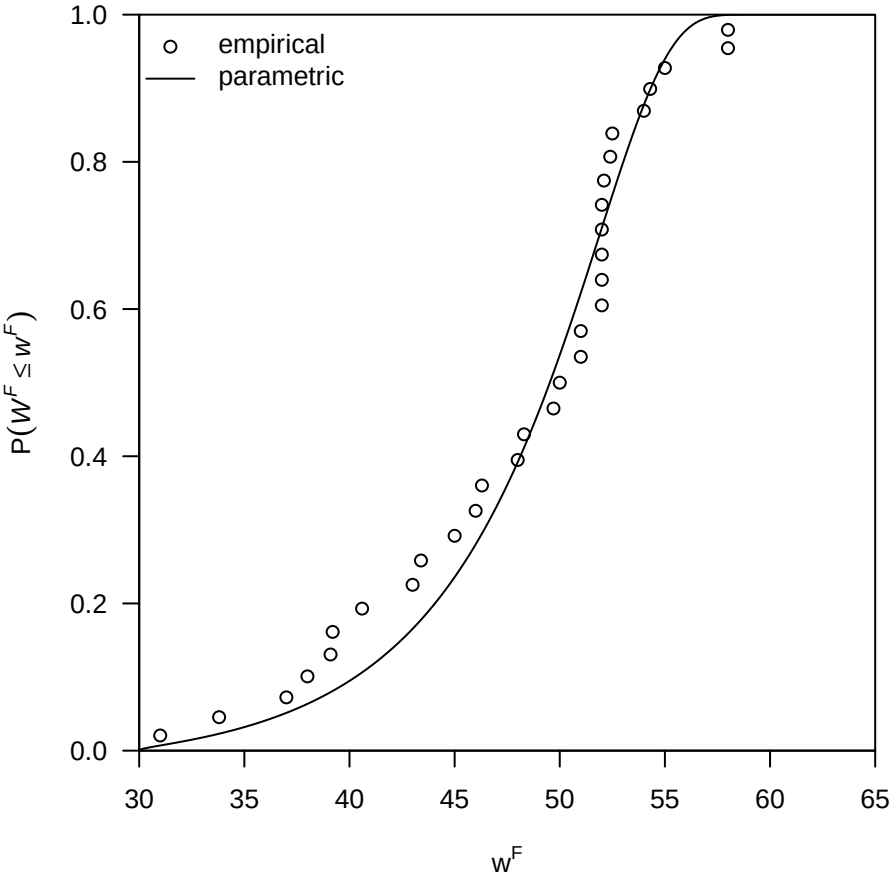


Figure 7.3 Empirical and parametric distribution functions of W^F.

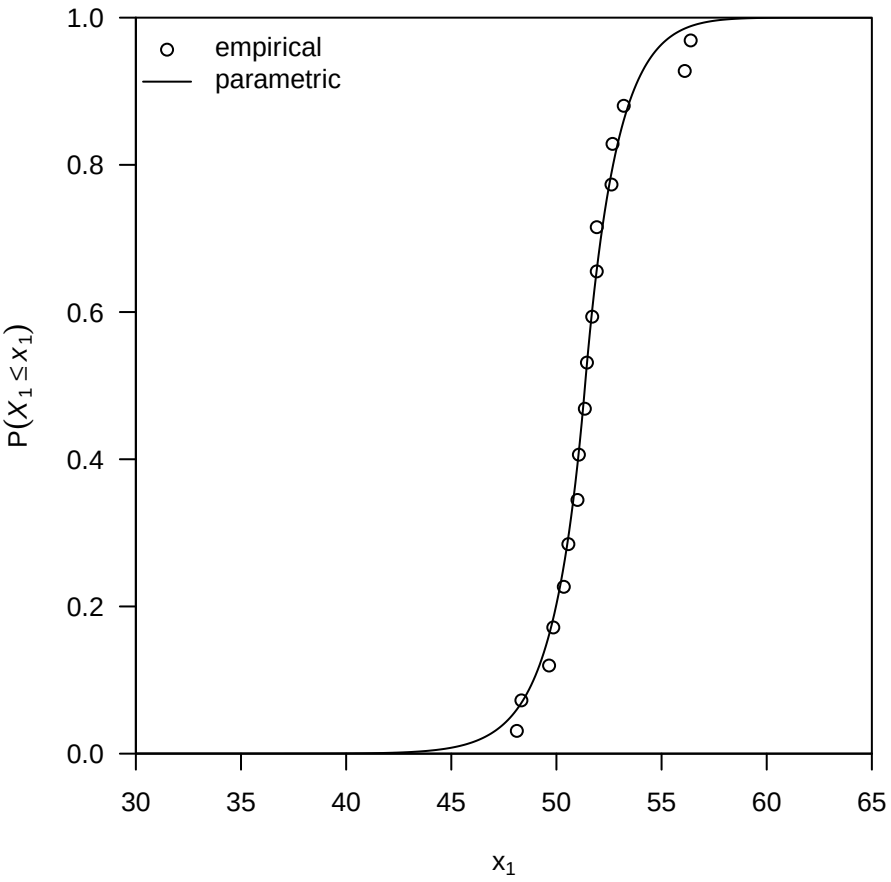


Figure 7.4 Empirical and parametric distribution functions of X_1.

Equations (7.5) and (7.6) are used to transform the original joint sample $\{(x_1, w^F)\}$ into $\{(z_1, v)\}$. Similar to the previous models developed, the NQT equations are:

$$V = Q^{-1}(G(W^F)), \tag{7.5}$$

$$Z_1 = Q^{-1}(\bar{K}_1(X_1)). \tag{7.6}$$

Table 7.1 shows the joint sample $\{(z_1, v)\}$ and figure 7.5-a shows the scatterplot of this joint sample along with the linear regression estimated using the least squares. The residuals of this regression are shown in figure 7.6-a and figure 7.6-b shows the QQ plot of these residuals. There is no clear indication of heteroscedasticity in the residuals.

The estimates of posterior parameters of the Bayesian meta-Gaussian model are

$$c_1 = 0.507, \quad c_0 = 0.759, \quad \text{and } T = 0.589.$$

These posterior parameters along with marginal distribution function of X_1 and the prior distribution function of W^F are sufficient to derive the family of posterior density and distribution functions. Figure 7.7 shows examples of posterior distribution and density functions, conditional on certain realizations of X_1. Suppose IBGE issues a deterministic forecast of $x_1 = 51$ for Alto Teles Pires (the microregion) and the grower wants to calculate the probability of $w^F > 48$. According to his historical records, this probability is equal to 0.61. However, using the posterior distribution function, conditional on $x_1 = 51$, the grower can take advantage of the deterministic forecast and update his uncertainty about the yield of this field. The posterior probability for this event is 0.83. Much higher than the historical records alone could predict.

The BPF model constructed so far utilizes only the historical record, but suppose the grower wants to incorporate his judgmental assessment about the yield of the current year. Chapter 4 and 5 discuss methods to assess and model the uncertainty of growers.

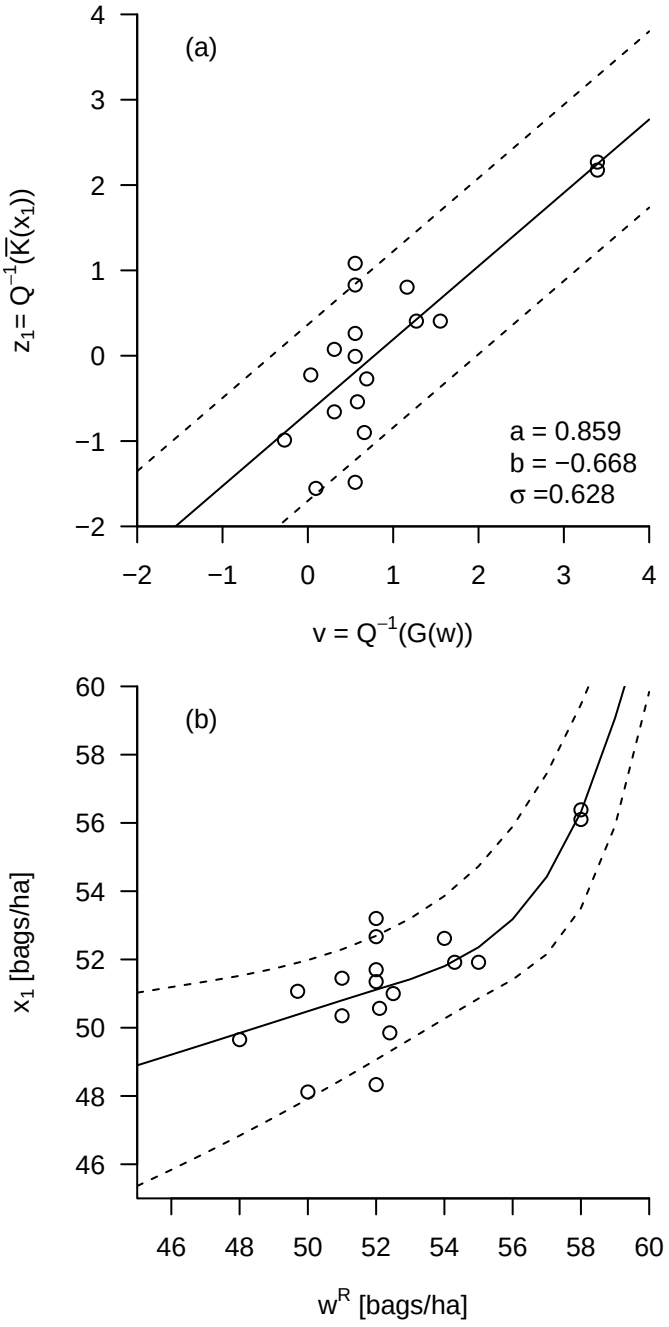


Figure 7.5 (a) Linear regression of Z_1 on V, and 90% central credible interval; (b) Bayesian meta-Gaussian median regression of the deterministic forecast X_1.

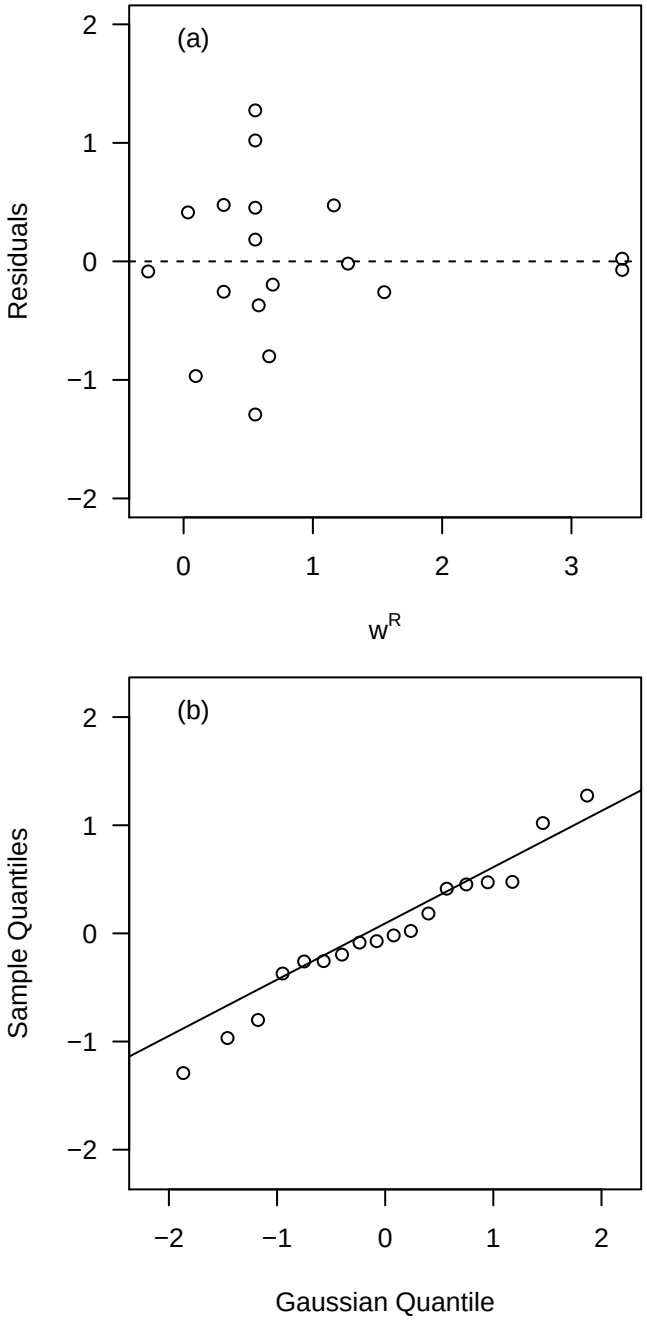


Figure 7.6 (a) Residuals from the linear regression of Z_1 on V; (b) QQ plot of the residuals constructed using the meta-Gaussian plotting positions.

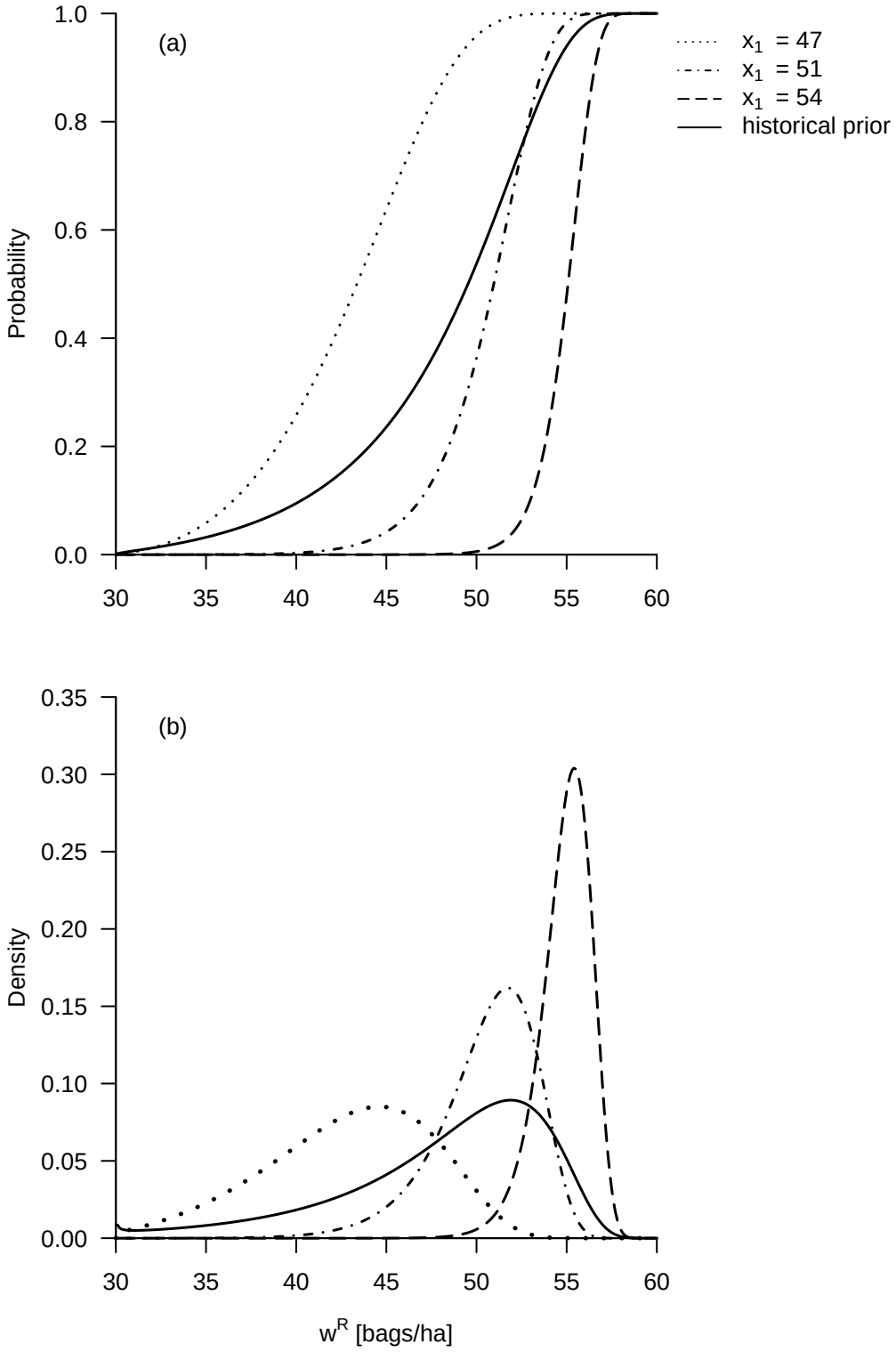


Figure 7.7 (a) Historical prior distribution function G, and posterior distribution function $\Phi(\cdot|x_1)$, given the deterministic forecast $x_1 = 47$, $x_1 = 51$, and $x_1 = 54$; (b) corresponding density functions $\phi(\cdot|x_1)$.

Table 7.3 Parameter values of the judgmental prior distribution function of W^F.

Function	Distribution	α	β	η_L	η_U	MAD
G_T^F	LC1–LL	0.4287	4.0827	30	65	0.0212

In this example, suppose the judgmental prior distribution function G_T^F assessed from the grower that owns the field is Log-reciprocal type I log-logistic (LC1–LL)

$$G_T^F(w^F|\alpha,\beta) = \left[1 + \left(\frac{y}{\alpha}\right)^{-\beta}\right]^{-1}, \tag{7.7}$$

where

$$y = ln\frac{\eta_U - \eta_L}{\eta_U - w^F}, \tag{7.8}$$

and the parameter values are shown in table 7.3

The posterior functions can be updated to incorporate this judgmental distribution function according to the methodology developed in section 3.3. Figure 7.8 shows a comparison between the historical and judgmental functions as well as the updated posterior functions of w^F, conditional on realizations $X_1 = x_1$. In this example, the grower seems to be confident that lower values of yield in year T are more likely to happen. For instance, the median in the historical prior is approximately 49.5 bags per hectare, while in his judgment, the median would be approximately 42.2 bags per hectare for year T. His assessments are pessimistic for the current year.

The updated posterior functions in figure 7.8 are quite different from the posterior functions in figure 7.7. The changes in the updated posterior functions reflect the judgmental assessments from the grower. For instance, the posterior probability of $w^F > 48$, conditional on $x_1 = 51$, was 0.83. The same event has probability of 0.64 according to the updated functions.

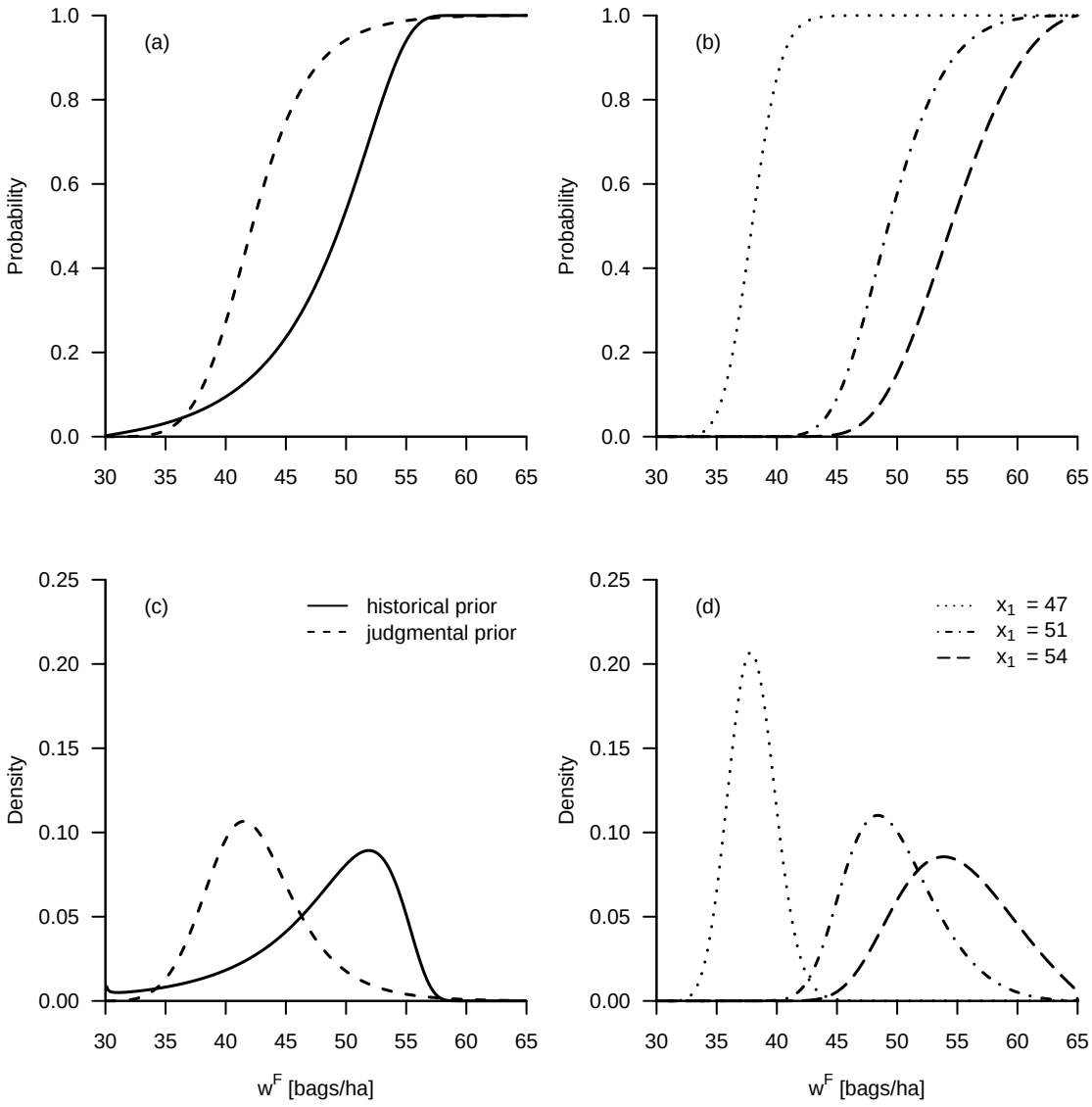


Figure 7.8 (a) Historical prior distribution function G and judgmental prior distribution function G_T^F, (b) posterior distribution functions $\Phi(\cdot|x_1)$, given the deterministic forecast $x_1 = 47$, $x_1 = 51$, and $x_1 = 54$, (c) corresponding historical and judgmental prior density functions, and (d) corresponding posterior density functions $\phi(\cdot|x_1)$.

7.3 Forecasting Regional Yield Using Judgmentally Assessed Prior

This section returns to the problem of forecasting the yield of a region. Suppose an analyst wants to forecast the soybean crop yield of the state of Mato Grosso in Brazil. The models developed in chapter 6 use exclusively observed data as inputs to the Bayesian forecaster. The models constructed in this section adapt the framework of the previous models to incorporate judgmental prior distribution functions.

Let the continuous variate W^R with sample space $\mathcal{W}^R$ be the *predictand* — the net harvested yield of the state of Mato Grosso, in bags per hectare . Its realization is denoted $w^R \in \mathcal{W}^R$, where $\mathcal{W}^R = \{w^R : 0 < w^F < \infty\}$. Let the continuous variate X_l with sample space $\mathcal{X}_l$ be the *predictor* — a deterministic forecast of the yield of Mato Grosso, in bags per hectare . Its realization is denoted $x_l \in \mathcal{X}_l$, where $\mathcal{X}_l = \{x_l : 0 < x_l < \infty\}$, for $l = 1, ..., 5$. Table 6.2 on page 81 contains the records of X_l, for $l = 1, ..., 5$, and W^R.

The steps to formulate the BPF are similar to the models developed so far up to the derivation of the posterior functions. The historical prior distribution function G is used in the NQT. Therefore, the steps similar to those performed in chapter 6 are omitted in order to present only the sensitivity of the models to the judgmental prior distribution functions. The reader must go through steps 1-4 for each model in section 6.5 and complete the modeling with the results described in this section.

Two types of judgmentally assessed prior distribution functions are evaluated. The first is a prior distribution function G_T^R modeled from the judgmentally assessed quantiles of W^R. This function is called *judgmental prior distribution function*, and it is described in chapter 4. The second is a prior distribution function G_T^S modeled from quantiles of W^R that are obtained from the judgmentally assessed quantiles of W^F using a field-region stochastic transformation. This function is called *transformed prior distribution function*, and it is described in chapter 5.

Table 7.4 Parametric models and parameter values of the historical prior distribution function G and the judgmental prior distribution functions G_T^R and G_T^S.

Function	Distribution	α	β	η_L	η_U
G	LC2–IW	1.0777	6.7246	30	90
G_T^R	LR1–LP	0.4154	-0.3489	30	90
G_T^S	LR1–LP	0.7369	-0.0329	30	90

Each judgmental prior distribution function was obtained by combining J individual prior distribution functions assessed from farmers in Mato Grosso. The parametric model for the judgmental and transformed prior distribution functions is LR1–LP

$$G_T^R(w^R) = G_T^S(w^R) = \begin{cases} \frac{1}{2}\exp\left(\frac{y-\beta}{\alpha}\right), & \text{if } y \leq \beta, \\ 1 - \frac{1}{2}\exp\left(-\frac{y-\beta}{\alpha}\right), & \text{if } \beta \leq y, \end{cases} \tag{7.9}$$

where

$$y = ln\frac{w-\eta_L}{\eta_U - w^R}. \tag{7.10}$$

Table 7.4 shows the parameter values for G, G_T^R, and G_T^S. Figure 7.9 shows the graphs of the respective distribution and density functions. Chapter 4 presents a more detailed discussion of the shapes of these prior distribution functions. This chapter analyzes the sensitivity of the posterior distribution and density functions to the different prior distributions.

Each BPF model is updated using the judgmental and transformed prior distribution functions. The BPF_{Oct} is updated using the methodology described in subsection 3.3.1. The BPF_{Feb} and BPF_{May} are updated according to the methodology in subsection 3.3.2. Therefore, two new sets of posterior density functions are obtained for each BPF.

Figure 7.10 shows the three prior density functions and examples of the resulting posterior density functions, conditional on different predictor values. The posterior density functions modeled using only the historical prior distribution function (column 0) are just

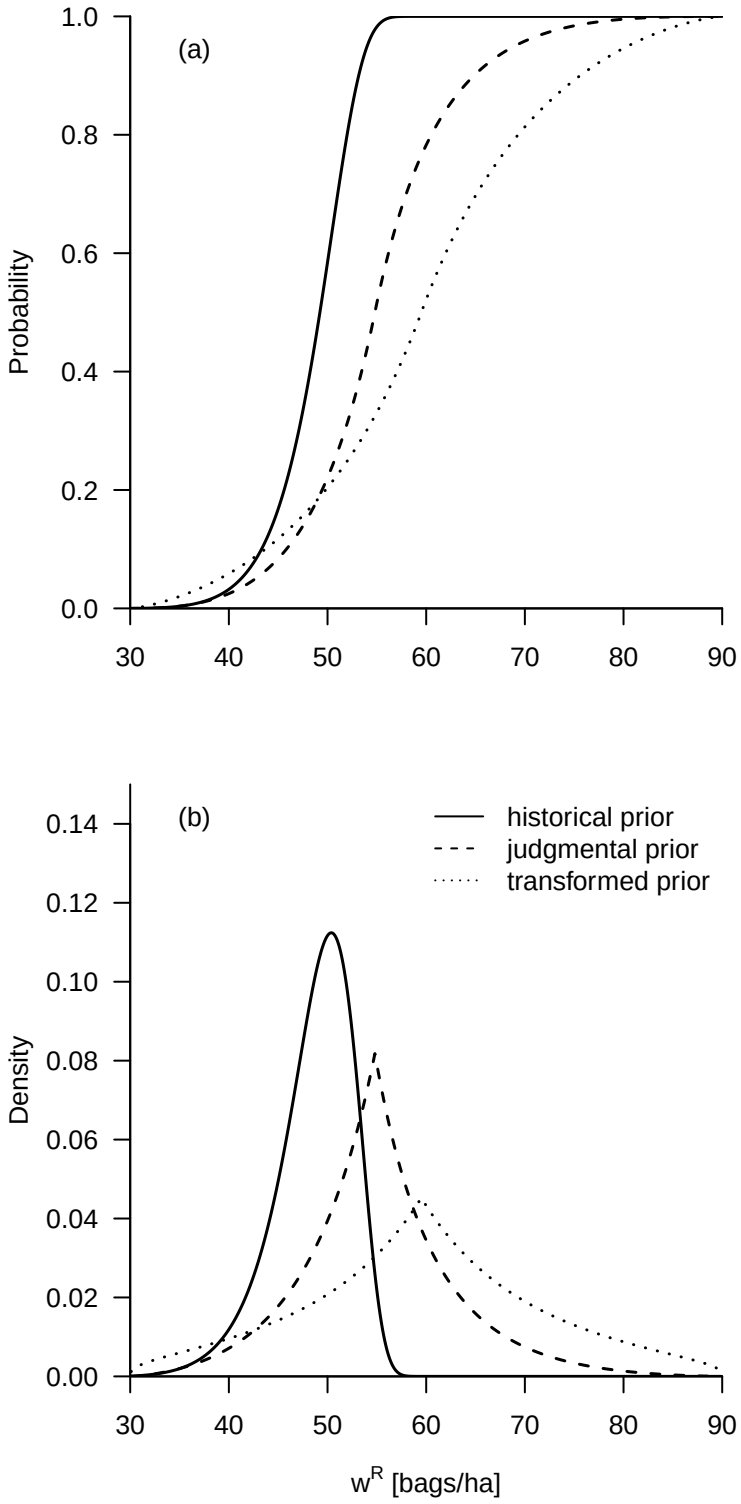


Figure 7.9 (a) Historical prior distribution function G, and judgmental prior distribution functions G_T^R and G_T^S; (b) corresponding density functions.

a reproduction of the results in chapter 6.

The impact of the judgmental and transformed prior distribution functions in the BPF_{Oct} is quite significant. These prior distribution functions suggest a greater variance than the historical prior, as well as higher medians. The posterior density functions are adjusted accordingly. The updated posteriors cover a wider range of possible values of yields, and they seem to have increase the uncertainty about W^R relative to the models developed in chapter 6.

Although the same judgmental and transformed prior distribution functions are used in the BPF_{Feb}, the impact seems to be weaker than in the BPF_{Oct}. There is a decrease in posterior uncertainty, which is due to a more informative set of predictors. The same comparison can be made between the BPF_{May} and the BPF_{Feb}.

These results illustrate the use of judgmentally assessed prior distributions in the Bayesian forecaster. The task to choose between one model or another will depend on a careful analysis of the performance of each farmer as a forecaster. Ultimately, the conventional BPF, using only the historical prior, can be used for a primary analysis, which next can be adjusted according to the judgmental prior distribution functions.

The transformed prior distribution function seemed to have increased the uncertainty about W^R when compared to the judgmental prior distribution function. The field-region stochastic transformation still holds as a viable method, but improvements can be made by collecting actual realizations w^F from each farmer, instead of using the yields of the microregion as proxies. Another potential improvement in this model lies in the elicitation of quantiles from farmers. The methodology used in this research tries to separate the farmer's assessment process from complex probability concepts.

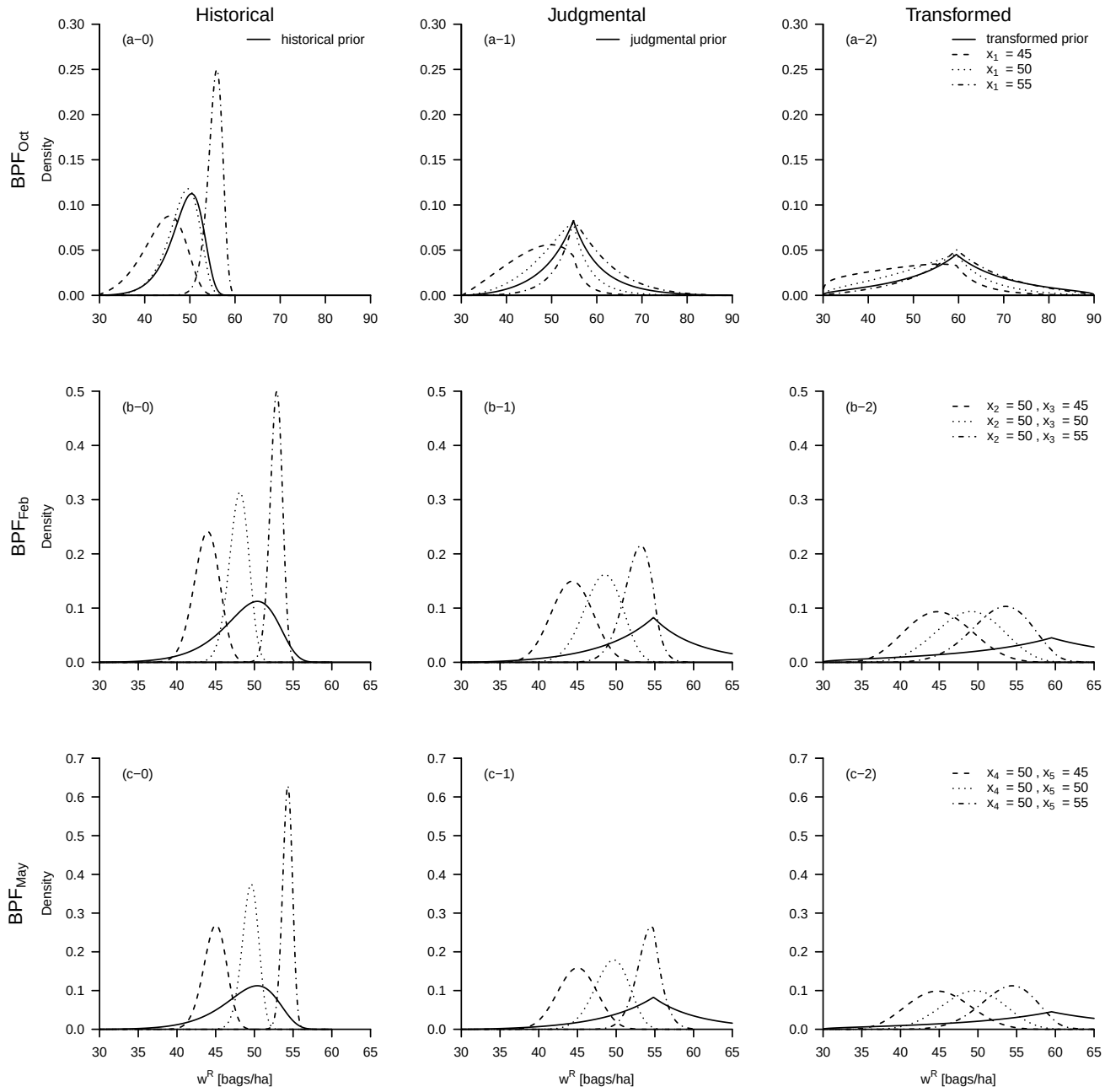


Figure 7.10 Prior density functions and examples of posterior density functions arranged in rows by type of BPF: (a) BPF_{Oct}, (b) BPF_{Feb}, and (c) BPF_{May}; and columns by type of prior distribution function utilized to derive the posterior functions: (0) historical prior, (1) judgmental prior, and (2) transformed prior.

7.4 Summary

This chapter analyzed the sensitivity of BPF models to the judgmental prior distribution function and to the transformed prior distribution function previously modeled in chapters 4 and 5. The first model developed provides a guideline to forecast the yield of a specific field. Farmers can used their own records to construct a BPF model and update it according to their judgments in a systematical way. This model can be constructed one time, and it can be used for many years until the additional sample size is large enough for updates. This methodology can be applied directly from this research or it can be coded into a software in order to facilitate use.

The updated BPF models incorporated the judgmentally assessed prior distribution functions. Although the shapes of these prior functions look invalid, it was possible to analyze their impact on the posterior density functions. In fact, these results illustrate the case of a judgmental prior distribution that seems incompatible with the shape of the historical prior distribution. This could happen due to underconfidence of the subjects, inaccurate understanding of the domain of the variate of interest, or flawed methodology to assess subjective quantiles.

8. Summary and Conclusions

This research has accomplished its objective of producing probabilistic forecasts of agricultural yields and incorporating judgmental information into the model. This chapter summarizes the contributions of this research, and suggests future research.

8.1 Summary of Contributions

The motivation for this research was to provide an alternative to the forecasting methods currently being applied in agriculture (i.e., to forecast soybean crop yield of a particular field or region). The available forecasts published by the Brazilian government are deterministic and they usually present no assessment of uncertainty. Developing a Bayesian forecaster for this problem allowed the production of probabilistic forecasts of the yield of a field or a region.

In addition to this goal, an expansion of the Bayesian Processor of Forecasts (BPF) was presented in order to incorporate information from farmers. Information coming from farmers, traders, weather forecasters, and other agents in the supply chain is widely assimilated in agriculture. In fact, many statistics published by official organizations such as IBGE and CONAB have a subjective component associated with individual or group forecasts. This research presented a methodology to collect judgmental assessments from farmers in the form of quantiles, and use them to model prior distribution functions. Specifically, the main contributions are related to four aspects of forecasting.

1. *Probability assessments.* Assessing subjective probabilities or other statistics from experts who do not have training in probability theory is challenging, and it has been a topic of interest for many decades. This research reviewed and applied a methodology to assess quantiles from experts. The online tool Qualtrics provided an interactive tool for farmers to assess quantiles of the yield.

2. *Forecast combination puzzle.* This term is commonly used to address the problem of combining forecasts of a variate of interest produced from a set of models. This research adds to this approach by exploring the combination of prior distribution functions using the Bayesian Model Averaging (BMA) theory that finds an optimum set of weights for each year. Appendix A presents three examples constructed to evaluate this algorithm in terms of calibration and informativeness. The BMA algorithm was able to reward more informative forecasters and punish uninformative ones, but it was not capable of discerning well calibrated forecasters from miscalibrated ones.

3. *Information fusion.* The credibility of information sources in agriculture is not always satisfactory. In fact, it is common in the market to observe a variety of deterministic forecasts for the same variate (e.g., the soybean crop yield for a region). The BPF developed in this research is able to quantify the uncertainties of multiple sources, such as CONAB and IBGE, and to fuse them. Decision maker can take advantage of both sources in a systematic manner without having to subjectively opt for a specific one.

4. *Probabilistic forecasts.* In summary, my research allows us to produce forecasts of the yield of a field or a region and have an immediate assessment of uncertainty about these variates, when the current methods often omits it. This supplemental information is useful to improve risk management and decision making in many parts of the food supply chain. The cases analyzed provide a guideline on how to apply the methodology and obtain probabilistic forecasts from deterministic forecasts and judgmental assessments. Instead of having a point estimate of the yield, decision maker can have an entire distribution function of the yield.

These contributions have the potential to improve decision making in agriculture by allowing the agents of the supply chain to explicitly take the risk into account in their

decisions during the crop season. In addition, contributions to the probability assessments and the forecast combination puzzle could be transferred to other fields, such as data analysis of large samples, machine learning, and risk analysis.

8.2 Future Research

Some of the topics presented here also represent potential areas for improvement and future research. Ample opportunities exist to continue the development of the Bayesian forecaster. Some of the suggestions for future research are:

1. *Quantile assessments.* Future research on this topic may further explore the use of data visualization tools to assess quantiles from experts, such as in Qualtrics. Many studies so far have tried to explain the concept of quantile or probability to the subjects prior to the interview, such as the study by Alpert and Raiffa (1982), but there is still space to explore research methods that circumvent this step. Moreover, the validation of the assessed quantiles is a step that could be further explored.

2. *Bayesian Model Averaging.* Further research is necessary to improve the use of the BMA approach in combining judgmental prior distribution functions, specifically in regards to the inability of the BMA to identify miscalibrated forecasters.

3. *Probabilistic forecasting of agricultural variates.* In this research, the variate of interest was the yield, but future research could analyze other types of agricultural variates, such as production, price, and so on. The same approach could be applied to other crops using different predictors, such as weather, disease incidence, or input applications, like fertilizer and chemicals.

Appendix A. BMA Simulation

This appendix develops 3 simulated examples to illustrate the application of the Bayesian Model Averaging (BMA) approach in obtaining the weights λ_j, for j = 1, ..., J, discussed in subsection 4.5.3. Additionally, these examples explore the response of the BMA approach to two characteristics of the forecasters, informativeness and calibration.

A.1 Example A

Example A analyzes the progression of simulated judgmentally assessed distribution functions of $J = 3$ farmers during $T = 8$ years. For each year, a set of weights λ_j is obtained using the BMA approach, for $j = 1, 2, 3$. Since the performance of each farmer in the first year $t = 0$ is unknown, a uniform set of weights is applied in that year. The first task is to compose 8 actual values w^R. Table A.1 shows the actual values w^R. These values are utilized to analyze the calibration of the forecasters.

The objective of this example is to verify whether the BMA can assign weights according to different levels of *informativeness* among the subjects. In order to discuss this problem, the simulation must satisfy certain conditions. The first condition refers to concept of *calibration of forecasters*.

According to Krzysztofowicz (2016), a forecaster is considered *well calibrated* if the frequency of events observed follows the frequency specified by the assessed quantiles. Therefore, the distribution functions generated for each farmer were constructed such that the frequency of events across $T = 8$ years followed the forecast probabilities simulated for each farmer. In other words, farmers j = 1, 2, 3 are well calibrated.

Table A.1 Simulated values w^R.

t	0	1	2	3	4	5	6	7
w^R	35	64	57	68	65	45	59	72

Adapting the definition explained in Krzysztofowicz (2016), a forecaster $j = 1$ is said to be more informative than a forecaster $j = 2$, with respect to predictand W^R if, for every rational decider, the economic value of forecaster $j = 1$ is at least as high as the economic value of the forecaster $j = 2$. In order to compare the forecasters, Krzysztofowicz (2016) defines the concept of *sufficiency characteristic* as

$$SC = \frac{|a|}{\sigma}, \tag{A.1}$$

where $|a|$ is called the "signal" and it is the slope of the regression of the forecast median $w_{j,0.5}(t)$ on the actual realization $w(t)$; and σ, called the "noise", is the standard deviation of the residuals around the regression line. Density functions with different shapes were constructed in order to simulate different levels of informativeness. These functions represented the uncertainty of each forecaster with respect to the predictand. The following profiles were then created using the concepts mentioned:

- *$j = 1$.* A forecaster well calibrated and highly informative. The density function for each year is narrow and tall.
- *$j = 2$.* A forecaster well calibrated and informative. The density function for each year is less narrow and tall than that of the highly informative forecaster, but it still reduces uncertainty about the predictand.
- *$j = 3$.* A forecaster well calibrated and uninformative. The density function of this forecaster is roughly the same over the entire time frame. Therefore, the forecaster adds no information from one year to another.

The comparison of these 3 profiles allows us to analyze whether the BMA algorithm assigns different weights according to the different levels of informativeness. The distribution functions created in this simulation follow the LC2–IW distribution as described in equations (6.1) and (6.2). Figures A.1 to A.6 show the simulated distribution and density functions for each profile for $t = 0, ..., 7$.

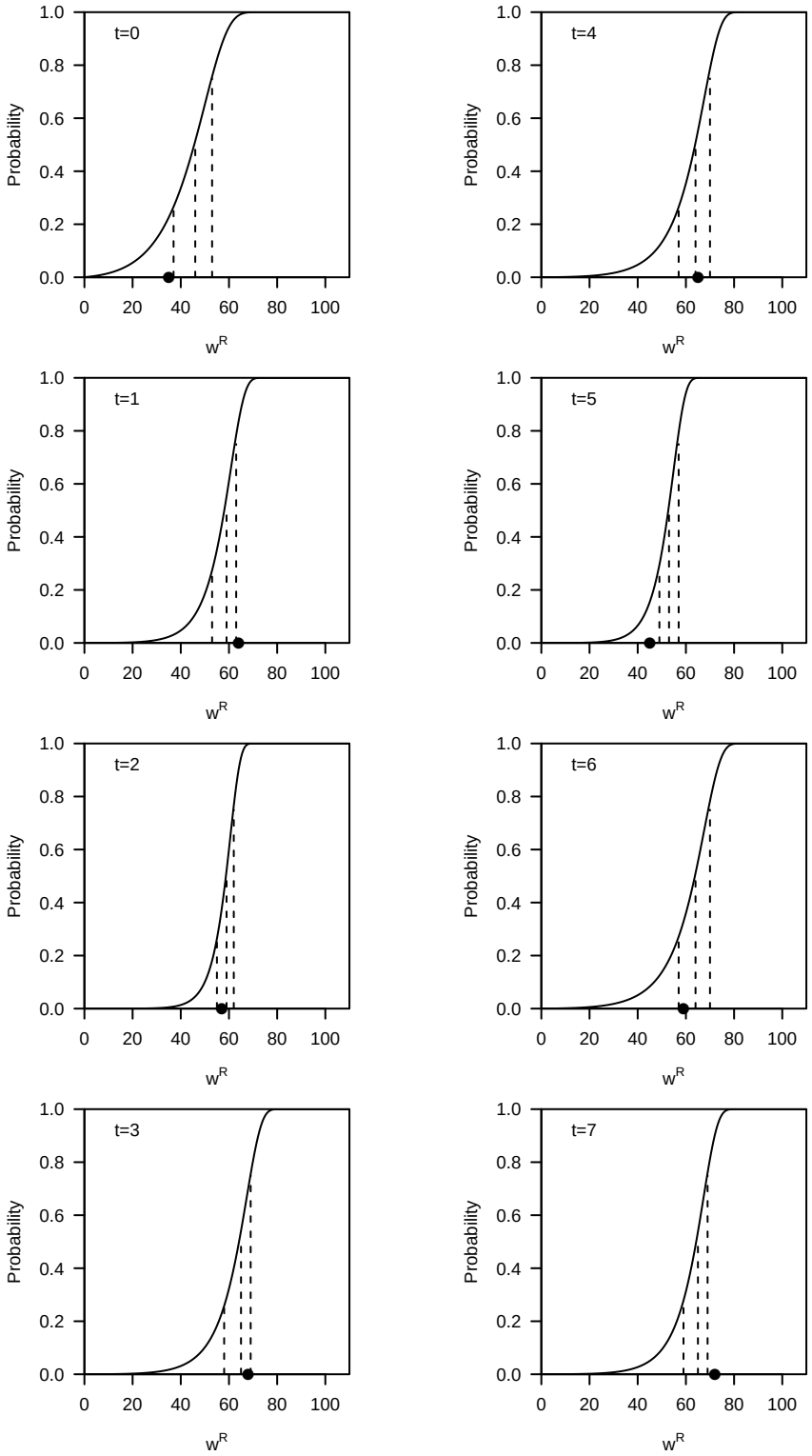


Figure A.1 Distribution functions simulated for $j = 1$, for $t = 0, ..., 7$.

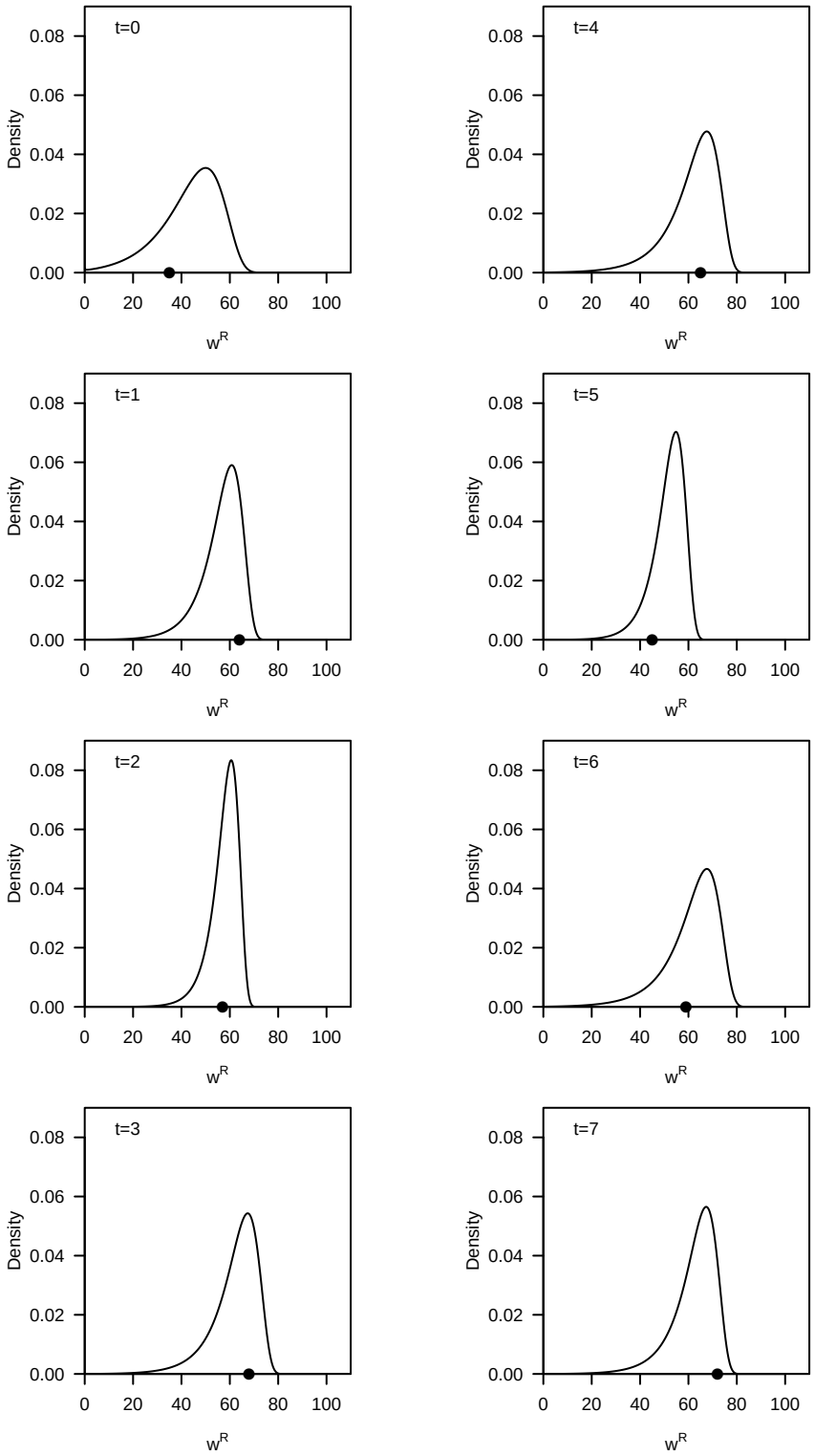


Figure A.2 Density functions simulated for $j = 1$, for $t = 0, ..., 7$.

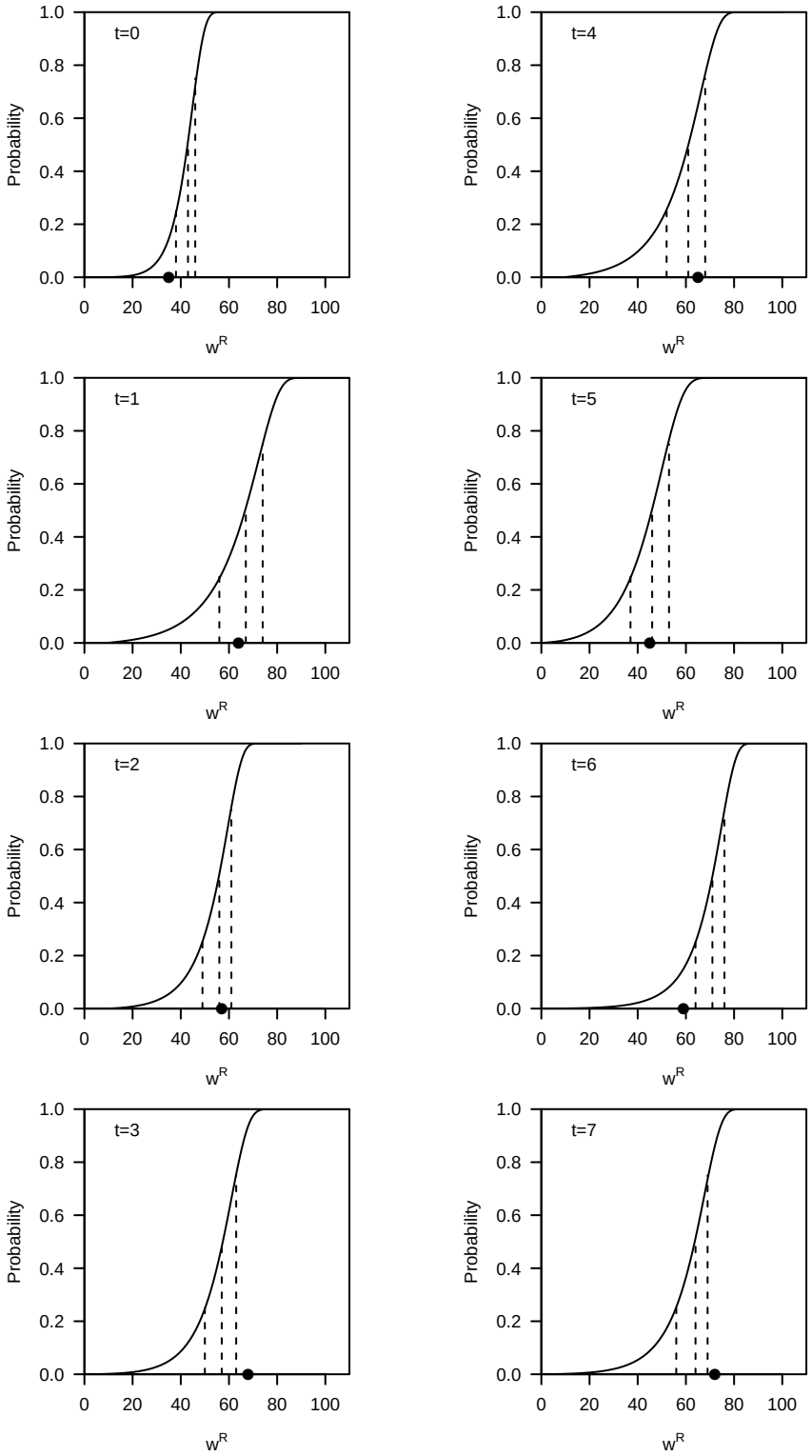


Figure A.3 Distribution functions simulated for $j = 2$, for $t = 0, ..., 7$.

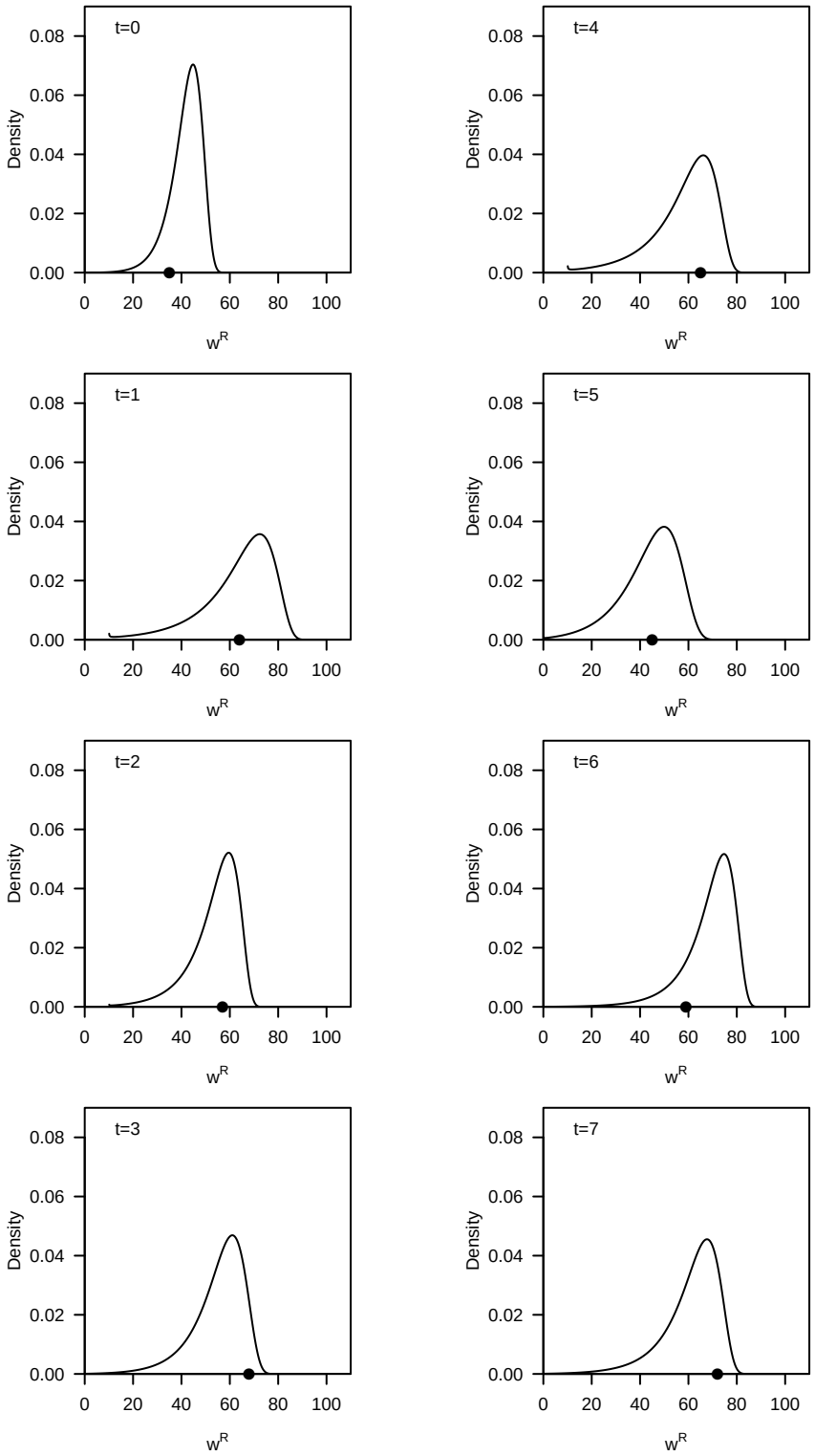


Figure A.4 Density functions simulated for $j = 2$, for $t = 0, ..., 7$.

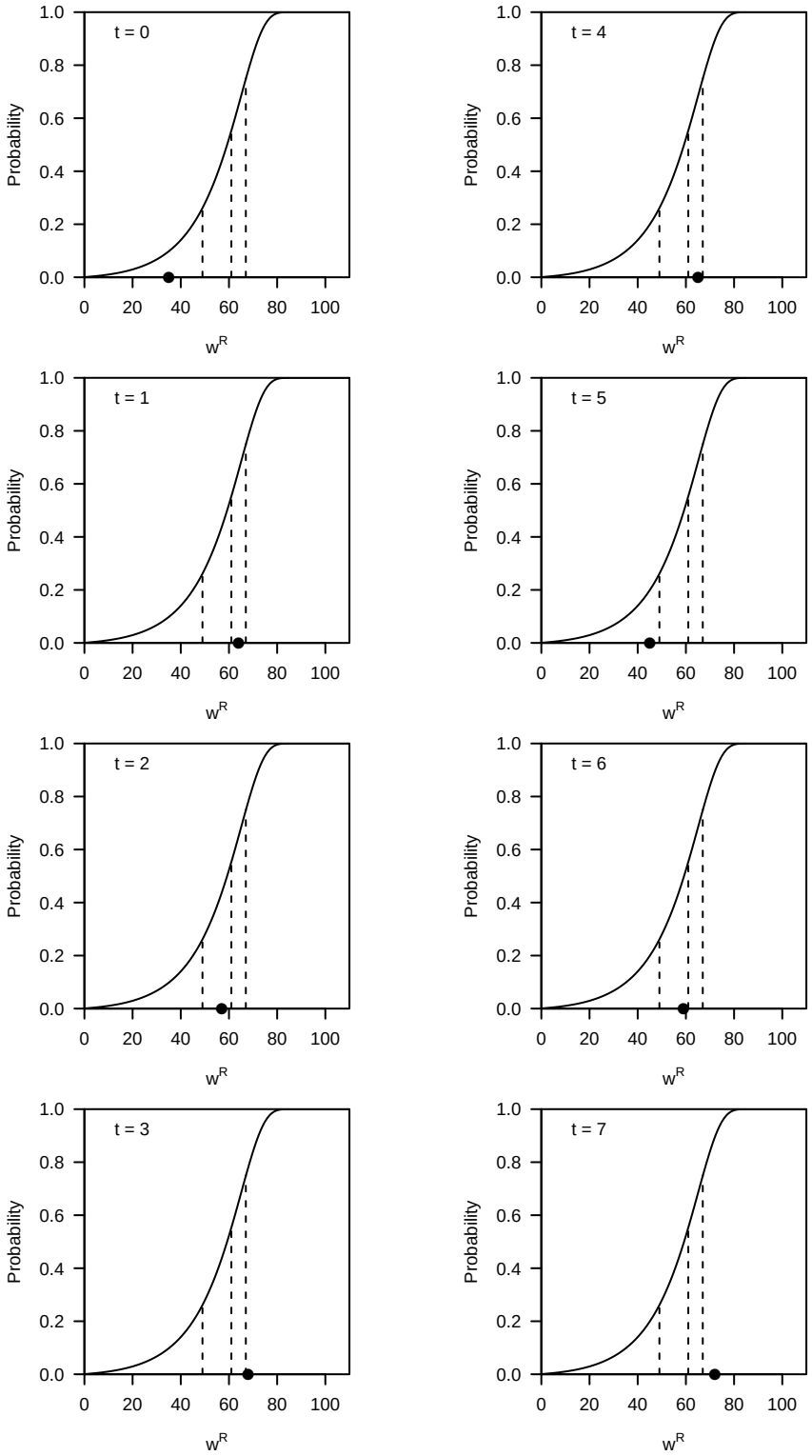


Figure A.5 Distribution functions simulated for $j = 3$, for $t = 0, ..., 7$.

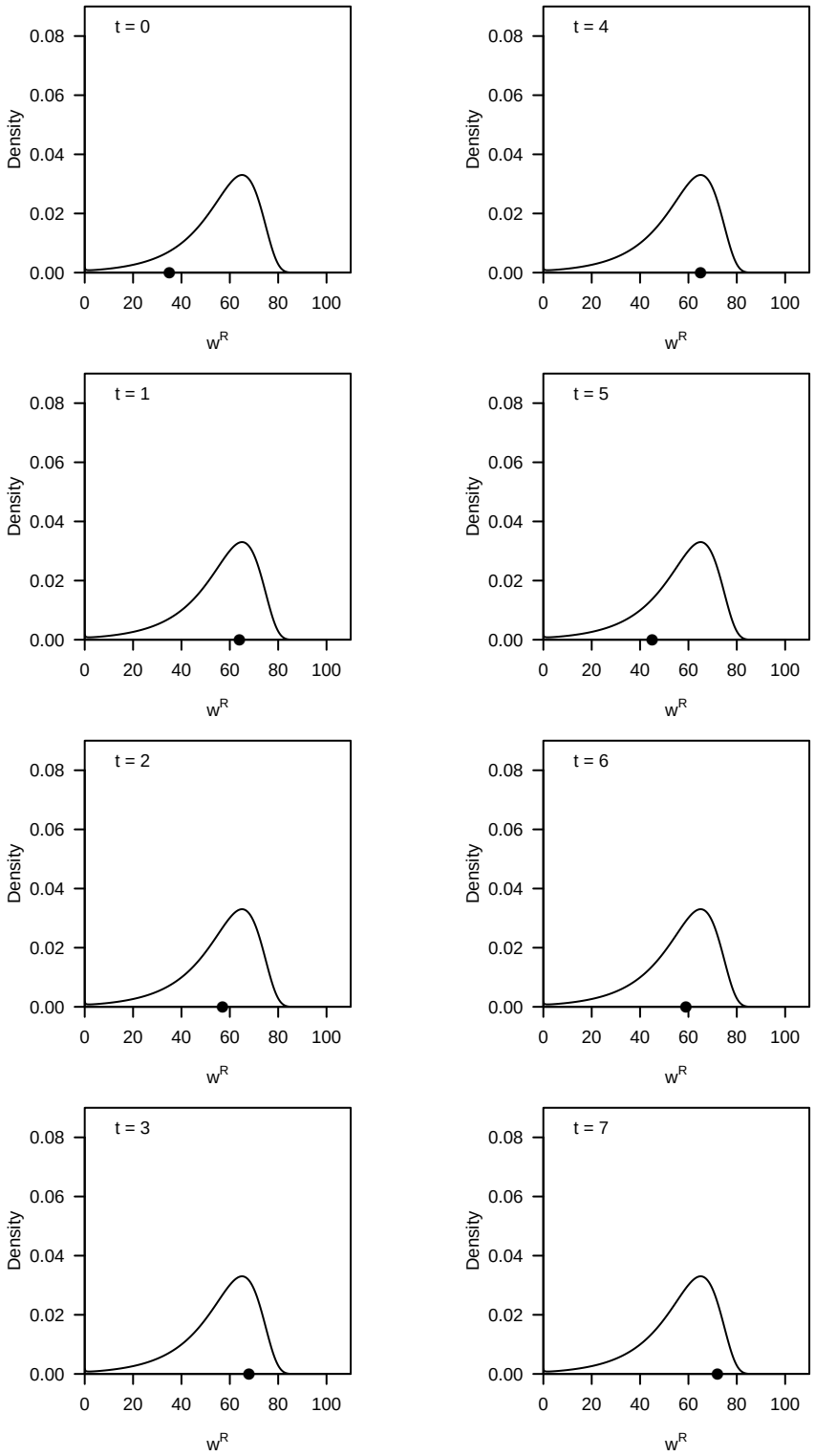


Figure A.6 Density functions simulated for $j = 3$, for $t = 0, ..., 7$.

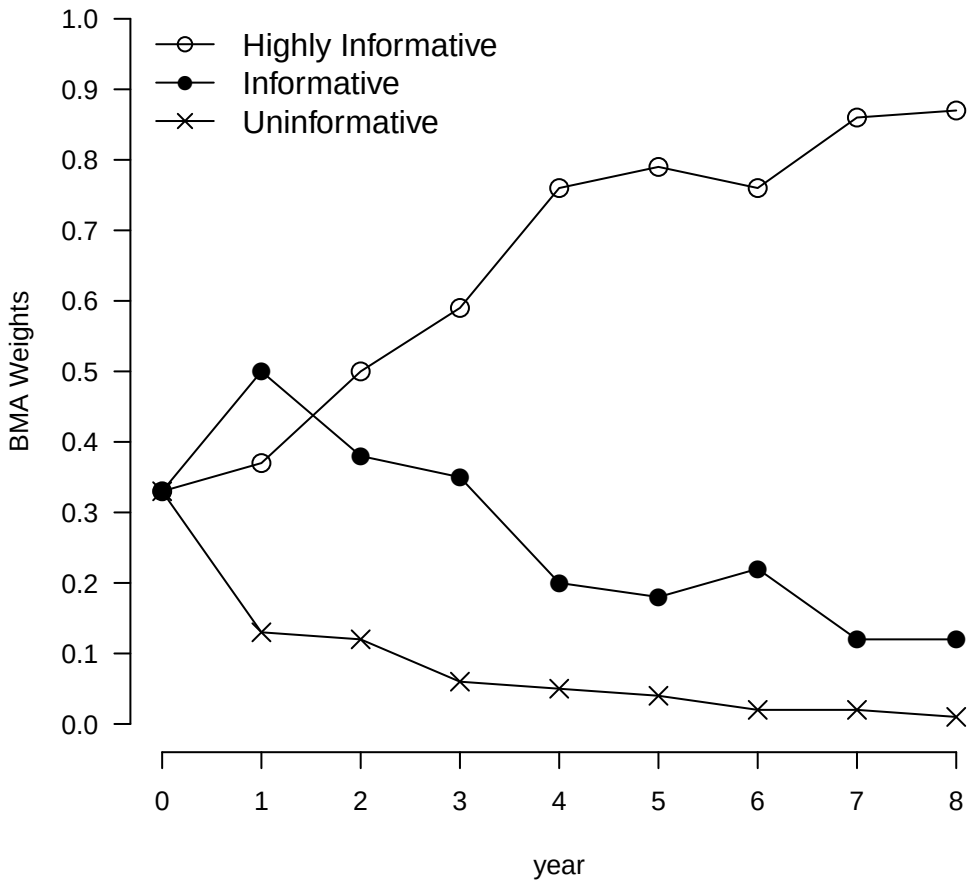


Figure A.7 BMA weights in Example A for $t = 0, ..., 7$.

Figure A.7 shows the weights obtained from the BMA algorithm. As mentioned, the forecasters start with a uniform weight since there is no information about their calibration and informativeness. However, once the subjects start providing their assessments, the BMA algorithm changes the weight of each person based on their performance.

After analyzing the density functions provided in year 0, the BMA algorithm assigns the highest weight to forecaster 2, and lower weights to forecasters 1 and 3. This result reflects the shape of the density functions provided that year. The density function of forecaster 2 was narrower around the median and the actual value was very close to it,

while the other forecasters presented density functions more flat.

However, forecaster 1 quickly became more informative than the other ones. The BMA consistently reallocated the weights from forecasters 2 and 3 to forecaster 1 over time. By year 8, the uninformative forecaster 3 has a weight very close to 0, while the weight for forecaster 1 is considerably higher than the weight for forecaster 2. Therefore, the BMA algorithm is able to penalize uninformative forecasters and favor the highly informative one, given that they are all well calibrated.

In addition, the algorithm is able to recognize improvements in the forecasts of certain subjects in certain years and adjust gradually over the future years. For instance, forecaster 2 presented years of improvements, such as in $t = 5$ and the algorithm increased his weight in the following year. This result is particularly interesting to the topic of training forecasters to improve their assessments. Suppose forecaster 3 decided to invest time in training and improving his understanding of the problem. The BMA algorithm would allow him to have higher weights, given observed improvements in his forecasts, while considering his historical performance.

Table A.2 shows the SC_j for each forecaster after the initial 4 years and at the end of the 8 years. Since forecaster 3 provides the same median for all years, the SC_3 is equal to 0. The difference between SC_1 and SC_2 is quite significant in the initial 4 years. Therefore, it is possible to conclude that forecaster 1 is more informative than forecaster 2. Although the difference between these two measurements decrease, the same conclusion holds after 8 years. However, the BMA algorithm did not seem to be affected by this change, as it continued to assign higher weights to the highly informative forecaster.

Table A.2 Sufficiency characteristics for $j = 1, 2, 3$.

years	SC_1	SC_2	SC_3
0 to 3	0.628	0.196	0
0 to 7	0.451	0.137	0

A.2 Example B

Similarly to previous case, Example B analyzes simulated distribution functions for $J = 3$ forecasters during $T = 8$ years. For each year, a set of weights λ_j is obtained using the BMA approach. The actual values w^R are the same as in table A.1.

The objective of this example is to analyze whether the BMA algorithm can identify a well calibrated forecaster and assign different weights accordingly. In order to analyze this problem, two constraints were imposed on the simulated data:

1. Each forecasters has a well-calibrated median. That is, after 4 years and after 8 years, 50% of the actual realizations w^R fell below the median and 50% above.

2. Each forecaster is about equally informative. Therefore, for this example, the constructed distribution functions must yield $SC_1 \approx SC_2 \approx SC_3$. The examples were constructed such that the sufficiency characteristic of each forecaster is approximately equal from years 0 to 3, 4 to 7, and for the entire range 0 to 7.

The following profiles were then created subject to the constraints mentioned above:

- *j = 1.* Well calibrated forecaster. The frequency of actual realizations w^R equals the probability specified by the simulated quantiles.

- *j = 2.* Overconfident forecaster. In this case, the actual realizations fell either below $w_{j,0.25}(t)$ or above $w_{j,0.75}(t)$.

- *j = 3.* Underconfident forecaster. In this case, the actual realizations fell between $w_{j,0.25}(t)$ and $w_{j,0.75}(t)$.

Figures A.8, A.9, and A.10 show the simulated distribution functions for $j = 1, 2, 3$. These distribution functions were constructed so that the forecasters are about equally informative after 4 years as well as after 8 years. The distribution functions simulated in this example follow the LC2–IW distribution as described in equations (6.1) and (6.2). Figure A.11 shows the set of weights λ_j for $j = 1, 2, 3$ over the years.

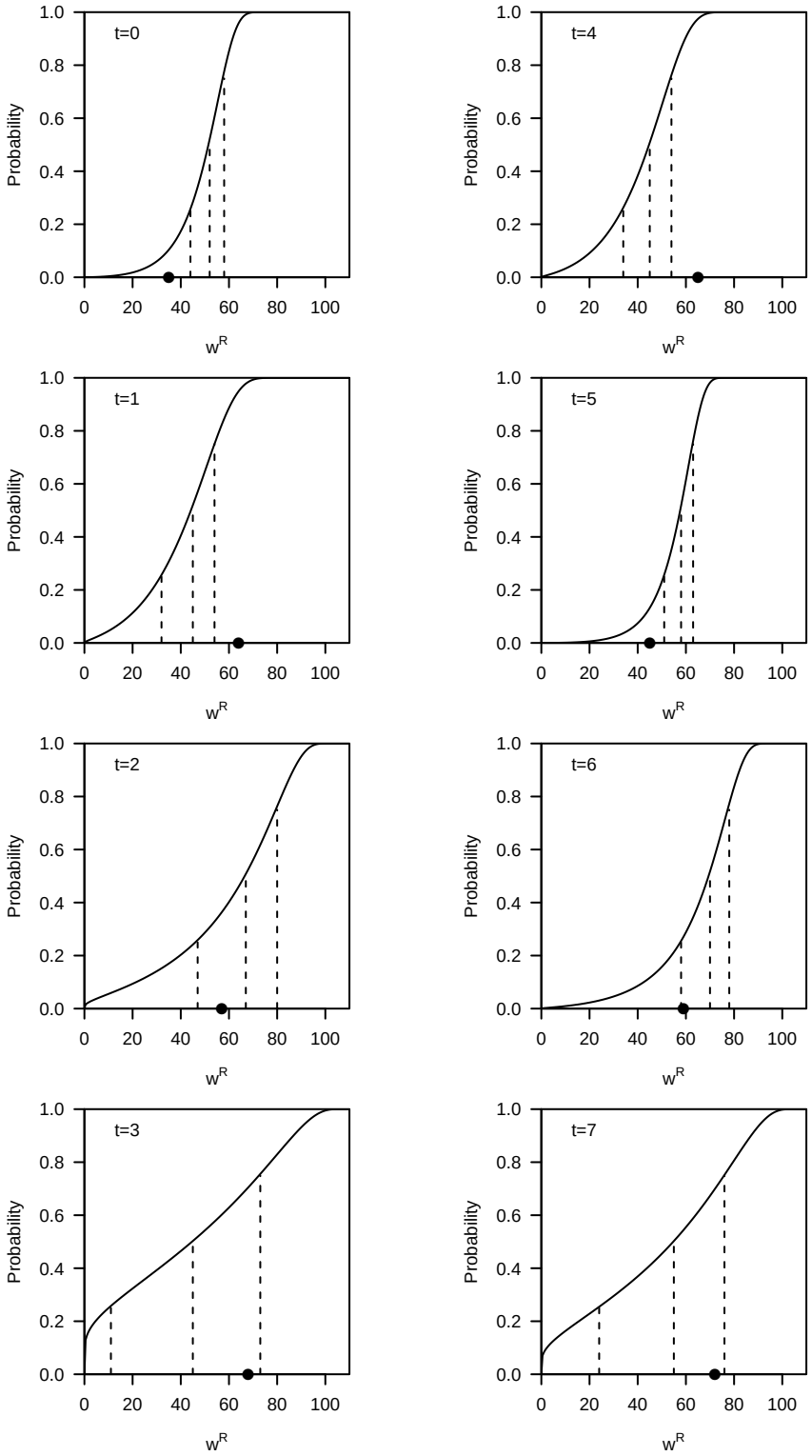


Figure A.8 Distribution functions simulated for $j = 1$, for $t = 0, ..., 7$.

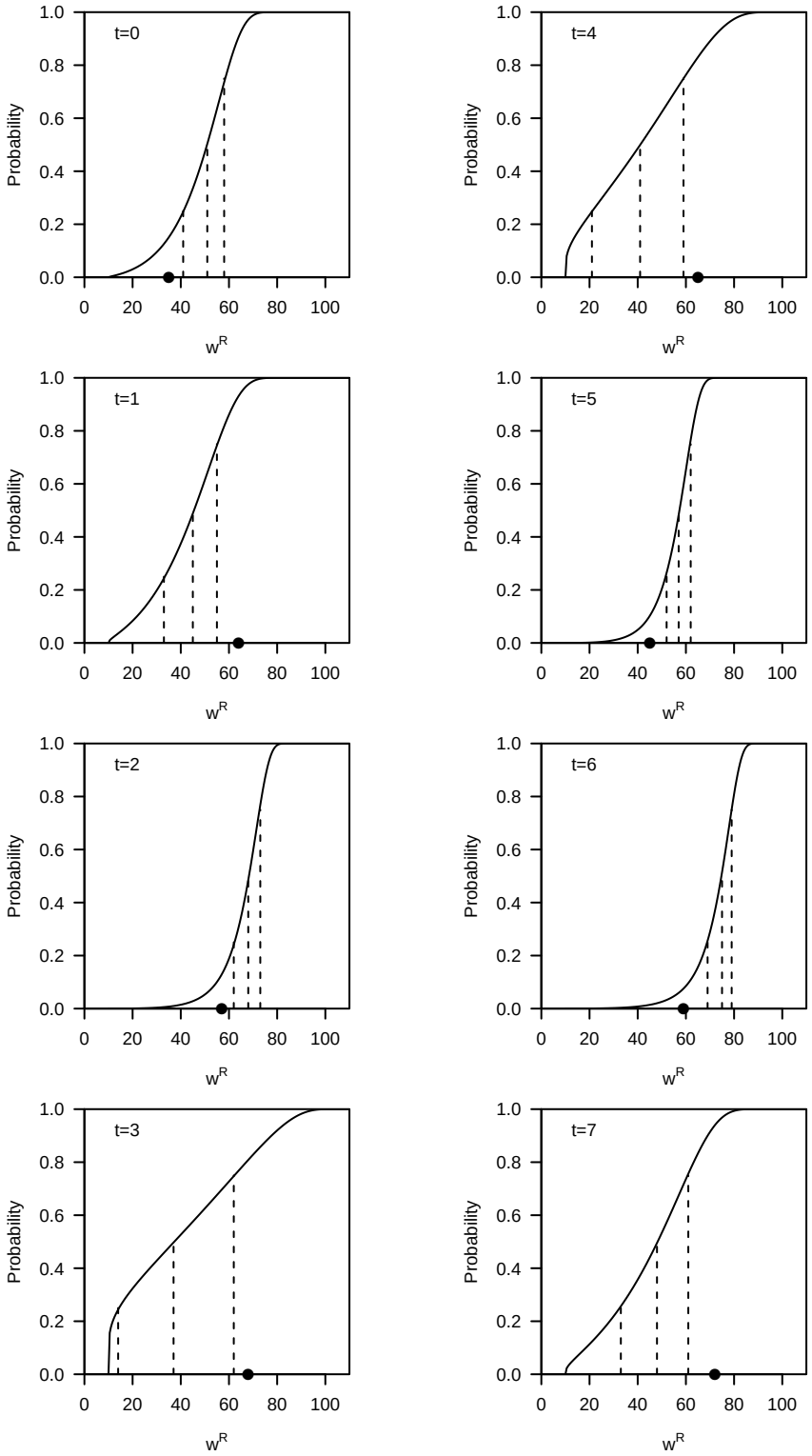


Figure A.9 Distribution functions simulated for $j = 2$, for $t = 0, ..., 7$.

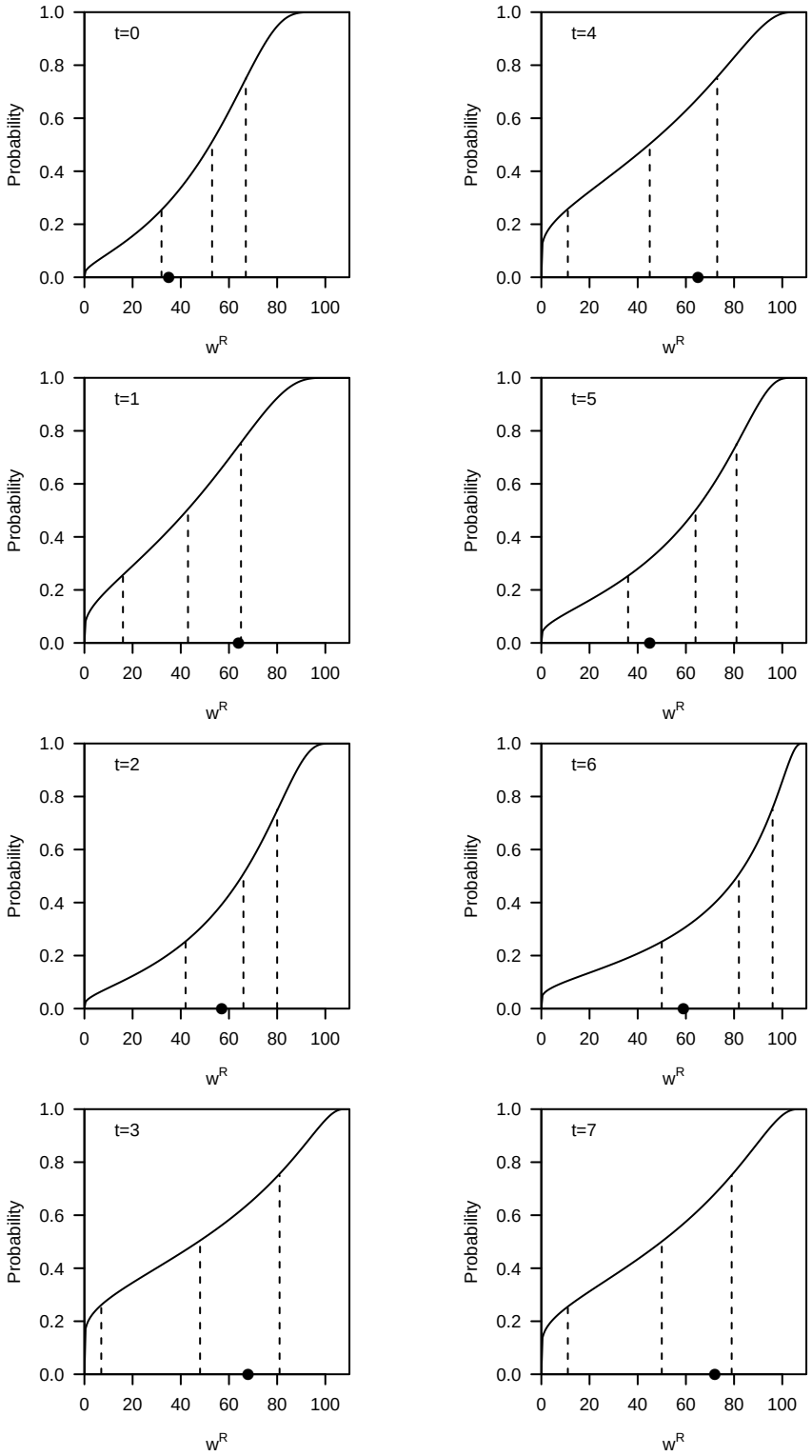


Figure A.10 Distribution functions simulated for $j = 3$, for $t = 0, ..., 7$.

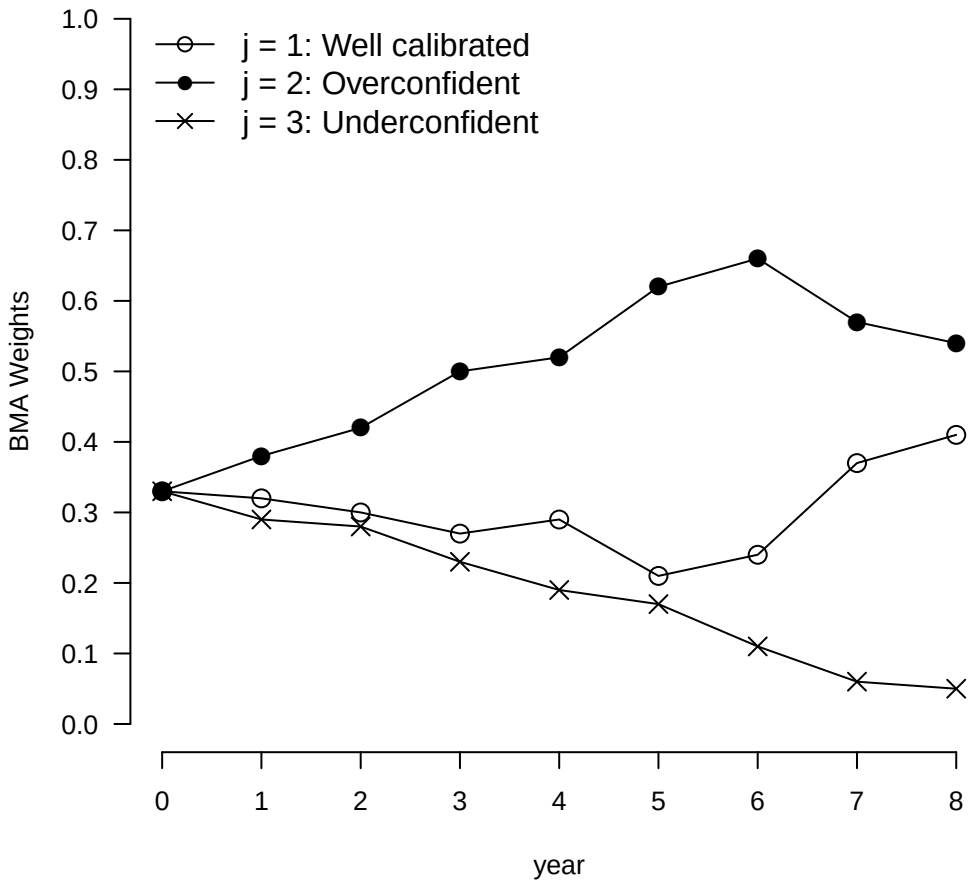


Figure A.11 BMA weights in Example B for $t = 0, ..., 7$.

The BMA algorithm consistently assigned higher weights to the overconfident forecaster, when compared to the other forecasters, during all 8 years. The well calibrated and underconfident forecasters had similar weights until year 5, then the BMA algorithm assigned higher weights to the well calibrated forecaster. Intuitively, a well calibrated forecaster is preferable to an overconfident or underconfident forecaster. However, the BMA algorithm was not able to capture this problem. Table A.3 shows the sufficiency characteristics for $j = 1, 2, 3$.

Table A.3 Sufficiency characteristics for $j = 1, 2, 3$.

years	SC_1	SC_2	SC_3
0 to 3	0.032	0.032	0.036
4 to 7	0.037	0.035	0.039
0 to 7	0.025	0.027	0.025

This example shows the importance of training forecasters towards improving their calibration, as the BMA algorithm is not capable of identifying this the miscalibration automatically. Therefore, the calibration of forecasts must be ensured independently from, and before the application of, the BMA the algorithm.

A.3 Example C

Example C builds on the previous example to analyze the sensitivity of the BMA algorithm to a single very unrealistic distribution function. Consider the same forecasters from example B: well calibrated, overconfident, and underconfident. Suppose the same distribution functions were assessed by each forecaster during the same eight years, except for the distribution function for year $t = 3$.

Suppose three different scenarios, where in each scenario a different forecaster assessed the unrealistic distribution function for $t = 3$ while the others remain the same as in example B. Figure A.12 shows the unrealistic distribution function. The actual value was $w^R = 68$ bags per hectare for $t = 3$. This distribution function gives probability 1 for $W^R < 68$. Therefore, it was far off from reality. The three scenarios constructed are:

- *Scenario 1.* Well calibrated forecaster provides unrealistic distribution function at $t = 3$, while other remain the same as in example B.
- *Scenario 2.* Overconfident forecaster provides unrealistic distribution function at $t = 3$, while other remain the same as in example B.
- *Scenario 3.* Underconfident forecaster provides unrealistic distribution function at $t = 3$, while other remain the same as in example B.

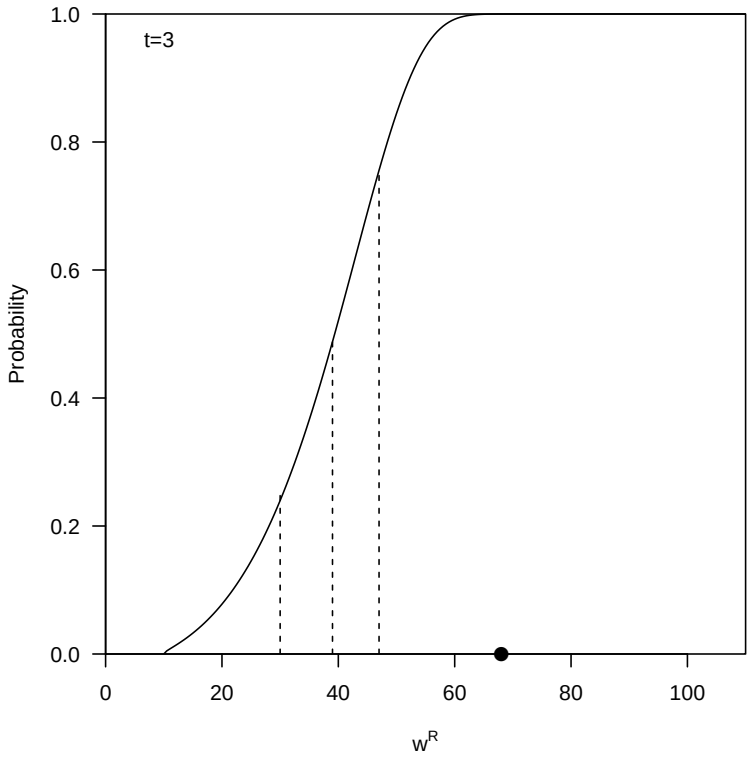


Figure A.12 Unrealistic distribution function for $t = 3$.

Figure A.13 shows the BMA weights for each scenario. All scenarios presented the same feature regarding the sensitivity of the BMA algorithm to the unrealistic distribution function. The weight of the forecaster with the unrealistic distribution immediately dropped to zero and the forecaster did not regain any participation up to the end of the time interval.

Scenario 2 can be compared to the results in Example B. In example B, the overconfident forecaster had consistently higher weight than the others, but a mistake in one year could have eliminated him from consideration as exemplified in scenario 2. The same comparison applies to the other forecasters.

In each scenario, the final set of weights considerably changed after $t = 3$. Further research should examine whether this merciless feature of the BMA should be automatically allowed to exclude the misguided forecaster, even though he could be well calibrated and informative in the remaining period.

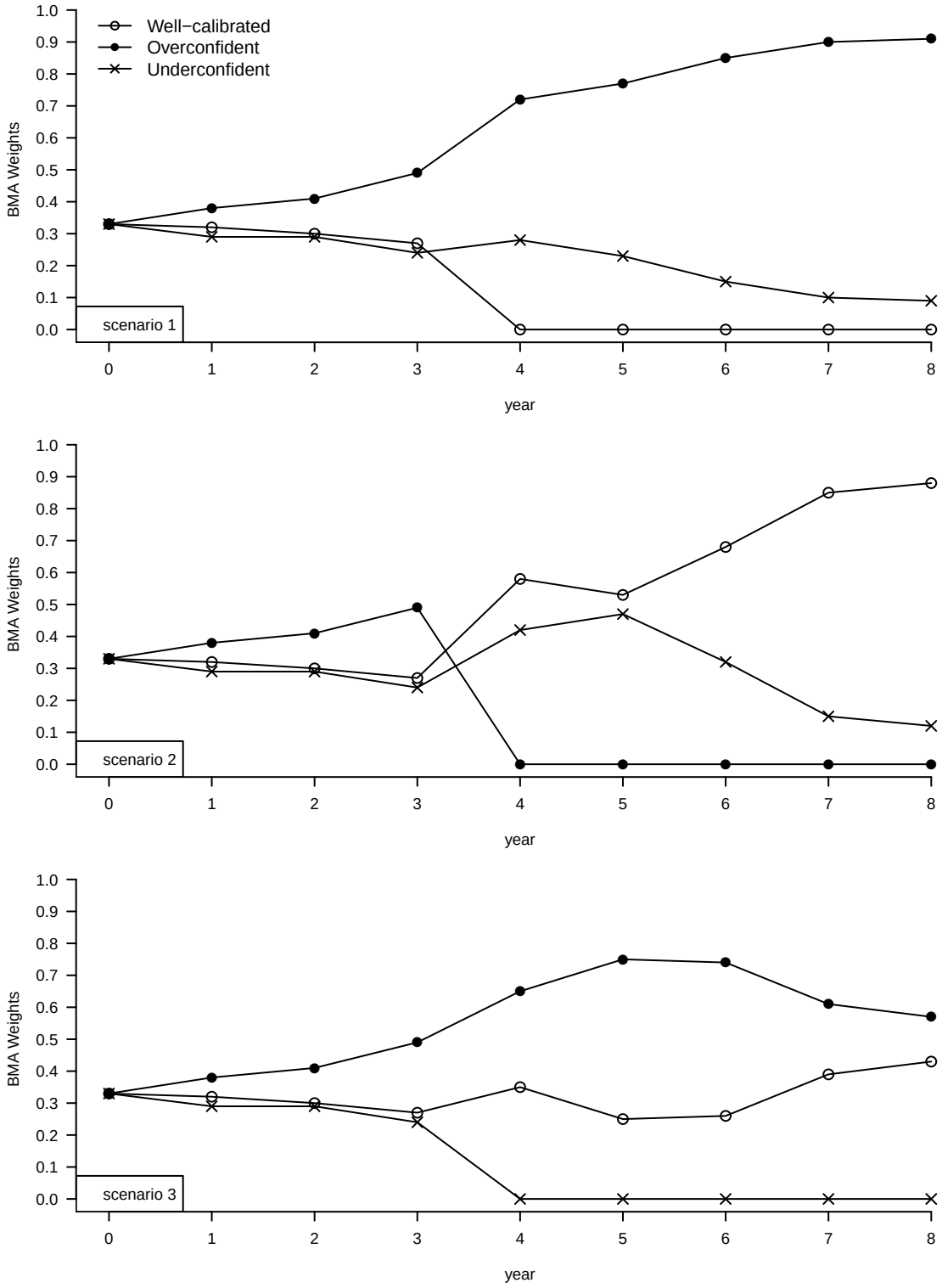


Figure A.13 BMA weights in Example B for $t = 0, ..., 7$ and scenarios 1, 2, 3.

A.4 Summary

The examples developed in this appendix illustrate the BMA approach to obtaining a set of weights for combining the judgmental distribution functions. It is desirable to assign higher weights to forecasters that present a better performance and lower weights to others. However, due to the nature of the problem in this research, it is difficult to obtain large data sets to evaluate the BMA in a short time frame. Therefore, simulating data is a solution.

Example A analyzed whether the BMA algorithm can capture different levels of informativeness and assign higher weights to more informative forecasters, when all are well calibrated. The results show a clear allocation of a higher weight to the more informative forecaster over the years.

Example B analyzed whether the BMA algorithm is able to identify well calibrated forecasters and assign different weights accordingly. The BMA algorithm was not capable of differentiating between the well calibrated forecaster and the others. This represents a disadvantage of using the BMA approach in an unsupervised way.

Example C analyzed the sensitivity of the BMA algorithm to an unrealistic distribution function. The algorithm quickly excluded the misguided forecaster, but became resistant to his recovery over time. This leads to another question of whether the BMA algorithm should be allowed to eliminate a forecaster who is unlucky one time, but otherwise is well calibrated and informative.

Appendix B. Modeling Distribution Functions

Modeling distribution functions from a data set is a key task in the application of the Bayesian Forecaster. This section describes a methodology that is used to model distribution functions throughout this research.

The methodology reviewed below is thoroughly described by Krzysztofowicz (2014). The distributions modeled are univariate, parametric, and have closed-form expressions for the distribution, density, and quantile functions. These expressions are convenient for the construction of the meta-Gaussian models developed in chapter 6.

Let us take as an example the modeling of the distribution function G of the continuous variate W, with sample space $\mathcal{W}$, and realization $w \in \mathcal{W}$. The goal is to find an expression for G that best represents the uncertainty related to W, given a sample of observed data. The same approach can be applied to modeling other distribution functions. The steps of this approach are as follows.

1. Construct an empirical distribution function of W.

   The empirical distribution function is created by pairing the observed data, sorted in ascending order, with a corresponding theoretical *plotting position*. This research uses the meta-Gaussian plotting positions calculated as follows (Krzysztofowicz, 2014):

$$p_n = \left[\left(\frac{N - n + 1}{n} \right)^{t_N} + 1 \right]^{-1}, \tag{B.1}$$

   where n is the rank of the ordered data, N is the sample size, and t_N is a constant calculated as follows for a sample size $11 \leq N \leq 20000$:

$$t_N = 1.9574 N^{-0.8039} + 1. \tag{B.2}$$

The ordered set of values $\{(w_{(n)}, p_n) : n = 1, ..., N\}$ specifies the empirical distribution function.

2. Hypothesize parametric models for G and estimate the parameters of each model.

   The parametric models considered in this research are available in a catalogue developed by Krzysztofowicz (2014). This catalogue provides the closed-form expressions for the distribution, density, and quantile functions of each model. The parameter values for any hypothesized model are estimated by minimizing the *maximum absolute difference* (MAD). The MAD is calculated as follows:

$$\text{MAD} = \max_{1 \leq n \leq N} \quad |p_n - G(w_{(n)})|. \tag{B.3}$$

   This optimization problem is solved numerically.

3. Compare the MAD of each hypothesized parametric model and select the one with the lowest MAD.

4. Evaluate the goodness-of-fit of the selected model.

   Krzysztofowicz (2014) suggests a validation of the selected model using a graphical analysis, the analysis of the calculated MAD, and the Kolmogorov-Sminorv (K-S) test. The graphical analysis involves plotting the empirical and the parametric distribution functions in order to judgmentally evaluate clear incompatibilities between the two functions.

The parametric distribution function obtained from implementing the steps above is a representation of the uncertainty associate with variate W. Appendix C describes the implementation of the distribution functions modeled in this research using the statistical software R.

Appendix C. Bayesian Forecasting Using R

The programming language and software environment R is an important tool for statistical analysis and data visualization. It is a powerful and free software available at `https://www.r-project.org/`. This section describes the R implementation used to model the Bayesian forecasters. Future users can add and improve this script to model their data.

C.1 Distribution, Density, and Quantile Functions

There are many packages in R to model and visualize distribution and density functions. However, in order to use these packages in the construction of a Bayesian forecaster, one must study the source code to be sure that the functions are in accordance to the theoretical framework. In order to avoid any complications, the functions in this study were constructed from the very beginning.

The following distribution functions were created based on the catalogue of univariate distributions in Krzysztofowicz (2014). They were utilized to construct the Bayesian forecasters in this research.

```
#Weibull density function
wb <- function(x, alpha, beta, eta){
  (beta/alpha)*((x-eta)/alpha)^(beta-1)*exp(-((x-eta)/alpha)^beta)
}

#Weibull distribution function
WB <- function(x, alpha, beta, eta){
  1 - exp(-((x-eta)/(alpha))^beta)
}

#Weibull quantile function
WBinv <- function(p, alpha, beta, eta){
  alpha*(-log(1-p))^(1/beta)+eta
```

```
}

#Inverted−Weibull density function
iw <− function(x, alpha, beta, eta){
  (beta/alpha)∗(alpha/(x−eta))^(beta+1)∗exp(−(alpha/(x−eta))^beta)
}

#Inverted−Weibull distribution function
IW <− function(x, alpha, beta, eta){
  exp(−(alpha/(x−eta))^beta)
}

#Inverted−Weibull quantile function
IWinv <− function(p, alpha, beta, eta){
  alpha∗(−log(p))^(−1/beta)+eta
}

#Laplace density function
lp <− function(x, alpha, beta){
  1/(2∗alpha)∗exp(−(abs(x−beta))/(alpha))
}

#Laplace distribution function
LP <− function(x, alpha, beta){
   ifelse (x <= beta, 0.5∗exp((x − beta)/(alpha)), 1 − 0.5∗exp(−(x−beta)/(alpha)))
}

#Laplace quantile function
LPinv <− function(p, alpha, beta){
   ifelse (p<=0.5, beta+ alpha∗log(2∗p), beta−alpha∗log(2∗(1−p)))
}

#LC1−WB density function
lc1wb <− function(y, alpha, beta, etaL, etaU) {
```

```
  x = log((etaU−etaL)/(etaU−y))
  1/(etaU−y)∗wb(x, alpha, beta, 0)
}

#LC1−WB distribution function
LC1WB <− function(y, alpha, beta, etaL, etaU) {
  x = log((etaU−etaL)/(etaU−y))
  WB(x, alpha, beta, 0)
}

#LC1−WB quantile function
LC1WBinv <− function(p, alpha, beta, etaL, etaU){
  etaU − exp(log(etaU − etaL) − WBinv(p, alpha, beta, 0))
}

#LC2−IW density function
lc2lw <− function(y, alpha, beta, etaL, etaU) {
  x = log((etaU−etaL)/(y−etaL))
  1/(y−etaL)∗lw(x, alpha, beta, 0)
}

#LC2−IW distribution function
LC2LW <− function(y, alpha, beta, etaL, etaU) {
  x = log((etaU−etaL)/(y−etaL))
  1 − LW(x, alpha, beta, 0)
}

#LC2−IW quantile function
LC2LWinv <− function(p, alpha, beta, etaL, etaU){
  exp(log(etaU − etaL) − LWinv(1−p, alpha, beta, 0)) + etaL
}

#LR1−LP density function
lr1lp <− function(y, alpha, beta, etaL, etaU) {
```

```
  x = log((y−etaL)/(etaU−y))
  (etaU−etaL)/((y−etaL)∗(etaU−y))∗lp(x, alpha, beta)
}

#LR1−LP distribution function
LR1LP <− function(y, alpha, beta, etaL, etaU) {
  x = log((y−etaL)/(etaU−y))
  LP(x, alpha, beta)
}

#LR1−LP quantile function
LR1LPinv <− function(p, alpha, beta, etaL, etaU) {
  (etaL + etaU∗exp(LPinv(p, alpha, beta)))/(1 + exp(LPinv(p, alpha, beta)))
}
```

The following function calculates the meta-Gaussian plotting positions developed by Krzysztofowicz (2014) and discussed in appendix B:

```
# meta−Gaussian plotting positions
ppointsMG <− function (n) {
  if (length(n) > 1L) n <− length(n)
  if (n>=2 & n<=3){
   t = 3.0193∗n^(−1.1018)+1
  }else  if (n>=4 & n<=5){
   t = 2.4035∗n^(−0.9096)+1
  }else  if (n>=6 & n<=10){
   t = 2.1408∗n^(−0.8423)+1
  }else  if (n>=11 & n<=20000){
   t = 1.9574∗n^(−0.8039)+1
  }else  if (n>20000){
   t = 1
  }
  if (n > 0)(((n − 1L:n +1)/(1L:n))^t +1)^(−1) else numeric()
}
```

Next, the function *normalPlot* produces a QQ plot using the meta-Gaussian plotting positions.

```
#QQ plots with meta−Gaussian plotting positions
normalPlot <− function (res) {
  res = sort(res)
  gaussianQuantiles = qnorm(ppointsMG(res))
  plot(gaussianQuantiles, res,
    main="", xlab="Gaussian Quantile", ylab="Sample Quantiles",
    las = 1)
  qqline(res)
}
```

C.2 Bayesian Forecasters

This subsection describes the implementation of two meta-Gaussian forecasters, the BPF_{Oct} and the $\text{BPF}_{Feb,}$ developed in section 6.5. The first model is constructed using one predictor and the second using multiple predictors. The remaining models developed in this research were similarly implemented by adjusting the functions and estimations according to the model.

```
### LOADING DATA
## Loading distribution functions parameters
dfPar<− read.csv("distPar.csv",sep=";", header=T)

## Loading predictor and predictand realizations
dfPnd<− read.csv("predictand.csv",sep=";", header=F)
dfPor1<− read.csv("conabOct.csv",sep=";", header=F)
dfPor2<− read.csv("conabFeb.csv",sep=";", header=F)
dfPor3<− read.csv("ibgeFeb.csv",sep=";", header=F)
dfPor4<− read.csv("conabMay.csv",sep=";", header=F)
dfPor5<− read.csv("ibgeMay.csv",sep=";", header=F)
```

```
### NORMAL QUANTILE TRANSFORMATION
dfNqtPnd<−qnorm(LC2IW(dfPnd,1.0777, 6.7246, 30,90), mean=0, sd=1)
dfNqtPor1<−qnorm(LC2IW(dfPor1,1.03, 16.07, 30, 90), mean=0, sd=1)
dfNqtPor2<−qnorm(LC1LL(dfPor2,0.45, 30.46, 30, 90), mean=0, sd=1)
dfNqtPor3<−qnorm(LC2IW(dfPor3,0.99, 18.99, 30, 90), mean=0, sd=1)
dfNqtPor4<−qnorm(LC2IW(dfPor4,1.05, 7.45, 30, 90), mean=0, sd=1)
dfNqtPor5<−qnorm(LC2IW(dfPor5,1.08, 6.76, 30, 90), mean=0, sd=1)

### OCTOBER BAYESIAN FORECASTER
##Fitting linear regression in the transformed space
regPor1<−lm(dfNqtPor1~dfNqtPnd[12:25])
b<−summary(regPor1)$coefficients[1]
a<−summary(regPor1)$coefficients[2]
sigma<−sqrt(moment(residuals(regPor1), order = 2, central = T))
resOct<−residuals(regPor1)

#Calculating posterior  coefficients
c1<− a/(a^2 + sigma^2)
c0<− (−a∗b)/(a^2 + sigma^2)
T2<− sigma^2/(a^2+sigma^2)

#Calculating 90 \% Central credible Interval: Transformed Space
CIU <− sigma∗qnorm(0.95, mean=0, sd = 1)
CIL <− sigma∗qnorm(0.05, mean=0, sd = 1)

#P−probability quantile function of X1
X1pProb <− function(p, w){
  Gh = LC2IW(w, 1.0777, 6.7246, 30, 90)
  Kinv = function(t){LC2IWinv(t, 1.0300, 16.0700, 30, 90)}
  Kinv(pnorm(a∗qnorm(Gh) + b + sigma∗qnorm(p)))
}

# Calculating 90 \% Central credible Interval: Original Space
```

```
CIUoriginal <− X1pProb(0.95, range2)
CILoriginal <− X1pProb(0.05, range2)
medianRegX1 <− X1pProb(0.5, range2)

# Posterior distribution function
PHIoct<−function(w, x){
  Gh = LC2IW(w,1.0777, 6.7246, 30, 90)
  Kbar = LC2IW(x, 1.03, 16.07, 30, 90)
  pnorm((qnorm(Gh) − c1∗qnorm(Kbar) − c0)/sqrt(T2))
}

# Posterior density function
phiOct<− function(w, x){
  gh = lc2iw(w,1.0777, 6.7246, 30, 90)
  Gh = LC2IW(w,1.0777, 6.7246, 30, 90)
  result = gh/(sqrt(T2)∗dnorm(qnorm(Gh)))∗dnorm(qnorm(PHIoct(w, x)))
}

# Posterior quantile function
PHIoctInv<−function(p, x){
  Kbar = LC2IW(x, 1.03, 16.07, 30, 90)
  LC2IWinv(pnorm(c1∗qnorm(Kbar)+ c0 + sqrt(T2)∗qnorm(p)),1.0777, 6.7246, 30, 90)
  }

#Forecasting using the BPFoct
seq1<−seq(30, 60, 0.1)
PHIoct47 <− PHIoct(seq1, 47)
PHIoct49 <− PHIoct(seq1, 49)
PHIoct52 <− PHIoct(seq1, 52)

### FEBRUARY BAYESIAN FORECASTER
##Mean, variance, and covariance
#manipulationg data set
numberLines<−max(length(dfNqtPor2), length(dfNqtPor3), length(dfNqtPnd))
```

```
dfFebFor = matrix(data=NA, nrow=numberLines, ncol=3)

start  = numberLines-length(dfNqtPor2)+1
for(i  in  start:numberLines){
    dfFebFor[i,  1]= dfNqtPor2[i-start+1]
}

start  = numberLines-length(dfNqtPor3)+1
for(i  in  start:numberLines){
    dfFebFor[i,  2]= dfNqtPor3[i-start+1]
}

start  = numberLines-length(dfNqtPnd)+1
for(i  in  start:numberLines){
    dfFebFor[i,  3]= dfNqtPnd[i-start+1]
}

dfFebFor = na.omit(dfFebFor)

#Function to compute ML mean
computeMean<-function(df){
  muQ = matrix(data=NA, nrow=ncol(df), ncol=1)
  for(i  in  1:ncol(df)){
    muQ[i, 1] = mean(df[,i],  na.rm=TRUE)
  }
  print(muQ)
}

#Function to compute ML variance
computeVar<-function(df){
  SigmaQ = mlest(df)$sigmahat
  print(SigmaQ)
}
```

```
#Function to compute multivariate density function fQ mean
computeMeanfQ<−function(df, v){
  #calculate MVN(muQ, SigmaQ) mean
  dim1=ncol(df)
  muQ = matrix(data=NA, nrow=dim1, ncol=1)
  for(i in 1:dim1){
    muQ[i, 1] = mean(df[,i], na.rm=TRUE)
  }
  #Calculate MVN(muQ, SigmaQ) var (require(mvnmle))
  SigmaQ = mlest(df)$sigmahat
  #calculate MVN(mufQ, SigmafQ) mean
  dim2=nrow(muQ)−1
  mufQ=matrix(data=NA, nrow=dim2, ncol=1)
  for(i in 1:dim2){
    mufQ[i, 1] = muQ[i, 1]+
      (SigmaQ[i, dim1]/SigmaQ[dim1, dim1])∗(v − muQ[dim1, 1])
  }
  print(mufQ)
}

#Function to compute multivariate density function fQ var
computeVarfQ<−function(df){
  #Calculate MVN(muQ, SigmaQ) var (require(mvnmle))
  SigmaQ = mlest(df)$sigmahat
  #Calculate MVN(mufQ, SigmafQ) var
  dim1=ncol(df)
  dim2=ncol(df)−1
  SigmafQ=matrix(data=NA, nrow=dim2, ncol=dim2)
  for(i in 1:dim2){
    for(j in 1:dim2){
      SigmafQ[i, j] = SigmaQ[i, j] −
        (SigmaQ[j, dim1]∗SigmaQ[i, dim1])/(SigmaQ[dim1, dim1])
    }
  }
```

```
  print(SigmafQ)
}

##Posterior Parameters
#Function to compute T
computePosT<-function(df){
  #Calculate SigmaQ)|(require(mvnmle))
  dim1=ncol(df)
  SigmaQ = mlest(df)$sigmahat
  #Calculate MVN(mufQ, SigmafQ) var
  dim2 = ncol(df)-1
  SigmafQ = matrix(data=NA, nrow=dim2, ncol=dim2)
  for(i in 1:dim2){
    for(j in 1:dim2){
      SigmafQ[i, j] = SigmaQ[i, j] -
          (SigmaQ[j, dim1]*SigmaQ[i, dim1])/(SigmaQ[dim1, dim1])
    }
  }
  #Calculate T
  sigma = SigmaQ[1:dim2, dim1]
  postT = sqrt((SigmaQ[dim1, dim1]^2)/(crossprod(sigma,solve(SigmafQ))%*%sigma +
     SigmaQ[dim1, dim1]^2))
  print(postT)
}
#Function to compute vector cT
computePostCT<-function(df){
  #Calculate SigmaQ)|(require(mvnmle))
  dim1=ncol(df)
  SigmaQ = mlest(df)$sigmahat
  #Calculate MVN(mufQ, SigmafQ) var
  dim2 = ncol(df)-1
  SigmafQ = matrix(data=NA, nrow=dim2, ncol=dim2)
  for(i in 1:dim2){
    for(j in 1:dim2){
```

```
      SigmafQ[i, j] = SigmaQ[i, j] -
        (SigmaQ[j, dim1]*SigmaQ[i, dim1])/(SigmaQ[dim1, dim1])
    }
  }
  #Calculate T
  sigma = SigmaQ[1:dim2, dim1]
  postT = sqrt((SigmaQ[dim1, dim1]^2)/(crossprod(sigma,solve(SigmafQ))%*%sigma +
    SigmaQ[dim1, dim1]^2))
  #Calculate Posterior CT
  postCT = c((postT^2)/SigmaQ[dim1, dim1])*crossprod(sigma, solve(SigmafQ))
  print(postCT)
}

#Function to compute vector c0
computePostC0<-function(df){
  #Compute muQ
  muQ = matrix(data=NA, nrow=ncol(df), ncol=1)
  for(i in 1:ncol(df)){
    muQ[i, 1] = mean(df[,i], na.rm=TRUE)
  }
  #Calculate SigmaQ|(require(mvnmle))
  dim1=ncol(df)
  SigmaQ = mlest(df)$sigmahat
  #Calculate MVN(mufQ, SigmafQ) var
  dim2 = ncol(df)-1
  SigmafQ = matrix(data=NA, nrow=dim2, ncol=dim2)
  for(i in 1:dim2){
    for(j in 1:dim2){
      SigmafQ[i, j] = SigmaQ[i, j] -
        (SigmaQ[j, dim1]*SigmaQ[i, dim1])/(SigmaQ[dim1, dim1])
    }
  }
  #Calculate T
  mu = muQ[1:dim2, 1]
```

```
  sigma = SigmaQ[1:dim2, dim1]
  postT = sqrt((SigmaQ[dim1, dim1]^2)/(crossprod(sigma,solve(SigmafQ))%*%sigma +
    SigmaQ[dim1, dim1]^2))
  #Calculate Posterior CT
  postCT = c((postT^2)/SigmaQ[dim1, dim1])*crossprod(sigma, solve(SigmafQ))
  #Calculate Posterior C0
  postC0 = postCT%*%(muQ[dim1,1]/SigmaQ[dim1, dim1]*sigma-mu)
  print(postC0)
}

#function to compute sufficiency Characteristic
computeSC<-function(df, i){
  #Calculate SigmaQ|(require(mvnmle))
  dim1=ncol(df)
  SigmaQ = mlest(df)$sigmahat
  #Calculate SC
  SC = SigmaQ[dim1, dim1]*((SigmaQ[dim1, dim1]*SigmaQ[i, i] -
    SigmaQ[i, dim1]^2))/(SigmaQ[i, dim1]^2)
  print(SC)
}

#function to compute the correlation gamma
computeGamma<-function(df, i){
  #Calculate SigmaQ|(require(mvnmle))
  dim1=ncol(df)
  SigmaQ = mlest(df)$sigmahat
  #Calculate correlation gamma
  gamma =
    abs(sign(SigmaQ[i, dim1])*(((SigmaQ[dim1, dim1]*(SigmaQ[dim1, dim1]*SigmaQ[i, i] -
    SigmaQ[i, dim1]^2))/(SigmaQ[i, dim1]^2))+1)^(-0.5))
  print(gamma)
}

##February Parameters
```

```
muFeb<−computeMean(dfFebFor)
sigmaFeb<−round(computeVar(dfFebFor), 3)
computeMeanfQ(dfFebFor, 40)
computeVarfQ(dfFebFor)
computePosT(dfFebFor)
ctFeb <− computePostCT(dfFebFor)
c0Feb <− computePostC0(dfFebFor)
computeSC(dfFebFor, 1)
computeSC(dfFebFor, 2)
computeGamma(dfFebFor, 1)
computeGamma(dfFebFor, 2)

#Posterior density function
phiFeb<− function(w, x2, x3){
  gh = lc2iw(w,1.0777, 6.7246, 30, 90)
  Gh = LC2IW(w,1.0777, 6.7246, 30, 90)
  Kbar2 = LC1LL(x2,0.4500, 30.4600, 30, 90)
  Kbar3 = LC2IW(x3, 0.9900, 18.9900, 30, 90)
  step1 = qnorm(Gh)
  step2 = ctFeb[1]∗qnorm(Kbar2)+ctFeb[2]∗qnorm(Kbar3)
  step3 = step1 − step2 − c0Feb
  result = gh/(sqrt(T2)∗dnorm(step1))∗dnorm(step3/sqrt(T2))
}

#Forecasting
seq1<−seq(30, 60, 0.1)
phiFeb1<−phiFeb(seq1, 50, 45)
phiFeb2<−phiFeb(seq1, 50, 50)
phiFeb3<−phiFeb(seq1, 50, 55)
phiFeb4<−phiFeb(seq1, 45, 50)
phiFeb5<−phiFeb(seq1, 50, 50)
phiFeb6<−phiFeb(seq1, 55, 50)
```

REFERENCES

Alpert, M. and Raiffa, H. (1982). A progress report on the training of probability assessors. In Kahneman, D., Slovic, P., and Tversky, A., editors, *Judgment under uncertainty: heuristics and biases*, chapter 21, pages 294–305. Cambridge University Press, Cambridge, United Kingdom.

Buchholz, D. D. (2004). Soil test interpretations and recommendations handbook. online, University of Missouri/College of Agriculture/Division of Plant Sciences. Originally written in 1983.

CEPEA (2018). PIB do Agronegócio Brasileiro de 1996 a 2018. `https://www.cepea.esalq.usp.br/br/pib-do-agronegocio-brasileiro.aspx`.

Clemen, R. T. (1996). *Making Hard Decisions: An Introduction to Decision Analysis*. Duxbury Press, 2nd edition.

Clemen, R. T. and Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2):187–203.

CONAB (2015). Acompanhamento da safra brasileira grãos. `https://www.conab.gov.br/info-agro/safras/graos`.

CONAB (2016). Acompanhamento da safra brasileira grãos. `https://www.conab.gov.br/info-agro/safras/graos`.

CONAB (2018). Acompanhamento da safra brasileira grãos. `https://www.conab.gov.br/info-agro/safras/graos`.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, Inc.

DeGroot, M. H. (1988). A Bayesian view of assessing uncertainty and comparing expert opinion. *Journal of Statistical Planning and Inference*, 20(3):295–306.

EMBRAPA (2013). Tecnologia de produção de soja - região central do brasil 2014. Technical Report 16, Empresa Brasileira de Pesquisa Agropecuária - Soja, Londrina, PR/Brazil.

Goldstein, D. G. and Rothschild, D. (2014). Lay understating of probability distributions. *Judgment and Decision Making*, 9(1):1–14.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (2000). *Understanding Robust and Exploratory Analysis*. John Wiley & Sons, wiley classics library 2000 edition.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417.

IBGE (2002). *Pesquisas Agropecuárias - Série Relatórios Metodológicos*. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brazil, 2nd edition.

IBGE (2017a). *Divisão Regional do Brasil em Regiões Geográficas Imediatas e Regiões Geográficas Intermediárias*. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brazil.

IBGE (2017b). Valor bruto da produção - lavouras e pecuária - brasil.

IBGE (2018). *Pesquisas Agropecuárias - Série Relatórios Metodológicos*. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brazil, 3rd edition.

IBGE (2019). Levantamento sistemático da produção agrícola: Tabela 6588 - série histórica da estimativa anual da área plantada, área colhida, produção e rendimento médio dos produtos das lavouras.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2017). *An Introduction to Statistical Learning with Applications in R*. Springer, springer texts in statistics edition.

Keeney, R. L., Sarin, R. K., and Winkler, R. L. (1984). Analysis of alternative national ambient carbon monoxide standards. *Management Science: Risk Analysis*, 30(4):518–528.

Kelly, K. S. and Krzysztofowicz, R. (1995). Bayesian revision of an arbitrary prior density. In *Proceedings of the section on Bayesian Statistical Science*, pages 50–53. American Statistical Association.

Kelly, K. S. and Krzysztofowicz, R. (1997). A bivariate meta-Gaussian density for use in hydrology. *Stochastic Hydrology and Hydraulics*, 11(1):17–31.

Krzysztofowicz, R. (1999). Bayesian forecasting via deterministic model. *Risk Analysis*, 19(4):739–749.

Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, (1-4):2–9.

Krzysztofowicz, R. (2014). Bayesian meta-Gaussian forecasters. Course pack for: SYS 7075 Bayesian forecast-decision theory. University of Virginia, Charlottesville, VA.

Krzysztofowicz, R. (2016). Probabilistic forecasts and optimal decisions. Course pack for: SYS 3060 Stochastic decision models. University of Virginia, Charlottesville, VA.

Krzysztofowicz, R. and Evans, W. B. (2008). Probabilistic forecasts from the national digital forecast database. *Weather and Forecasting*, 23(2):270–289.

Krzysztofowicz, R. and Kelly, K. S. (2000). Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research*, 36(11):3265–3277.

Krzysztofowicz, R. and Reese, S. (1991). Bayesian analyses of seasonal runoff forecasts. *Stochastic Hydrology and Hydraulics*, (4):295–322.

Lee, C., Herbek, J., Murdock, L., Schwab, G., Green, J. D., and Martin, J. (2007). Corn and soybean production calendar. Technical report, Cooperative Extension Service - University of Kentucky - College of Agriculture.

Liu, J. (2018). *Bayesian System Averaging: A Theory Unifying Bayesian Forecasting System and Bayesian Model Averaging Methods*. PhD thesis, University of Virginia.

Mallarino, A. P., Sawyer, J. E., and Barnhart, S. K. (2013). *A General Guide for Crop Nutrient and Limestone Recommendations in Iowa*. Iowa State University Extension and Outreach.

Maranzano, C. J. (2006). *Bayesian Meta-Gaussian Models For Data Analysis And Probabilistic Forecasting*. PhD thesis, University of Virginia.

Maranzano, C. J. and Krzysztofowicz, R. (2008). Bayesian re-analysis of the challenger o-ring data. *Risk Analysis*, 28(4):1053–1067.

MDIC (2018). Sistema de análise das informações de comércio exterior. Ministério do Desenvolvimento, Indústria e Comércio Exterior (MDIC).

Morris, P. A. (1974). Decision analysis expert use. *Management Science*, 20(9):1233–1241.

Morris, P. A. (1977). Combining expert judgments: A Bayesian approach. *Management Science*, 23(7):679–693.

USDA (1999). Understanding USDA crop forecasts. Miscellaneous Publication 1554, United States Department of Agriculture - National Agricultural Statistics Service and Office of the Chief Economist / World Agricultural Outlook Board, Washington, D.C.

USDA (2006). How the WASDE is prepared.

USDA (2012). The yield forecasting program of NASS. Technical report, United States Department of Agriculture - National Agricultural Statistics Service - Statistical Methods Branch.

USDA (2019). Foreign agricultural service: World agricultural production. *Agricultural Service - Circular Series*, WAP 2-19.

Winkler, R. L. (1968). The consensus of subjective probability distributions. *Management Science*, 15(2):B61–B75.

Winterfeldt, D. V. and Schweitzer, E. (1998). An assessment of tritium supply alternatives in support of the us nuclear weapons stockpile. *Interfaces*, 28(1):92–112.