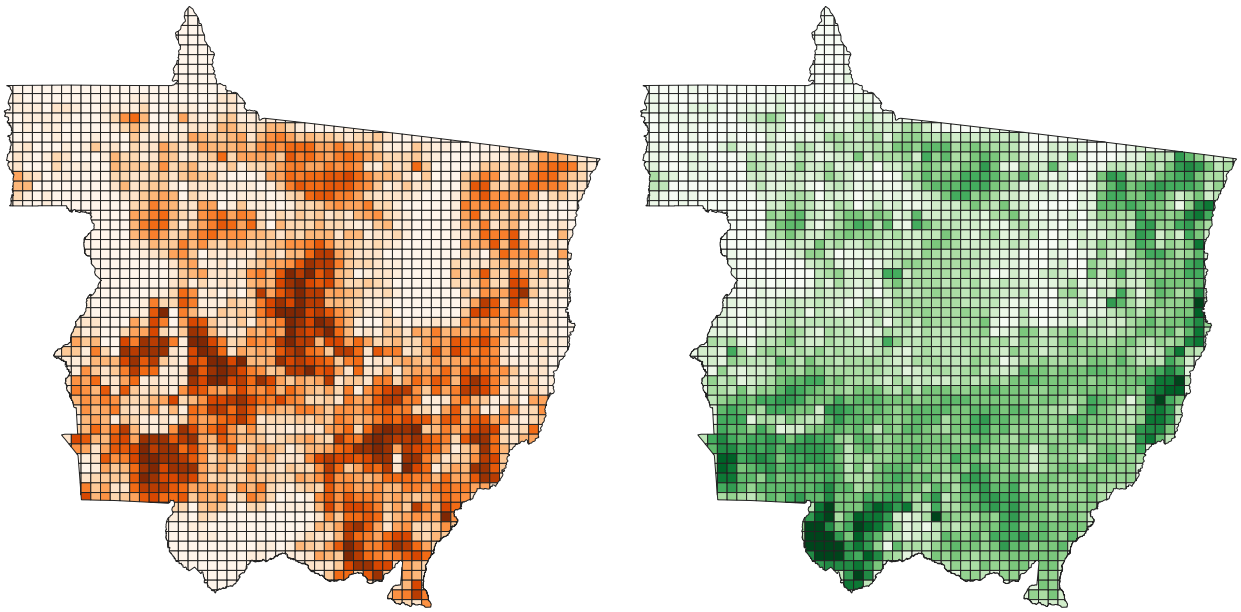
CHALMERS
UNIVERSITY OF TECHNOLOGY



Spatial deaggregation of agricultural statistics using constrained cross entropy minimisation

Master's thesis in Complex Adaptive Systems

GABRIELLA GRENANDER

Department of Space, Earth and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020

Master's thesis 2020

Spatial deaggregation of agricultural statistics using constrained cross entropy minimisation

GABRIELLA GRENANDER

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Space, Earth and Environment
Physical Resource Theory
Chalmers University of Technology
Gothenburg, Sweden 2020

Spatial deaggregation of agricultural statistics using constrained cross entropy minimisation
GABRIELLA GRENANDER

Supervisor: Martin Persson, Space, Earth and Environment
Examiner: Kristian Lindgren, Space, Earth and Environment

Master's Thesis 2020
Department of Space, Earth and Environment
Physical Resource Theory
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Cropland (orange) and pasture (green) statistics distribution in the Brazilian state Mato Grosso found with means of cross entropy minimisation

Typeset in LaTeX
Gothenburg, Sweden 2020

Spatial deaggregation of agricultural statistics by means of constrained cross entropy minimisation

GABRIELLA GRENANDER
Department of Space, Earth and Environment
Chalmers University of Technology

Abstract

Every year, millions of hectares of forests disappears in the tropics, something that has impacts on both local and global scale. Locally, the forest loss impacts ecosystem services such as water, energy and food security. Globally, the tropical forest loss releases large amounts of carbon dioxide into the atmosphere, making it the second largest driving factor of climate change, after combustion of fossil fuels.

In research aiming to catalogue the main driving factors behind deforestation linked to land usage, there is a need to spatially downscale agricultural statistics from large statistical reporting units into smaller, in order to increase the accuracy of the driver identification. This thesis will investigate how one feasibly can do this deaggregation. Specifically, we have used constrained cross entropy minimisation, a method which aims to minimise the difference between a target distribution and a prior assessment of the spatial distribution of land usage, while also taking certain limiting constraints into account.

For our investigations, we have chosen to focus on the Brazilian state Mato Grosso, a region that has experienced deforestation due to the spread of agricultural land uses. The prior will be created from land cover maps, generated from satellite imagery. Here we evaluated different prior preprocessing methods, finding that a rescaling of the land cover classifications using the land cover map's confusion matrix was the method with the most promising result, capturing the general shape of the true distribution while also having a reasonable area distribution. It was however also noted that this method has a difficulty, common when dealing with land cover classification, in separating certain similar land uses from one another. It is suspected that this is due to the quality of the prior, and it would be interesting to investigate further how adjustments to the prior creation would improve the results found in this thesis.

Keywords: Cross entropy minimisation, agricultural statistics, deaggregation, deforestation, land use, prior, certainty parameter, confusion matrix.

Acknowledgements

First and foremost, I would like to thank my supervisors, Martin Persson and Florence Pendrill. You have both been really supportive through the ups and downs of the writing of this thesis, brought great comments and new ideas to try, and helped me when I have torn my hair while QGIS has had a mind of its own. I am very thankful to have had the opportunity to work with you. I also would like to thank Kristian Lindgren for your help in understanding information theory and our discussions regarding this thesis' method.

A heartfelt thanks to all the wonderful maths teachers I have met over the years, who were committed to their students and always tried to inspire, and show how cool and useful maths can be. Special thanks to Marie and Stina who knew I wanted to be an engineer before I even knew it, and brought the idea to my mind.

My gratitude to Akademihälsan, Studienämnden Fysik, and all masseurs and physiotherapists I have had the pleasure to meet. Student mental and physical health is so important, and so often put aside as irrelevant by those it should concern. I am glad to have had you looking out for my well-being during these years.

I am grateful for all the incredible people I have encountered during my time at Chalmers, through school activities and through extracurricular activities. You are too many to mention all by name, but know I am happy to have you in my life. Without all lovely moments I have had with you while singing, sewing, board gaming, coffee drinking and simply goofing around, my time here at Chalmers would have been so very dull. I would especially like to thank Emmy Lundberg, Anna Petersson, Ida Pettersson, Victoria Johansson, Joanna Langer, Ellinor Hansson and Annie Olsson; you have been a great support during the production of this thesis, and in life in general. I love you all.

Special thanks to my family; Annika, Peter, Alexander, Anja, Ingrid, Bengt, Gert, Göran. You have always supported me, both in my academic endeavours and in life. I am really grateful to have you as my stable anchor in this world.

Thanks to Kasper, who with a buff with his nose and a happy meow can bring sunshine on the cloudiest of days. And lastly, I would like to thank Hampus Renberg Nilsson. Your optimistic outlook on life, encouraging words and support mean a lot to me. I am very happy to have you with me every day.

Gabriella Grenander, Gothenburg, August 2020

Contents

List of Figures

List of Tables

1

Introduction

Global forest cover is on the decline; millions of hectares is lost annually, with the largest net loss being found in South America [1], [2]. This can partly be attributed to commodity-driven deforestation - permanent conversion of forests to for example agriculture or mining - and forestry - forest management within for example tree plantations - where each constitutes a quarter to a third of the the global forest cover [1]. In South America, commodity-driven deforestation accounts for 64% of the region's total tree cover loss [1].

This has several implications, locally around the deforestation, but also globally. An increase in farmland and industrialised agriculture brings with it an increased usage of fertilisers and pesticides [3]. This leads to decreased water quality due to the chemicals from these products seeping into freshwater sources, which in turn causes, for example, limited access to freshwater for humans and fish habitats disappearing [4]. Soil erosion increases with disappearing forests, which leads to the release of mercury into rivers, which makes fish in the region dangerous to eat [5]. Large scale deforestation of rain forests might yield a warmer and drier climate in the region [6]. Habitats of local communities of species changes with the changing land use, leading to some species - some which might only exist in this specialised region - diminishing in numbers, or even become extinct[7]–[9]. Habitat modification might on the other hand also lead to increased populations of species that might be hazardous to humans living in the area; it has been found that e.g. mosquitoes and ticks adapt well to the changes, leading to increased cases of malaria and Lyme disease in affected areas [5]. New viruses, such as the SARS-CoV-2 virus which has lead to the for this thesis current COVID-19 pandemic, spills from their host animals to humans either by directly handling these animals, or indirectly via farm animals. Tropical forest edges are a major area where such viruses take hold in humans, and with an increase in habitat fragmentation and deforestation, humans are more likely to come into contact with wildlife, leading to higher risk of transmission of diseases [10].

Since the effects of deforestation varies spatially (for example due to carbon dioxide emissions and loss of biodiversity), a better approximation of where the loss of forest occurs and what kinds of land use is driving it might generate a better assessment of the environmental effects, a better decision basis for policymakers, companies and consumers. It is therefore of interest to explicitly determine these factors: what the driving factors of deforestation are, where they are located spatially, and the magnitude at which they are driving it. Pendrill, Persson, Godar, *et al.* [11] has proposed a model which quantifies to what extent different agricultural produce and forest plantations are driving deforestation

(*land-balance model*). This model is at the moment on an aggregated country-based level, but has been noted that the model's performance varies with the degree of spatial deaggreation, meaning smaller areas better represents the land uses that are driving the deforestation [12]. Since many statistics regarding land usage is found on a country based level at its highest resolution, it is desirable to downscale these statistics. The aim of this thesis is to investigate how one feasible can deaggregate agricultural statistics from larger statistical reporting units (countries, states etc) into smaller areas, while also taking limiting factors such as area restrictions and estimations of spatial distribution into consideration.

Deaggregation of aggregated agricultural statistics has been done in other papers, such as by You and Wood [13] and You, Wood, Wood-Sichra, *et al.* [14], where they by means of cross entropy minimisation downscaled country level statistics into smaller regions. Their investigations however concerned only cropland and mainly different kinds of crops and production systems, while we will look at cropland as a whole, alongside with pastures. Their prior estimation of the land use distribution, which is needed for cross entropy minimisation, was based on population density and suitability, while we will create priors from land cover spatial distribution data generated from satellite imagery.

There are several ways to down-scale aggregated data, such as for example artificial neural networks and linear regression. We chose to use cross-entropy minimisation, inspired by You and Wood [13], in this work. Since Pendrill, Persson, Godar, *et al.* [11] identified cropland and pastures to be two of the main driving factors of deforestation, it is these deforestation driving land uses that will be considered in this work. There are other potential drivers such as forest plantations, logging and mining, but they will be excluded from this study. We will focus our investigation the state Mato Grosso in Brazil. Mato Grosso has experienced a lot of deforestation over the last twenty years - about 19% of its humid primary forest has been lost between the years of 2002 and 2019 [15] - mainly due to cropland- and pasture expansion, which made it a topical example. We also have validation data for the region, which will aid us in our evaluation of the performance of our method [16].

2

Theory

2.1 A short introduction to information theory

Before looking at the method used in this thesis, it is helpful to have some understanding of information theory. Briefly summarised, information theory is concerned with measures of information associated with distributions of random variables, and we will later see that this is one of the ideas that is behind the method chosen in this work.

2.1.1 What is information?

When a person makes an observation of a system with an unknown state, that person receives information about that system. How much information is received - how "surprising" that event is - depends on how likely that observation was. For example, as is illustrated in Figure 2.1, a weather report predicting warm weather in a tropical area does not contain a lot of information, but a report predicting a snowstorm in the same location contains a lot of information. This of course assumes that one has a previous understanding of how the weather usually is in the tropics - it usually is warm in the tropics, so it is the most likely outcome of an observation. If one has no previous knowledge of the tropics and its climate, it would be a reasonable assumption to make initially that warm and cold weather has an equally likely chance of being observed, thus either observation would give us the same amount of information about the system.

In order to reach a mathematical definition of information, let X be a random discrete variable describing the state of the system of interest, with a set of outcomes $\mathcal{X} = \{x_1, x_2, ..., x_n\}$ (often in information theory called an alphabet). The random variable has a probability mass function (the discrete equivalent of probability density function) $p(x) = \Pr(X = x),\ x \in \mathcal{X}$, which is the a priori probability that an event in the event space would occur. The measure of *information* that is gained by observing an event $x \in \mathcal{X}$ is defined as

$$I(x) \triangleq \log_b \left(\frac{1}{p(x)} \right). \tag{2.1}$$

The base b of the logarithm defines the unit of the information and can take any value; some choices are 10 (with unit name *dits*), e (*nats*) and 2 (*bits*). More often, one does not write the base of the logarithm, but instead simply write

$$I(x) = \log \left(\frac{1}{p(x)} \right), \tag{2.2}$$

(a) *Warm weather in a tropical area - low amount of information gained about the system.*

(b) *Cold weather in a tropical area - high amount of information gained about the system.*

Figure 2.1: *The amount of information gained by observing a system depends on its state and that state's likelihood of happening. The more unlikely an event is, as for example observing snow instead of sun in a tropical area, the more information is gained.*

and let the unit used at the end of the calculation signal what logarithm base was used.

2.1.2 Entropy in information theory

From Equation 2.2 it is concluded that the amount of information that is gained by observing a system is dependent on what system state that is observed. It can be useful to know the average information gained by observing an event from our random variable; if information is a measure of how surprising an observed event is, this proposed measure would then be a measure of how surprising the entire random variable is. This measure is called the *entropy* of X and is defined as

$$H(X) \triangleq \sum_{x \in \mathcal{X}} p(x) \log \left(\frac{1}{p(x)} \right) = \sum_{x \in \mathcal{X}} p(x) I(x). \tag{2.3}$$

In any case where $p(x) = 0$, we will say that $H(X) = 0$ (for proof of why it is reasonable to say so, see Appendix A). Thus, the entropy is a weighted sum of the information gained from each of the possible observable events and the probability of that event happening.

We can rewrite this expression for entropy in terms of expected value instead; we remember that the expected value of a random variable X is defined as

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x p(x), \tag{2.4}$$

and for a random variable $g(X)$

$$\mathbb{E}\left[g(X)\right] = \sum_{x \in \mathcal{X}} g(x)p(x). \tag{2.5}$$

If we take $g(X) = \log \frac{1}{p(X)}$, the entropy can be written as an expected value of the random variable X:

$$\begin{aligned} H(X) &= \sum_{x \in \mathcal{X}} p(x) \log\left(\frac{1}{p(x)}\right) = \\ &= \mathbb{E}\left[\log\left(\frac{1}{p(X)}\right)\right] = \\ &= \mathbb{E}\left[\log I(X)\right]. \end{aligned} \tag{2.6}$$

The entropy of a random variable can thus be interpreted as the average amount of information that is required to describe the random variable; in other words, the lack of knowledge we have about the system which needs to be found to fully determine its state.

We see from Equation 2.6 that the entropy will always be larger than or equal to 0 (since $p(x)$ is between 0 and 1).

2.1.3 Relative entropy

Now that we have a way of measuring the average information gained by observing an event from a probability distribution, it is useful to have a measure that compare two probability distributions with each other. The *relative entropy*, which is also known as *Kullback-Leibler divergence*, between two probability distributions of the same random variable is a measure of the distance between the distributions - how similar they are to one another. For a discrete random variable X, the relative entropy between the distributions $p(x)$ and $q(x)$ is defined as

$$K(p;q) = \sum_{x \in \mathcal{A}} p(x) \log \frac{p(x)}{q(x)} = \mathbb{E}_p\left[\log\left(\frac{p(X)}{q(X)}\right)\right] \tag{2.7}$$

where $\mathbb{E}_p$ is the expected value of random variable X with respect to distribution $p(x)$. Another, perhaps more intuitive, way of seeing the Kullback- Leibler divergence is: "How much more information would be needed in order to accurately describe a random variable if we use distribution q instead of the true distribution p for the variable?".

It can be shown that the relative entropy between to probability mass functions must be larger or equal to 0, with equality if and only if the two mass functions are equal for all possible input values. A proof of this can be found in Cover and Thomas [17, p.26].

2.2 Cross-entropy minimisation

Suppose that one has a system with a set number of possible states of which the system can take form. The true probabilities of these states occurring is not known, but an estimation of the probabilities is available. Later on, new information is obtained concerning the limits - the constraints - of the system and its possible states. It is desirable to update

our understanding of the states' probabilities, which is where cross entropy minimisation comes in.

To put this in mathematical terms, let the set of the possible states of the system be denoted $\mathcal{X} = \{x_1, x_2, ..., x_n\}$ and the true distribution of the probabilities of these states to be observed be denoted $p^*(x)$. The true distribution is unknown, but a prior $\pi(x)$ exists. The new information obtained are constraints on the distribution p^*, and the aim is now to obtain a distribution $p(x)$ that is the best estimate of p^* given the information that is available. The cross entropy minimisation is then

$$\min_{p(x)} \quad \sum_{x \in \mathcal{X}} p(x) \log(p(x)) - \sum_{x \in \mathcal{X}} p(x) \log(\pi(x)) \tag{2.8a}$$

$$s.t. \quad \text{All constraints are satisfied,} \tag{2.8b}$$

meaning that we wish to minimise the relative entropy between our target distribution $p(x)$ and our prior estimation of the distribution $\pi(x)$ with respect to our target distribution, while satisfying our limiting constraints.

There are several names for this minimisation method - *Kullback-Leibler-divergence minimisation* and *relative entropy minimisation* are two such examples. Since You and Wood [13] and You, Wood, Wood-Sichra, *et al.* [14], the articles inspiring our choice of using this method, calls this method *Cross entropy minimisation*, we will do so as well through the continuation of this thesis.

2.3 Confusion matrix

When classifying data there almost always a risk that there will be some error in the process, where elements that belong to a certain class is misclassified as another class. This can happen due to - among other things - lack of enough information about the element to make a correct classification, two classes being very similar to one another and there is a difficulty to distinguish between them, or simply human error.

In order to visualise how well any specific algorithm performs at its classification, it is common to utilise a confusion matrix (also called error matrix). In order explain this tool, we will use an example. Let us say that we have an algorithm with a purpose to distinguish images of lions, tigers, jaguars, leopards, and snow leopards from one another. In order to know the performance, we run the algorithm on a dataset of images (in this case 143 images) where the true classification is already known. We then summarise all the classifications the algorithm made and compare them to the true classification. Illustrated in Table 2.1 is a numerical example displaying this summary.

We can read this matrix in several ways:

- On the diagonal (elements marked in green) we see the count of images that were correctly classified. For example, 23 images were correctly classified as *Lion*.

- Reading down along the *Lion* column, we see that 7 images that should have been classified as *Lion* were classified as something else; these classifications are called *omission errors* or *false negatives*.

- Reading across the *Lion* row, we see that 7 images that should have been classified as something else were classified as *Lion*; these classifications are called *commission errors* or *false positives*

- Displayed along the lowest row is the *Product accuracy*. This visualises the probability that an element belonging to each given class is correctly classified. The *Lion* class has a producer accuracy of 77% in this algorithm.

$$\text{Product accuracy of class } i = \frac{\#\text{ Elements correctly classified as class } i}{\#\text{ Elements belonging to class } i}$$

- Displayed along the rightmost column is the *User accuracy*. This visualises the probability that an element classified as each given class actually belongs to that class. The *Lion* class has a user accuracy of 72% in this algorithm.

$$\text{User accuracy of class } i = \frac{\#\text{ Elements correctly classified as class } i}{\#\text{ Elements classified as class } i}$$

- The element in the bottom right corner (marked in blue) is the *overall accuracy* of the algorithm. It visualises the amount of elements overall that were correctly classified. This algorithm has an overall accuracy of 53%.

$$\text{Overall accuracy} = \frac{\#\text{ Elements correctly classified}}{\#\text{ Elements}}$$

Table 2.1: *Example of a confusion matrix created by running an algorithm designed to distinguish images of lions, tigers, jaguars, leopards, and snow leopards on a dataset of 143 images with known true classification.*

		Truth						
		Lion	**Tiger**	**Jaguar**	**Leopard**	**Snow Leopard**	**Total**	**User acc. (%)**
Prediction	**Lion**	23	4	3	2	0	32	**72**
	Tiger	4	14	2	1	1	22	**63**
	Jaguar	1	2	14	7	7	31	**45**
	Leopard	2	4	12	20	8	46	**43**
	Snow Leopard	0	0	3	4	5	12	**42**
	Total	30	24	34	34	21	**143**	
	Product acc. (%)	**77**	**58**	**41**	**59**	**24**		**53**

3

Methods

3.1 Spatial allocation model

The aim of this thesis is to disaggregate information on a coarser scale into a finer scale; to downscale low resolution data into a higher resolution. More specifically, our goal with the spatial allocation model is to take agricultural statistics gathered in larger reporting units, such as on a state level illustrated in Figure 3.1a, and disaggregate the data into pixels on a gridded map of the same reporting area, as illustrated in Figure 3.1b.

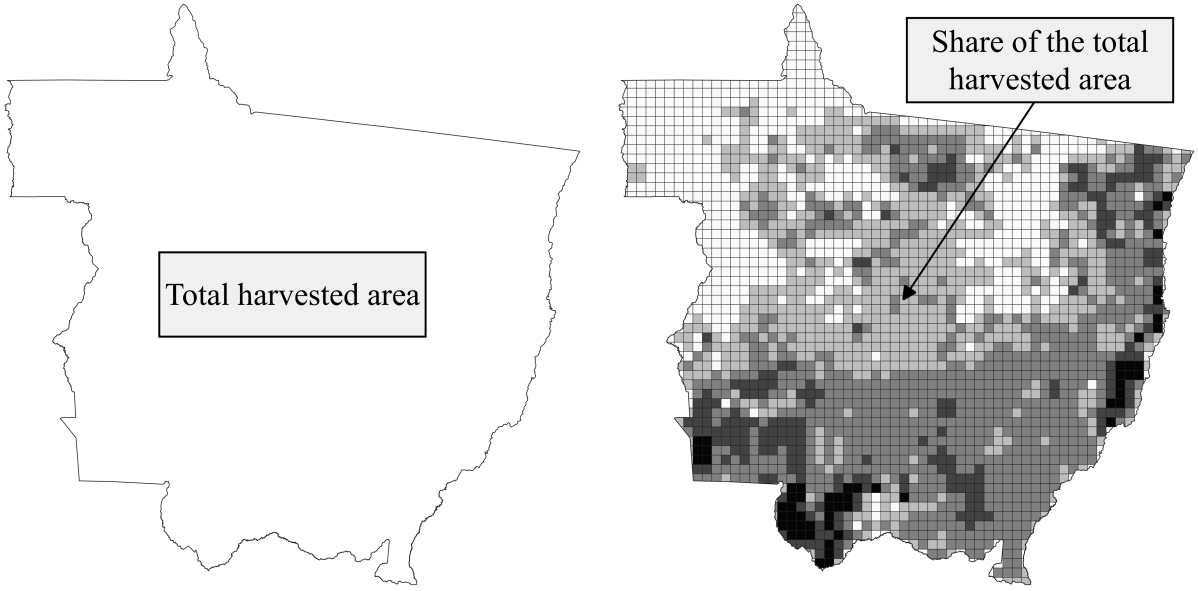


Figure 3.1: *Agricultural statistics is often available in large statistical reporting units, such as on a state level in Brazil as seen in (a). Our aim with the spatial allocation model is to disaggregate these statistics into smaller units, as seen in (b).*

Reported harvested areas of cropland and pasture land use in the region is collected from IBGE as described in Section 3.2.1. In our model we will call the respective land use areas L_j, where j represents the various land uses.

With the agricultural statistics in hand, we wish to have a measure of how much of the total harvested areas in the whole area that should be assigned to each of the grid pixels.

Let σ_{ij} be the share of the total area of land use j assigned to pixel i. It must be the case that

$$\sum_i \sigma_{ij} = 1, \ \forall j, \tag{3.1}$$

meaning that σ_{ij} can be seen as a percentage of the total area of land use j assigned to pixel i. We can also, due to the fact that σ_{ij} sums to one for every j, view each column j in the matrix σ as a distribution of the total land use area in the region of interest over the pixels i.

The area in pixel i assigned to land use j, which we will denote α_{ij}, are then, with our notation, defined to be

$$\alpha_{ij} = \sigma_{ij} \times L_j. \tag{3.2}$$

The ESA-CCI global land cover map (described in further detail in Section 3.2.2) is translated into pixelated area shares of the whole area of investigation of cropland, pasture and other land uses; a detailed explanation of the three methods used in this thesis to this end can be found in Section 3.3. This will serve as our prior assessment of the area shares of land use j in pixel i, which we will denote π_{ij}.

In order to find the area shares σ_{ij}, we then minimise the cross-entropy for the problem according to the following objective function and constraints:

$$\min_{\sigma_{ij}} \quad \sum_i \sum_j \sigma_{ij} \log(\sigma_{ij}) - \sum_i \sum_j \sigma_{ij} \log(\pi_{ij}) \tag{3.3a}$$

$$s.t. \quad \sum_i \sigma_{ij} = 1, \qquad \forall j \tag{3.3b}$$

$$\sum_j \alpha_{ij} \leq A_i, \qquad \forall i \tag{3.3c}$$

$$0 \leq \sigma_{ij} \leq 1, \qquad \forall i, \forall j \tag{3.3d}$$

where

- $i : i = 1, 2, 3, ...$ defines the pixels in the region of interest
- $j : j = \{\text{cropland, pastures}\}$ defines the land uses of interest

Equation 3.3a is the objective function of the minimisation problem with optimisation variable σ_{ij}, where the difference between the entropy of the area share distribution σ_{ij} and the cross entropy between the area share distribution σ_{ij} and the prior assessment of the distribution p_{ij} is desired to be as small as possible. Equation 3.3b is an equality constraint which ensures that the area shares σ_{ij} for each land use j add up to 1, making certain that we allocate the exact amount of reported area to the pixels i. The third constraint, Equation 3.3c, is an inequality constraint which ensures that we do not allocate more land use area in a pixel than what is available in that pixel (A_i). Lastly, Equation 3.3d, constrains the optimisation variable σ_{ij} to a range that is feasible.

The minimisation is done in MATLAB, using the built in function `fmincon` - a function designed to find minima of constrained nonlinear multivariate functions [18]. The function finds the optimum with the Interior Point algorithm. All options are set to the default settings given by the function, with the exceptions of:

- `MaxFunctionEvaluations`: 600000. We observed that `fmincon` often ended its optimisation before any minimum was found with the default conditions, so a higher number of function evaluations was chosen to increase the chances of a successful optimisation.

- `OptimalityTolerance`: 1.0000e-05. We loosened the tolerance of first-order optimality due to the fact that the optimisation runs often ended in not finding an optimum, in order to increase the chances of a successful run.

An initial value is needed in order for `fmincon` to be able to run. This is created by drawing uniformly random numbers $\tilde{\sigma}_{0;ij}$ for the number of optimisation variables we have (number of grid pixels times the number of land uses), and normalising them so that they for each land use sum to 1, such as:

$$\sigma_{0;ij} = \frac{\tilde{\sigma}_{0;ij}}{\sum_i \tilde{\sigma}_{0;ij}}, \quad \forall j, \tag{3.4}$$

where σ_0 denotes the initial conditions, and the subscripts i and j denotes grid pixel i and land use j respectively.

When `fmincon` ends with a successful run, we know that a minimum has been found. We can however not know that the global minimum has been found, only that the first-order optimality measure was less than the optimality tolerance and that the maximum constraint violation was less than the constraint tolerance. In order to increase the likelihood that we find the global minimum, we run the optimisation algorithm 20 times for different vales on σ_0 for each run and taking the result from the run that generated the lowest optimisation result.

3.2 Data sources

In order to run the spatial allocation model, we need a region that has agricultural statistics that we wish to deaggregate. We chose to focus our study to the Brazilian state Mato Grosso, a region that has experienced deforestation as well as a spread in agricultural land uses over the last 20 years, making it a relevant subject of investigation. In this section the various data sources used in this thesis will be presented, alongside motivations of why these particular sources have been chosen.

3.2.1 Land use statistics

This thesis will use two sets of agricultural data provided by *Instituto Brasileiro de Geografia e Estatística* (IBGE), the Brazilian Institute of Geography and Statistics. Firstly, we will utilise agricultural census data as our statistics that we wish to downscale. Secondly, we will use a map over the spatial distribution of the aforementioned census data as a basis for our validation.

3.2.1.1 Agricultural census data

Censo Agropecuário is a dataset containing a census of agricultural establishments in Brasil for the years of 1995/1996, 2006 and 2017 [19]. We chose to focus on the year

Table 3.1: *Variables used from IBGE data set* Censo Agropecuário *for the year 2006. The statistics downloaded were for the state Mato Grosso.*

Variável: Área dos estabelecimentos agropecuários (Hectares) **Variable: Area of agricultural establishments (Hectares)**		
Utilização das terras **Land use**		
Class name (Portuguese)	**Class name (English)**	**Aggregated statistics**
Lavouras - permanentes	Cropland - permanent	Cropland
Lavoras - temporárias	Cropland - temporary	
Lavouras - área plantada com forrageiras para corte	Cropland - animal feedstuff	
Lavouras - área para cultivo de flores (inclusive hidroponia e plasticultura), viveiros de mudas, estufas de plantas e casas de vegetação	Cropland - area for growing flowers (including hydroponics andplasticulture), seedling nurseries, and greenhouses	
Pastagens - naturais	Pastures - natural	Pastures
Pastagens - plantadas degradadas	Pastures - degraded	
Pastagens - plantadas em boas condições	Pastures - good condition	
Sistemas agroflorestais - área cultivada com espécies florestais também usada para lavouras e pastoreio	Agroforestry systems - area cultivated with forest also used for crops and grazing.	50% Cropland 50% Pastures

2006, due to the fact that IBGE also provides a map over the spatial distribution of the census data for this year, which we can use for validation. The dataset provides information about many variables concerning agriculture; the variable of interest to this work was the total area of cropland and pastures in our area of investigation [20].

The variables chosen are detailed in Table 3.1, and the statistics downloaded were for the state Mato Grosso. The land use class values marked under *Aggregated statistics* as *Cropland* and *Pastures* were summed together and used as our statistic of how much cropland respective pastures (in hectares) existed in 2006 in Mato Grosso. The value of the class *Agroforestry systems*, which concerns a mixture of forests, cropland and pastures, were divided evenly between and added to our *Cropland* and *Pasture* statistics.

3.2.1.2 Map over agricultural census data

In order to evaluate or results, it is desirable to have a map over the actual spatial distribution of the agricultural land uses. However, no such true image map was found in our investigations. Satellite imagery maps such as MapBiomas, which offer pixelated land cover maps over Brazil, are based on a classification process where the land uses are estimated from different data sources, with a possibility of error in classification [21]. Maps where census data gatherers have been in an area and determined its land use on a high resolution scale are very labour intensive, which might explain the lack of such data

sources. The closest we have found is *Mapa da utilização da terra no Censo Agropecuário* [16], which is a map generated by the data gathered from the data set *Censo Agropecuário* presented in Section 3.2.1.1. Each segment in the map is a census tract, a territorial unit used by the IBGE [22]. Each census tract is assigned a class according to the following rules [23]:

- If a land use class occupies more than 50% of the census tract, that census tract is assigned to be classified as that land use class. For example, if the majority of a census tract is found to be cropland, the whole census tract is said to be *Cropland*.

- If no land use class occupies more than 50% of the region, one of the mixture classes that best fit the land use distribution within the region will be assigned to be the census tract's classification. For example, if a third of a census tract is found to be cropland, a quarter to be pasture, and the rest miscellaneous classes, the whole census tract will be classified as *Cropland + Pasture*.

This map is visualised in figure 3.2 with a legend with English translations of the land use classes. Further details about the map and its variables can be found in Appendix B.2. All classes related to cropland are visualised in Figure 3.3a and all classes related to pastures in Figure 3.3b.

3.2.2 Land cover maps for the purpose of prior creation

In order to construct a prior estimation of what the land use of the region of interest is, a land cover map is needed. This map does not need to represent the true spatial distribution of the land cover, but we wish for it to be a well-informed estimation. The criteria that was set up for what such a land cover map should include were:

- Land cover mapping of the whole world, since deforestation is a phenomenon throughout the world.

- Contain land cover information regarding agricultural statistics.

- Be updated frequently, preferably at least yearly, in order to be up to date with current deforestation drivers.

- Have small classification pixels/regions, in the hope of increased accuracy.

With these criteria in mind, the map provided by the European Space Agency's (ESA) Climate Change Initiative (CCI) - and subsequently the European Union's Copernicus Climate Change Service (C3S) - was chosen.

ESA-CCI produced land cover maps of the world during the years of $1992-2015$ [24]. The European Union's Copernicus Climate Change Service (C3S) has continued this venture and has at the moment land cover maps for the years 2016-2018 which are consistent with the existing land cover maps produced by ESA-CCI [25]. We will use the land cover maps provided by ESA-CCI for the year 2006, since we have validation data for that year.

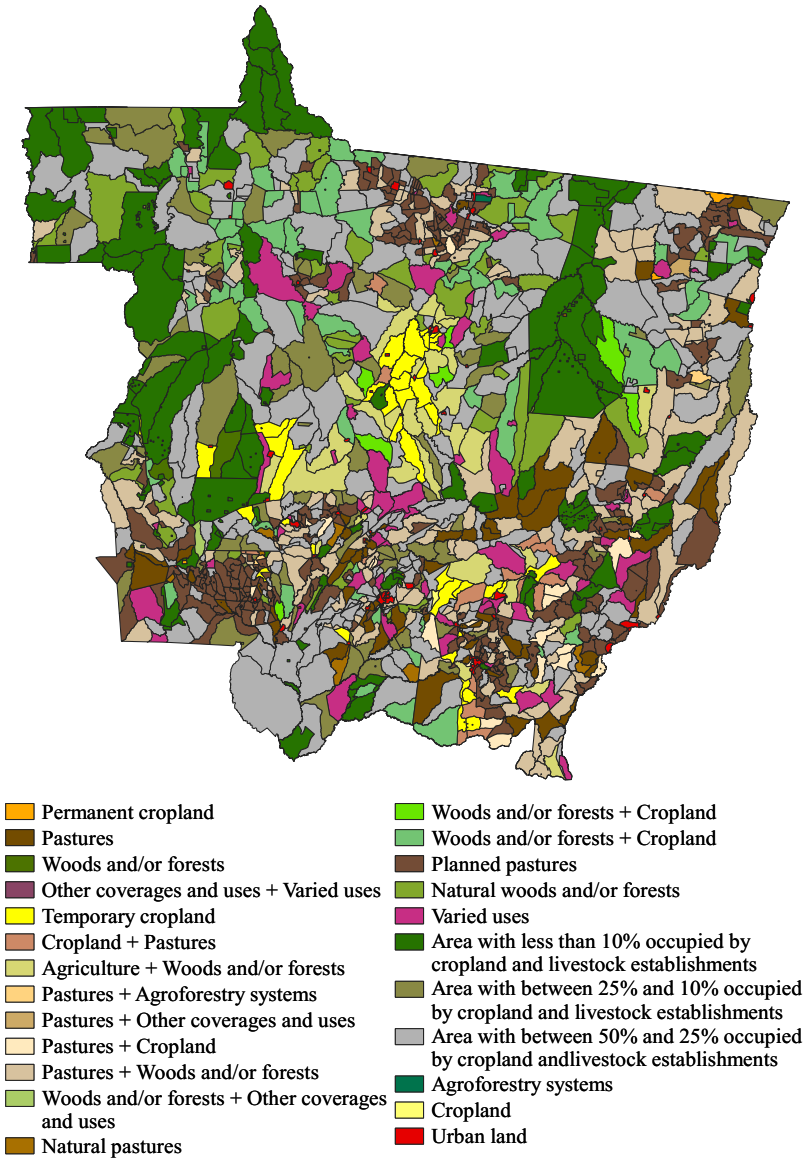
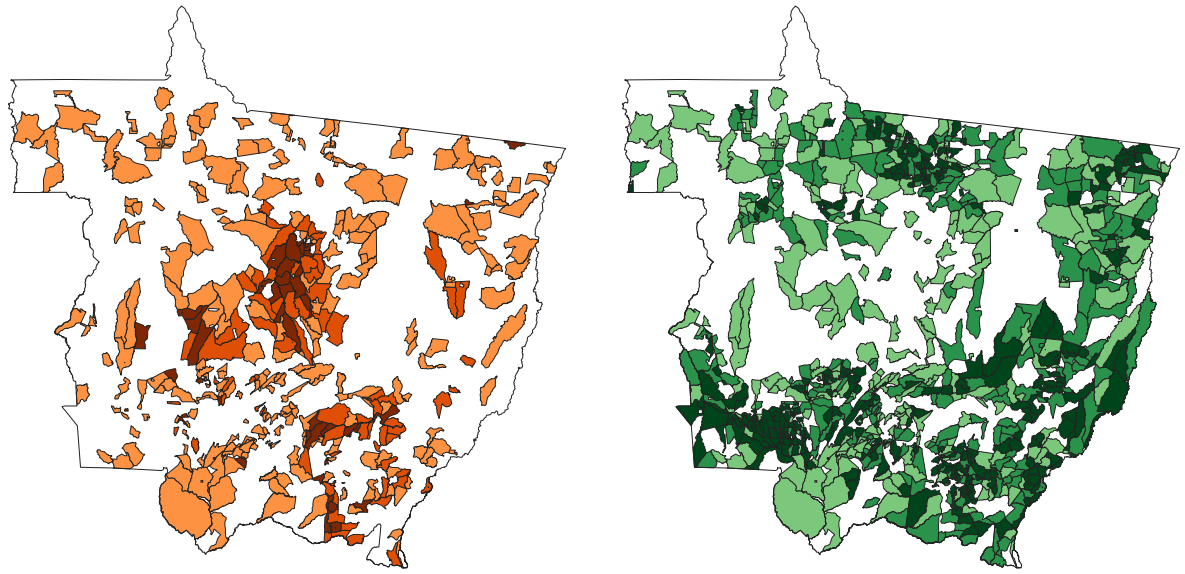


Figure 3.2: *Map over land use spatial distribution according to IBGE's 2006 census. A legend explaining the land use classes and the colours associated to them can be found in Appendix B.2.*

The land cover maps consists of pixelated maps of the world where each pixel has been assigned a class according to data gathered by means of satellite imagery, and a classification process improved by the use of validation data. The land cover maps has a pixel size of 0.002 778°, which corresponds to approximately 300 m at the equator.

The classes of the land cover maps is based on the Land Cover Classification System (LCCS) developed by the United Nations (UN) Food and Agricultural Organization (FAO). The maps are intended to be consistent over the whole world, while at the same time not lose any additional information that is available on a regional level. Thus, the legend classes is divided into two levels: global and regional. Here we use the information given by the global classes; any pixels classified with classes from the regional level have been counted towards their global counterparts.



(a) **(b)**

Figure 3.3: *Map over cropland (a) and pasture (b) spatial distribution according to IBGE's 2006 census. The darkest colours represents areas with homogeneous classes, the middle colour represents mixture classes, and the lightest colours represent areas stated to have less than 50% of agricultural use.*

A legend of all the classes on both the global and regional level can be found in Appendix B.1. Available there are also the classes' colour coding used in figures throughout this thesis. The colours are the same as the ones provided by CCI and C3S, with the exceptions of the classes on a regional level. Since only the global classes are of interest to the thesis, we have substituted the colours provided for them by CCI and C3S by the global counterpart classes.

3.3 Constructing the prior

In order to run the cross entropy minimisation algorithm, found in Equation 3.3, a prior estimation of the land use is needed. In this thesis, this estimation has been constructed from ESA-CCI land cover data, which consists of a pixelated map of Earth where each pixel has been assigned a land use classification based on satellite data; for further explanation of the data, see Section 3.2.2.

The ESA-CCI land cover data is loaded into a GIS (Geographical Information System) application. Due to its large file size, the raster file is cut into a shape where it only has raster information of the region of interest, in order to make it easier to handle. Since the only classes of interest for this work are the ones containing information regarding farm land and pastures, a reclassification of the ESA-CCI data was conducted.

The ESA-CCI class chosen to represent pastures in this work is:

- *Grassland*

The ESA-CCI classes chosen to represent cropland in this work are:

- *Cropland, rainfed.* Included in this class are the regional classes *Cropland, rainfed, herbaceous cover* and *Cropland, rainfed, tree or shrub cover.*
- *Cropland, irrigated or post-flooding.*
- *Mosaic cropland(*$> 50\%$*)/natural vegetation (tree, shrub, herbaceous cover)(*$< 50\%$*).*
- *Mosaic natural vegetation (tree, shrub, herbaceous cover)(*$> 50\%$*)/cropland(*$< 50\%$*).*

These classes are summarised in Table 3.2. The spatial distribution of these classes in the region of interest can be found in Figures 3.4a and 3.4b. All other classes in the ESA-CCI land cover data set were set as the class *Other.*

Table 3.2: *Classes in the ESA-CCI land cover data set used here. The cropland class used in this work consists of the pixels classified by ESA-CCI with the global values of* $\{10, 20, 30, 40\}$ *(The regional values* $\{11, 12\}$ *are counted as the global value 10). The pasture class in this work consists of pixels classified by ESA-CCI with the global value of 130.*

Label		Value		Colour
Global	**Regional**	**Glo.**	**Reg.**	
Cropland, rainfed		10		
	Cropland, rainfed, herbaceous cover		11	
	Cropland, rainfed, tree or shrub cover		12	
Cropland, irrigated or post-flooding		20		
Mosaic cropland ($> 50\%$) / natural vegetation (tree, shrub, herbaceous cover) ($< 50\%$)		30		
Mosaic natural vegetation (tree, shrub, herbaceous cover) ($> 50\%$) / cropland ($< 50\%$)		40		
Grassland		130		

Due to the high resolution of the land cover pixels, a cross entropy optimisation run directly on this data would quickly take too much computer power for investigations of larger regions. Therefore, we wish to aggregate the land cover pixels into larger and more manageable pixels. At the same time, high resolution gives a better description of the spatial distribution, so we also wish to keep the aggregated pixels as small as possible in the process. An equidistant grid was placed over the raster file, as seen in Figure 3.4c; the grid resolution is in the example and in all images of our results set to 20 km. The grid pixels represents the areas for which we want to aggregate data. The grid is cut to only include grid pixels that are included in or intersects the border region.

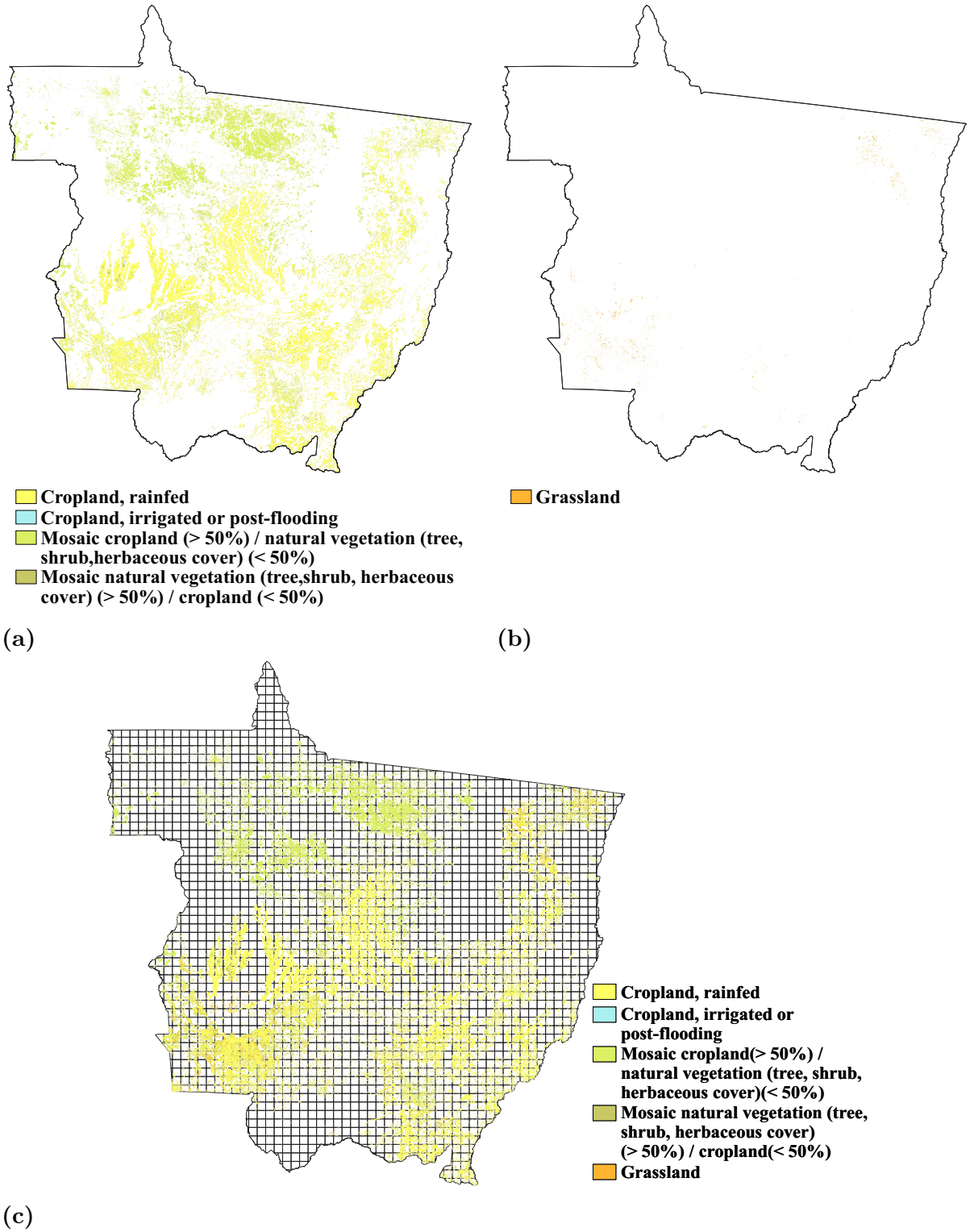


Figure 3.4: *Spatial distribution of the classes used in this work to denote farmland and pasture. The pixels in figure (a) has been clumped together into the farmland class used in this work, and the pixels in figure (b) represents the pasture class in our work. The figure in (c) illustrates the land cover data, superimposed with grid with resolution* 20 km *were each pixel represents the areas for which we wish to aggregate our data. The legends states the name of the class as provided by ESA-CCI. These images illustrate the distribution of land uses in Mato Grosso for the year 2006 in specific.*

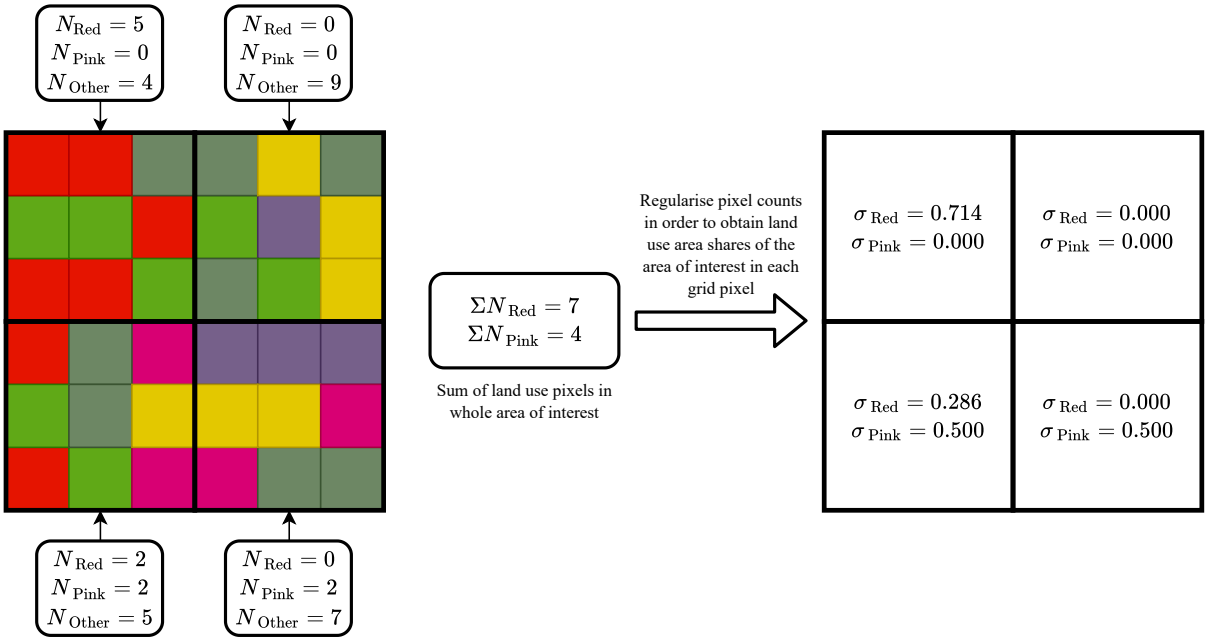


Figure 3.5: *Illustration over the aggregation process, where the coloured pixels illustrate land cover classes and the superimposed larger pixels bordered in black lines represent the grid pixels. In this example, the red and pink classes are the ones of interest. A count of occurrences of classes of interest in each grid pixel is done. The number of each class in the whole area of interest are counted and used as a normalising factor when calculating the class share of the total area of interest in each grid pixel.*

In order to calculate the area share of each land use class j in each grid pixel i, a count is performed of the number of land use pixels of each classification (where classes not of interest are clumped together as *Other*) from the land cover data. We then count the number of occurrences of each class in the whole area of interest and use it as a regularising factor to calculate the class share of the total area of interest in each grid pixel; this process can be seen in Figure 3.5, illustrated with an example.

However, as can be seen in Figure 3.5, we might obtain zero-values for our area shares of interest with this method, since it is possible that no land use pixels of our classes of interest are located in a grid pixel. Since we can see from Equation 3.3a that the prior π_{ij} cannot contain zero-valued elements if the corresponding optimisation variable σ_{ij} does not also equal zero, we need to manipulate the data somewhat in order for the algorithm to be able to give feasible results. Two ways of going about this were identified:

1. Create an initial guess of the area shares σ_{ij} (the optimisation variables) such that all σ_{ij} equals zero where the prior π_{ij} equals zero.

2. Modify the prior π_{ij} to not contain zero-valued elements.

The first listed way implicates that we believe that our prior construction is an informa-

tive prior, meaning that we say it expresses definite information about the distribution. However, since the nature of the ESA-CCI land cover data is such that each pixel is classified to be only one class (when it could be a mixture of many classes and the class assigned to the pixel is merely the major land use in the pixel), and since the classification process is not perfect, this approach is not ideal. A modification of the prior was therefore our chosen method.

There are several ways this modification can be done; one could for example add a constant positive value or a random uniformly distribute positive value to all prior distribution elements. Merely adding would however introduce more total area than what is available in the area of interest, thus some kind of normalisation is needed. In this work we tested two different ways of performing this preprocessing of the prior - *Certainty parameter* and *Confusion matrix* - which will be described in the sections below.

3.3.1 Certainty parameter preprocessing method

To tie in with the inherent uncertainty of the classified data provided by ESA-CCI, this method introduces a certainty parameter γ in order to redistribute data among the land use pixels. The idea is that we assume that, to a certain percentage, we are sure that the area in the land use pixel is actually of the class that it has been assigned. The percentage of the area of which we assume we are not certain of its classification is distributed evenly to all other classes, including the ones that does not concern cropland or pastures.

Similarly to the procedure described in Figure 3.5, we count the number of land use pixels of each class (where classes that does not concern cropland or pasture are clumped together as *Other*) in each grid pixel; we will call it N_{ij}, where i is an index over the grid pixels and j an index over the land uses. Each count N_{ij} is multiplied by the certainty parameter γ in order to obtain the number of pixels we assume have a correct classification,

$$N_{ij,\text{Cert}} = \gamma \cdot N_{ij}, \quad \forall i, j. \tag{3.5}$$

Note that the counts $N_{ij,\text{Cert}}$ are not necessarily integers at this point, and we cannot really say it is a count anymore. However, since the land use pixels represents an area, having real valued counts would translate as counting a part of the pixel areas.

The pixels which we have reason to believe, according to the certainty parameter, to not be correctly classified,

$$N_{ij,\text{NotCert}} = (1 - \gamma) \cdot N_{ij}, \quad \forall i, j, \tag{3.6}$$

are summed according to their initial classification,

$$\Sigma N_{j,\text{NotCert}} = \sum_i N_{ij,\text{NotCert}}. \tag{3.7}$$

We now wish to evenly distribute these pixels to the other land use class counts N_j. We will do this by creating a transition matrix, which has the width and height of the number of classes in our data set and where all elements contain the same number, $\frac{1}{\#\text{classes}-1}$, except the elements on the diagonal, which contains 0. The class represented on each row, will distribute its pixel counts of uncertainly classified pixels $N_{ij,\text{NotCert}}$ to the other classes, represented by each column in that row, according to the factor found

in that corresponding element. No pixels will be distributed from a class to the same class; since we have a zero on the diagonal, where the transition matrix states what factor of pixels should be transferred from one class to itself, we make sure that no pixels we have assumed are misclassified goes back to the class they came from.

The sum of all pixels redistributed to class j from all other classes, as descided from the transition matrix, will be noted as δN_j. Lastly, the new land use class j counts in each grid pixel i, $\tilde{N}_{ij}$, is set to be

$$\tilde{N}_{ij} = N_{ij,\text{Cert}} + \delta N_{\text{j}}. \tag{3.8}$$

For our specific problem setup, the count of class 130 - *Grassland* - is said to be the number of pixels classified as *Pasture* in the grid pixel, and the rest of the classes that is not classes $\{10, 20, 30, 40, 130\}$ is said to be the number of pixels classified as *Other*. We however need to make some adjustments for the classes concerning cropland, due to the fact that two of the cropland classes concerns land use that are mixed between cropland and other land uses, and we are only interested in the cropland. Thus:

- For class 30, *Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)*, we keep 75% of the pixels of this class, and say that the rest are *Other*.

- For class 40, *Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)*, we keep 25% of the pixels of this class, and say that the rest are *Other*.

All counts of the classes 10 (*Cropland, rainfed*), 20 (*Cropland, irrigated or post-flooding*), and the pixel counts left after compensating for the mixed use in classes 30 and 40 are summed together and said to be *Cropland*.

The area shares σ_{ij} are then calculated with the new classification counts $\tilde{N}_{ij}$ as described in Figure 3.5.

In this work, we decided to include results for three values of the certainty parameter, $\gamma = \{0.714, 0.850, 0.995\}$. In the documentation of the data provided by ESA-CCI, they state the overall accuracy of their classifications as 71.4% if validation deemed as homogeneous (consisting of one land use class) are taken into account [24]. If one takes into account validation points that are both homogeneous and heterogeneous (consisting of several land use classes) cells, the overall accuracy is calculated as 75.4%[24]. Runs of the algorithm for both of these values set for γ were made, but it made no big differences between the results. Due to space and time considerations, a decision was made to only include results for $\gamma = 0.714$. We also wanted to include a high certainty parameter value in order to show the spectrum of what the algorithm can do, and a value in between; thus, the inclusion of $\gamma = 0.850$ and $\gamma = 0.995$.

3.3.2 Confusion matrix preprocessing method

In this method, we use the information provided by ESA-CCI about their land cover data and its weaknesses in order to preprocess our prior information. This information is given in the form of a confusion matrix; for further explanation of this concept see Section 2.3.

The confusion matrix is found in the Product Quality Assessment Report documentation when downloading the data [26, p.18], but can also be found enclosed in Appendix D of this thesis. In the calculations below, we will denote the confusion matrix with Γ.

We begin by creating a transition matrix Π from the data given in the confusion matrix Γ. For each row i (representing a classification made by ESA) in the confusion matrix, we divide each element j by the sum of all elements in that row:

$$\Pi_{ij} = \frac{\Gamma_{ij}}{\sum_i \Gamma_{ij}}, \quad \forall i. \tag{3.9}$$

In the example provided in Section 2.3, this would mean that for the *Lion* row, we would take each of the elements $\{23, 4, 3, 2, 0\}$ and divide by 32. This generates a matrix which visualises the percentage of pixels classified as that rows class (*Lion* in this example) which should really be classified as any of the classes, including a percentage of correct classifications.

The ESA-CCI land cover data is loaded in, and all pixels classified as local classes are counted to be their parental global class instead, due to the fact that the confusion matrix only concerns the global classes. For each grid pixel i in the area of investigation, we then adjust the land use pixel counts N_{ij} by, for each class j, multiplying the grid pixel value for that attribute with the transition matrix row for that class. After we have gone through all classes, the results generated by this procedure is summed together and taken as our new grid pixel value:

$$\tilde{N}_{ij} = \sum_j N_{ij}\Pi_{ij}, \quad \forall i \tag{3.10}$$

The procedure described above were almost entirely effective of removing zeros from the cropland and pasture classes, but unfortunately there were a handful that were zero in either one or both of the classes. The smallest element of each of these classes was identified and its value divided by 4; this value was then added as a small noise to all pixel counts in the respective classes.

We adjust the pixel counts for the mixed cropland classes in the same fashion as described for the certainty parameter method. The calculation of area shares σ_{ij} are then done in as illustrated in Figure 3.5.

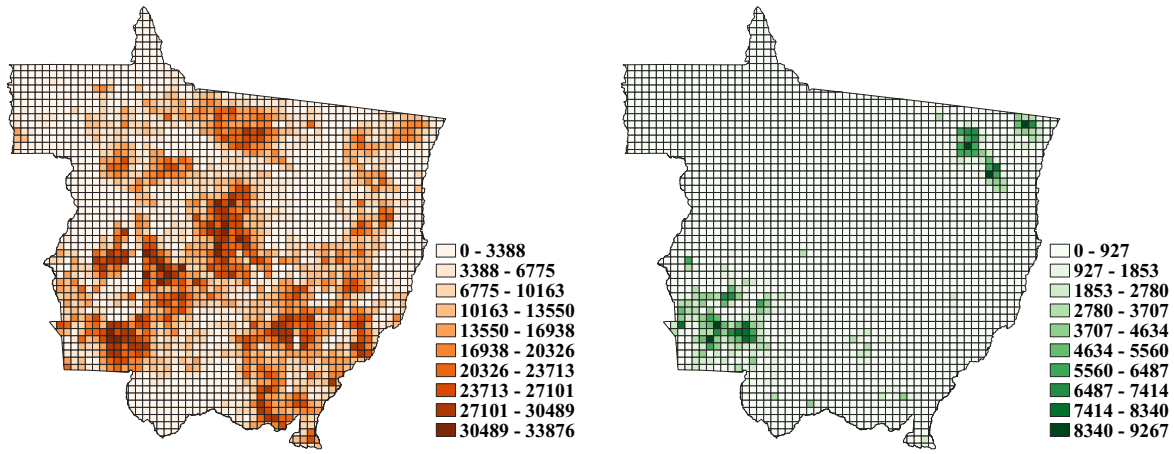
4

Results

4.1 Priors

In this section, we will illustrate the priors obtained from the different prior preprocessing methods used in this thesis. All figures will be visualised with equal interval classification with 10 classes, where the minimum and maximum value of the classes are set to the smallest and largest values found in our priors and results. The pixel size of the images are 20 km×20 km, and all legend entries show land use area measured in hectares.

4.1.1 No prior preprocessing



(a) **(b)**

Figure 4.1: *Prior estimation of land use area spatial distribution in Mato Grosso for the year 2006 obtained before using any prior preprocessing method (Cropland (a), Pasture (b)). Visualisation are done with equal interval classification.*

Illustrated in Figure 4.1 are the priors generated from the ESA-CCI spatial distribution data set (Cropland (a), Pasture (b)) before any preprocessing of the prior has been made, meaning that there are zero-valued pixels in the images. Note that the images in this figure does not share the same class divisions as the the images (priors, result from the cross entropy minimisation algorithm) in the rest of this report. This is due to the fact that the values in these priors are on different scales of values, and the shapes of the spatial distribution would be difficult to distinguish otherwise.

We see the general shape that the raw data from the ESA-CCI data says the cropland and pasture areas in Mato Grosso should be located. Note that the unprocessed prior says that the pasture area, in Figure 4.1b, are located in just a few pixels. The pasture prior does not share similar shapes as the pasture census map in Figure 3.3b except for the small points where the census map seems to have a high concentration of pastures. The sum of the areas assigned as pastures does not seem reasonable when keeping in mind that pasture land use is one of the largest land uses in Mato Grosso. The cropland prior captures the general clusters of the cropland census map, see Figure 3.3a. Since a grid pixel has an area of 40 000 ha, the cropland prior tells us that the pixels with the darkest colour in this map almost exclusively consists of cropland.

4.1.2 Certainty parameter preprocessing

Delineated in Figure 4.2 are the priors generated from the ESA-CCI spatial distribution data set using the certainty parameter preprocessing method with $\gamma = \{0.714, 0.850, 0.995\}$ (cropland (a), (c) and (e), pasture (b), (d) and (f)).

First, we observe that the both the cropland and pasture priors for $\gamma = 0.714$ and $\gamma = 0.850$ have a seemingly uniform distribution when using the certainty parameter method. It is not until, in our results, when gamma is as high as $\gamma = 0.995$ that we can see clustering in the cropland prior (Figure 4.2e) similar to the cropland census map in Figure 3.3a. We only see a very small cluster in the lower left corner in the pasture prior (Figure 4.2f) that barely distinguish it self from the surrounding pixels in the same region where the pasture census map (Figure 3.3b) seem to have a high concentration of pasture area.

4.1.3 Confusion matrix preprocessing

Shown in Figure 4.3 are the priors priors generated from the ESA-CCI spatial distribution data set using the confusion matrix preprocessing method (cropland (a), pasture (b)).

In this prior, we can see clustering of where the cropland and pasture area is located in the area of investigation. We can see similarities between the large cluster in the middle of the cropland census map, as seen in Figure 3.3a, in our cropland prior, as well as similarities between the cluster in the lower right region of the census map and the cropland prior. In the pasture prior, we can see the general shape that runs along the lower half and right side of the pasture census map, as seen in Figure 3.3b, in our pasture prior.

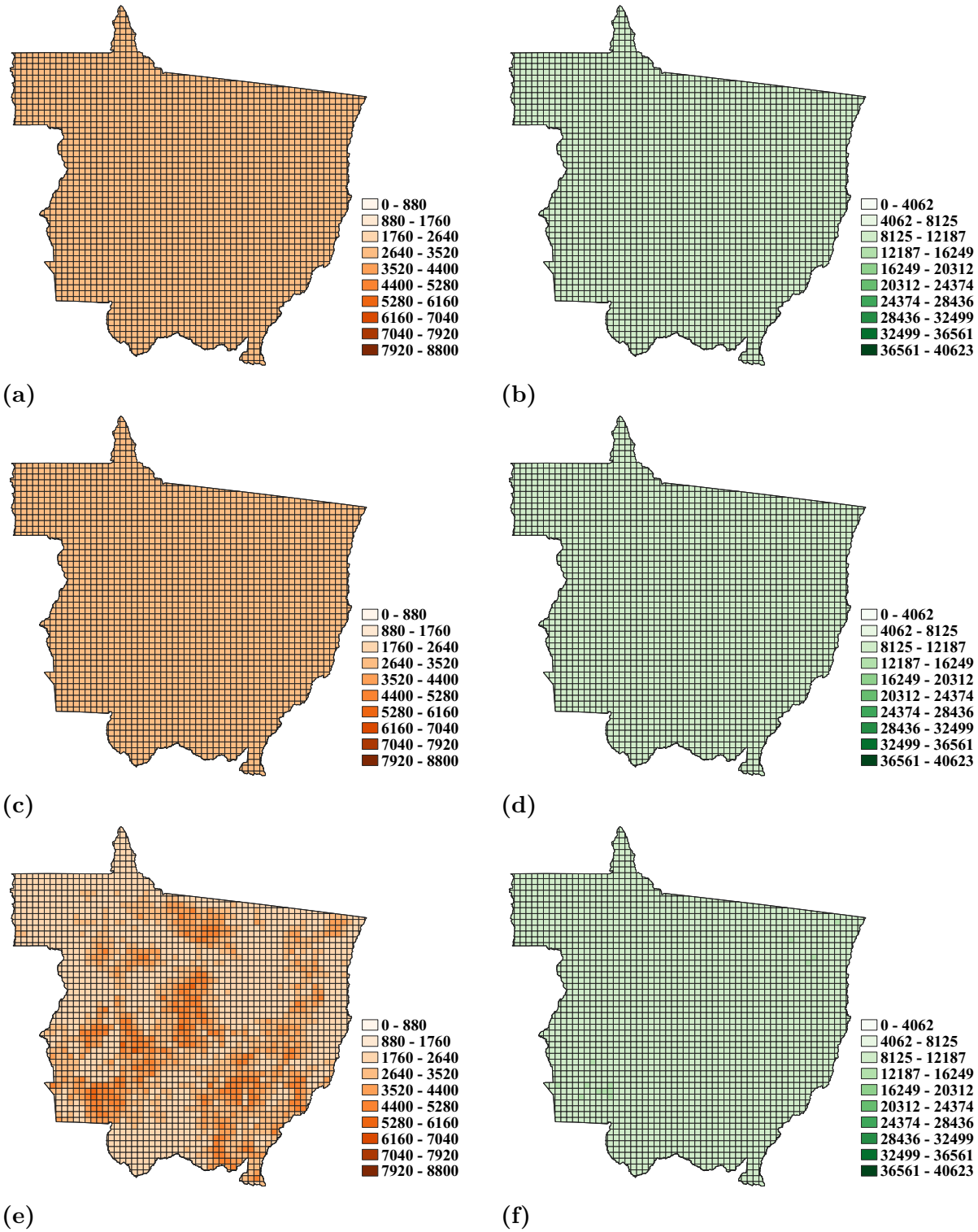
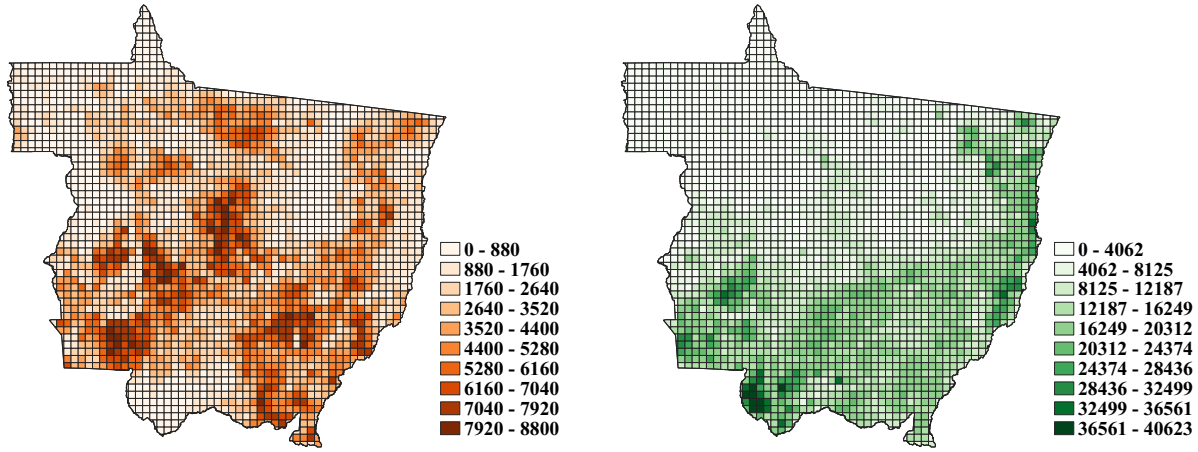


Figure 4.2: *Prior estimation of land use area spatial distribution in Mato Grosso for the year 2006 obtained using the certainty parameter preprocessing method with* $\gamma = 0.714$ *(Cropland (a), Pasture (b)),* $\gamma = 0.850$ *(Cropland (c), Pasture (d)) and* $\gamma = 0.995$ *(Cropland (e), Pasture (f)). Visualisation is done with equal interval classification.*



(a) **(b)**

Figure 4.3: *Prior estimation of land use area spatial distribution in Mato Grosso for the year 2006 obtained using the confusion matrix preprocessing method (Cropland (a), Pasture (b)). Visualisation are done with equal interval classification.*

4.2 Certainty parameter

Below, the results from the cross entropy minimisation algorithm run with the certainty parameter prior preprocessing method for $\gamma = \{0.714, 0.850, 0.995\}$, as well as histograms over the value distribution for the land uses, are presented. All results are generated by running the cross entropy minimisation algorithm 20 times for randomly generated initial conditions for σ_0 to increase the likelihood for us to find the global minimum while not using a global search method. All legend entries are displayed in hectares.

In Figure 4.4 we can see the result of cross entropy minimisation using the certainty parameter preprocessing method. We observe that for $\gamma = 0.714$ and $\gamma = 0.850$ we have a seemingly uniform distribution. It is not until, in our results, when γ is as high as $\gamma = 0.995$ that we can see clustering in the cropland prior (Figure 4.2e) similar to the cropland census map in Figure 3.3a. We only see a very small cluster in the lower left corner in the pasture prior (Figure 4.2f) that barely distinguish itself from the surrounding pixels in the same region where the pasture census map (Figure 3.3b) seem to have a high concentration of pasture area.

In Figure 4.5, we can see histograms over the respective land use area distributions in each cross entropy optimisation grid pixel in Mato Grosso for the year 2006 using the certainty parameter preprocessing method with $\gamma = \{0.714,\ 0.850,\ 0.995\}$ (cropland (a), (c) and (e), pasture (b), (d) and (f)). The data was divided into 100 bins, where the horizontal axis represents the amount of hectares the land use in question occupies in each grid pixel and the vertical axis the amount of pixels that fall within each bin.

We note from the histograms that the variance of the value distribution is not big for $\gamma = 0.714$ and $\gamma = 0.850$ for either land use; almost all pixels have the same value. For $\gamma = 0.995$ the cropland distribution seem to have more variance, but the pasture distribution is still in a spike along one value.

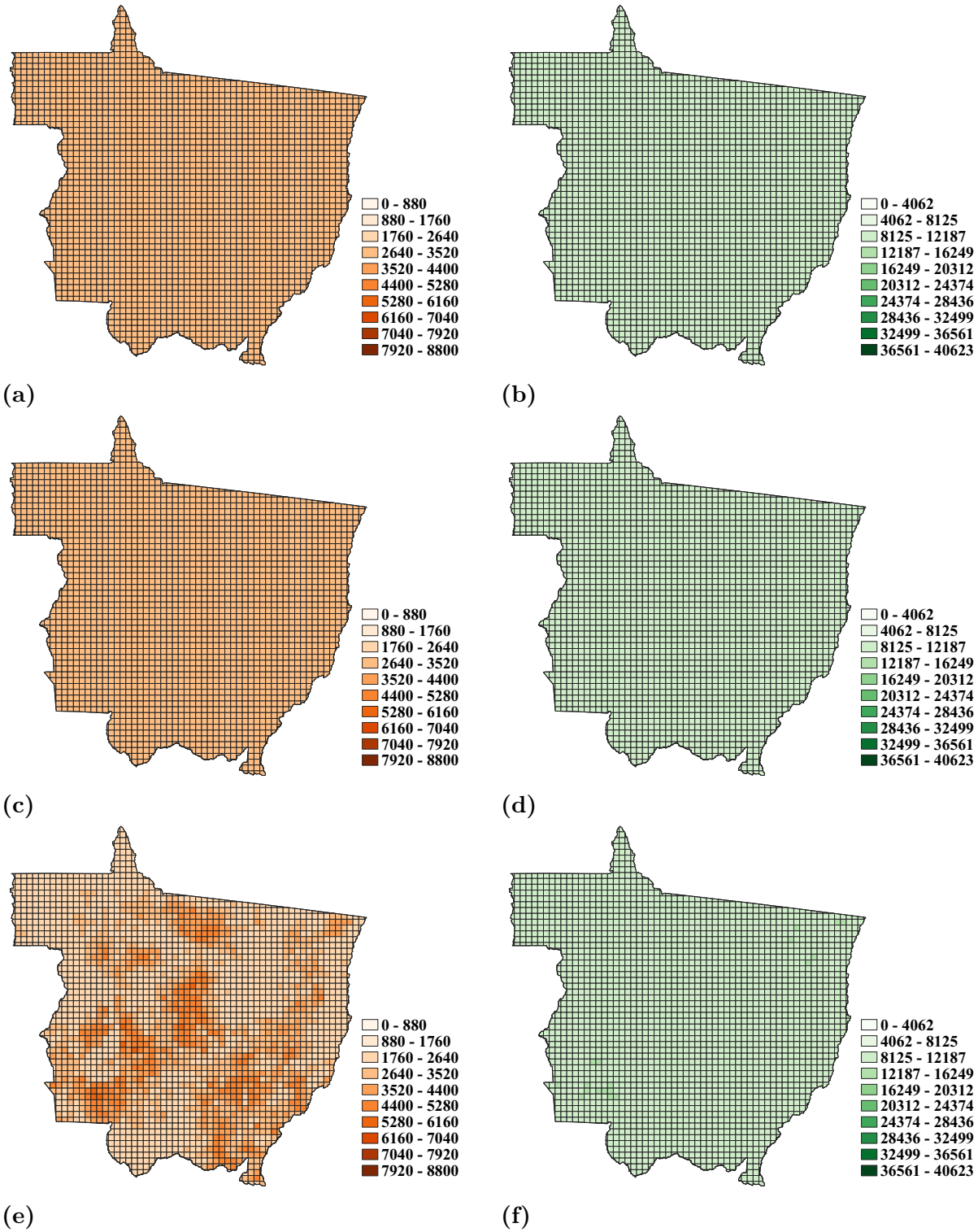


Figure 4.4: *Land use area distribution obtained using the certainty parameter preprocessing method of the prior with* $\gamma = \{0.714,\ 0.850,\ 0.995\}$ *(cropland (a), (c) and (e), pasture (b), (d) and (f)). Visualisation are done with quantile classification. All legend entries are displayed in hectares.*

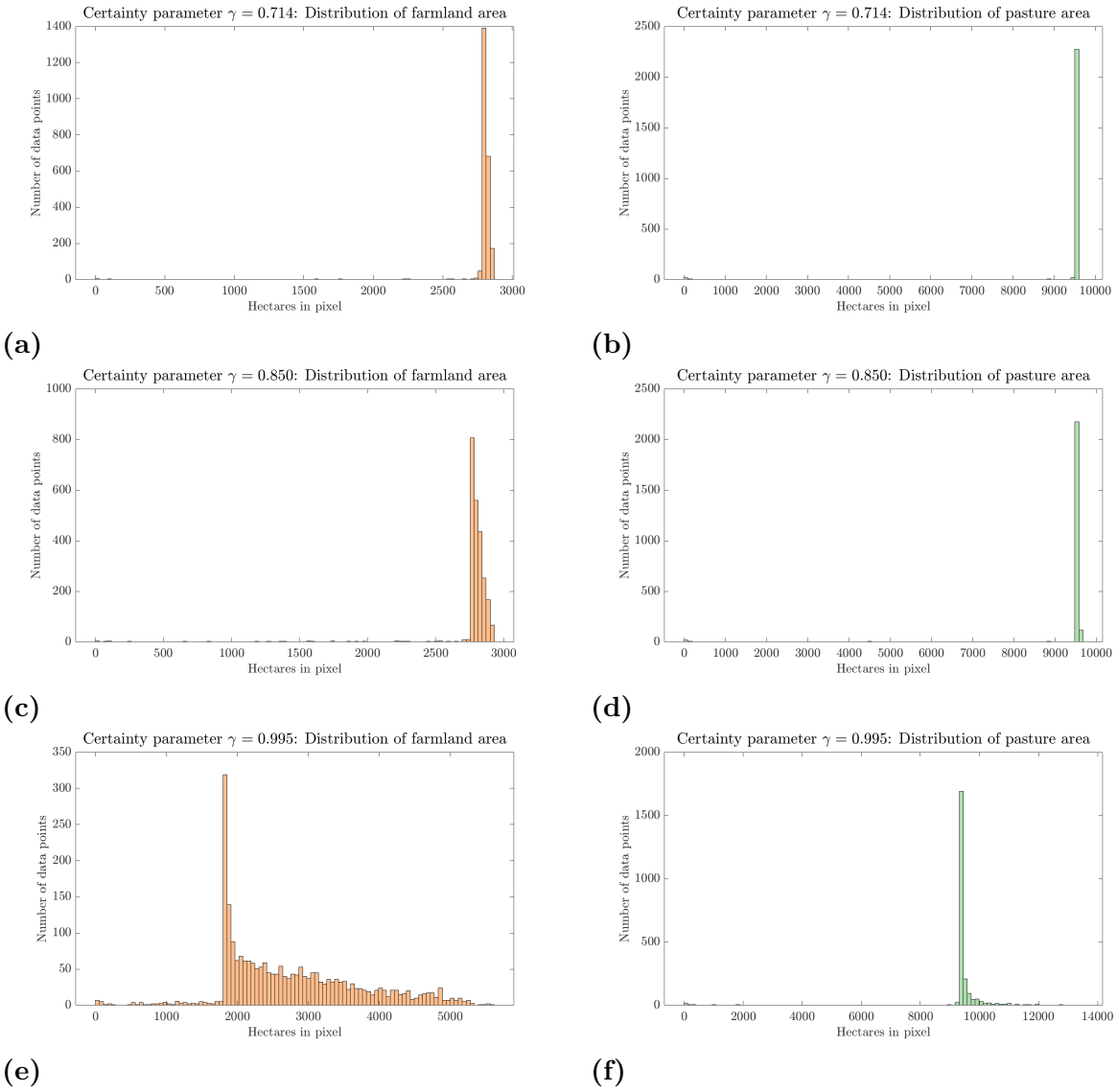


Figure 4.5: *Histograms over the respective land use areas in each cross entropy optimisation grid pixel on Mato Grosso for the year 2006 using the certainty parameter preprocessing method with* $\gamma = \{0.714,\ 0.850,\ 0.995\}$ *(cropland (a), (c) and (e), pasture (b), (d) and (f)). The data was divided into 100 bins, where the horizontal axis represents the amount of hectares the land use in question occupies in each grid pixel and the vertical axis the amount of pixels that fall within each bin.*

Since only $\gamma = 0.995$ shows any kind of clustering, we will only compare those results with the validation data. When doing so, as in Figure 4.6, we see that some of the shapes of the cropland distribution are caught in our results, such as the cluster in the middle of the region and in the lower right corner of the region. The cluster in the lower left corner is however not found, and we have a cluster in the top middle of our results which does not exist in the census map. For the pasture results, we see almost no clustering in the results; the clusters in the lower left corner and upper right corner are somewhat visible, but areas which the census map states should contain more pasture area than others are classified to contain about the same amount of pasture area in our results.

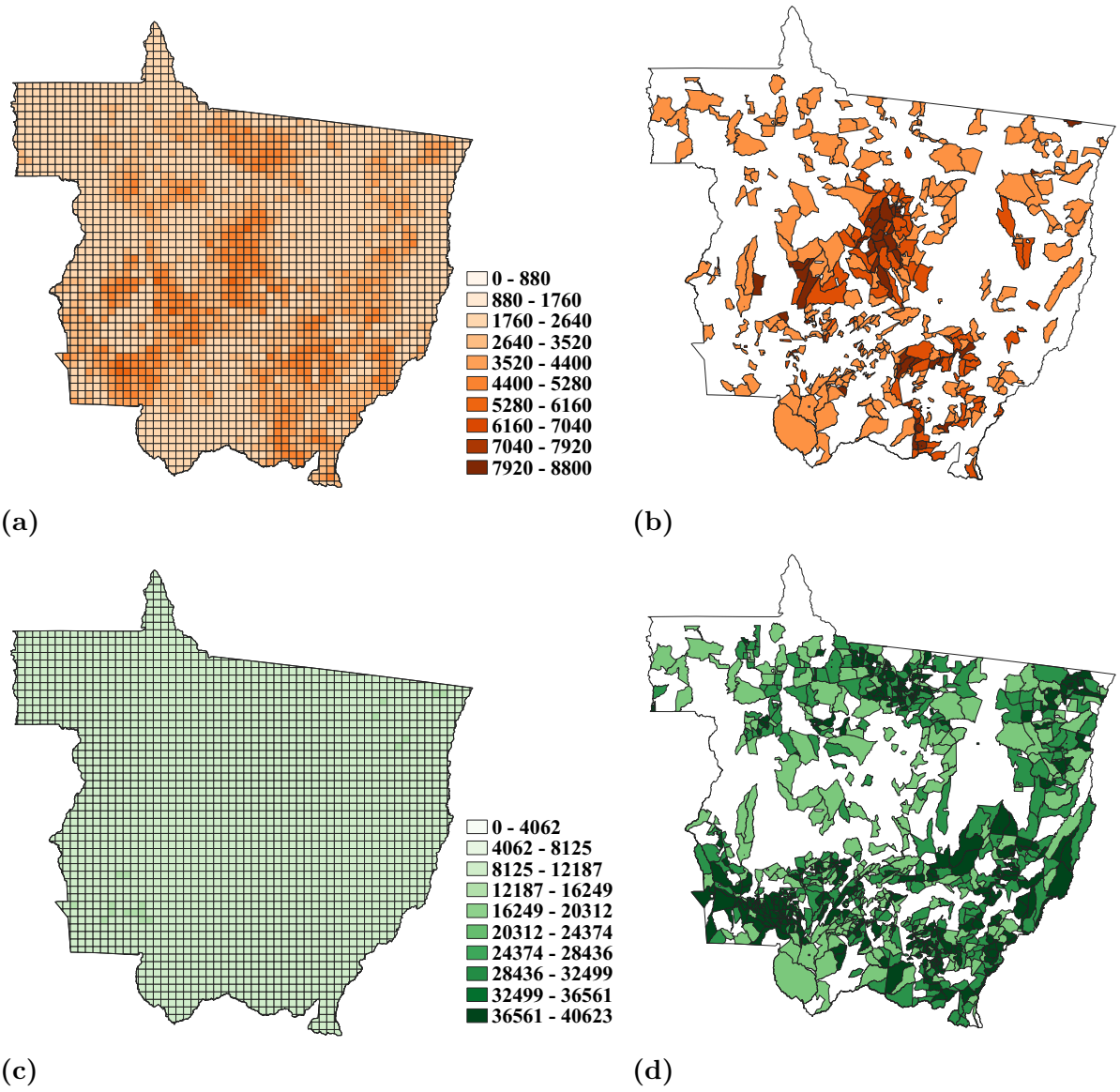


Figure 4.6: *Land use area spatial distribution results using certainty parameter preprocessing method with $\gamma = 0.995$ (cropland (a), pasture (c)), visualised alongside the true land use spatial distribution (cropland (b), pasture (d)).*

4.3 Confusion matrix preprocessing method

Below, the results from the cross entropy minimisation algorithm run with the confusion matrix prior preprocessing method, as well as histograms over the resulting value distribution for the land uses are presented. All results are generated by running the cross entropy minimisation algorithm 10 times for randomly generated initial conditions σ_0 order to increase the likelihood for us to find the global minimum while not using a global search method. All legend entries are shown in hectares. The pixel size of the images are $20\,\text{km} \times 20\,\text{km}$.

In Figure 4.7 we see a comparison between the results obtained with the cross entropy algorithm using the confusion matrix prior preprocessing method and the true image given by the census map. We see that for the cropland, the cluster in the middle of the

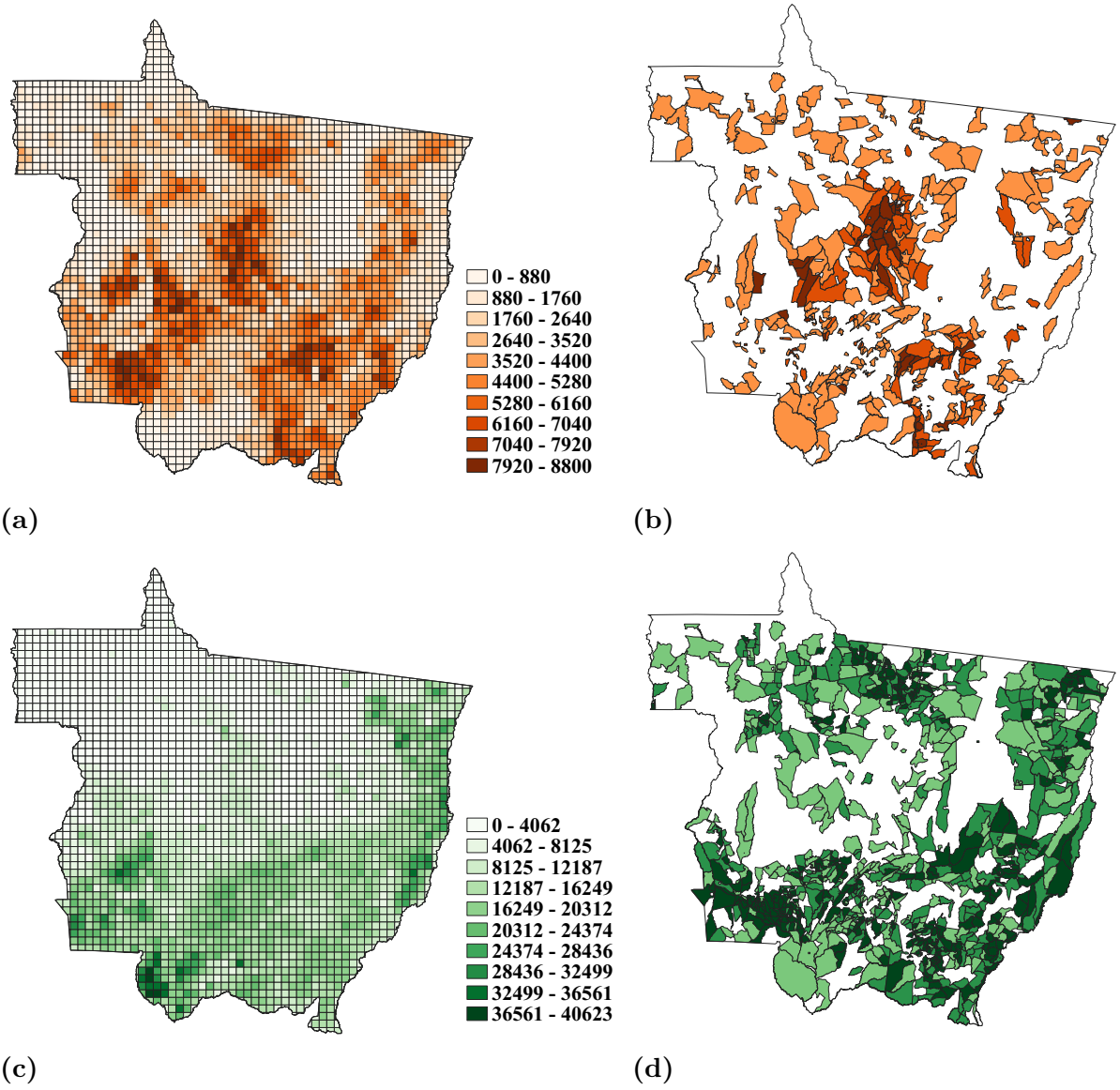
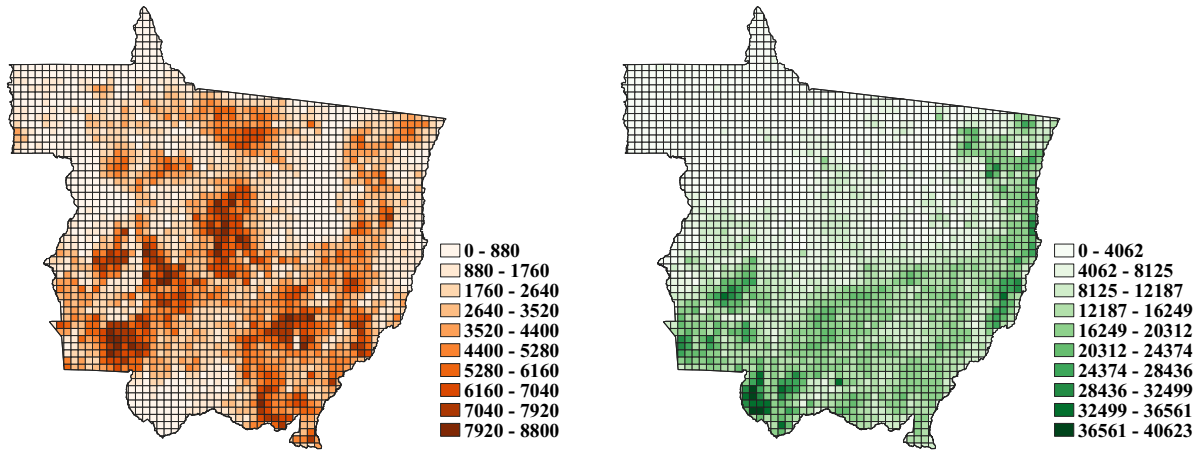


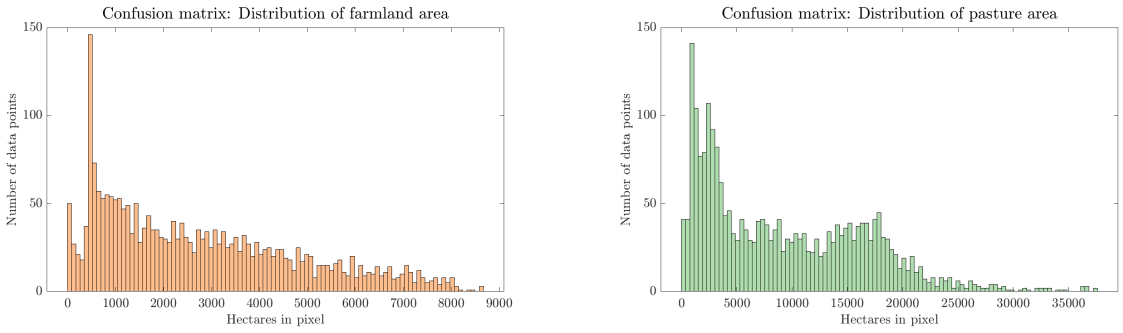
Figure 4.7: *Land use area spatial distribution results using confusion matrix preprocessing method (cropland (a), pasture (c)), visualised alongside the true land use spatial distribution (cropland (b), pasture (d)).*

region in the true image (Figure 4.7b) are present in our results, as well as the clusters in the lower right corner of the region. We do however still have a larger cluster in our cropland results in the top middle and the lower left corner which is not present in the true image. The pasture spatial distribution in our results seems to capture more of the general shape of the true image using the confusion matrix method compared to the certainty parameter method. The larger cluster in the bottom half and up along the right side of the region in the true image in Figure 4.7d is present in the pasture results obtained as seen in Figure 4.7c. The cluster in the top middle in the true image is however not seen in the pasture results.



(a) **(b)**

Figure 4.8: *Land use area spatial distribution in Mato Grosso for the year 2006 obtained using the confusion matrix preprocessing method of the prior (cropland (a), pasture (b)). Visualisation are done with quantile classification. All legend entries are displayed in hectares.*



(a) **(b)**

Figure 4.9: *Histograms over the respective land use areas in each cross entropy optimisation grid pixel on Mato Grosso for the year 2006 using the confusion matrix preprocessing method (cropland (a), pasture (b)). The data was divided into 100 bins, where the horizontal axis represents the amount of hectares the land use in question occupies in each grid pixel and the vertical axis the amount of pixels that fall within each bin.*

5

Discussion and conclusion

5.1 Prior preprocessing methods and cross entropy minimisation

In this section we will evaluate and discuss the performance of the two prior preprocessing methods used in this thesis, as well as the performance of the cross entropy minimisation algorithm.

When looking at the certainty parameter results visualised in Figure 4.4, one might assume that the cropland and pasture spatial distribution for $\gamma = 0.714$ and $\gamma = 0.850$ are completely uniform; with an equal interval visualisation, the whole map is found to be of the same class. It is true, as we have seen in the histograms of Figure 4.5, that there is a small variance of the distribution values, but the cross entropy minimisation algorithm does actually create non-uniform results. Looking at the images visualised with a quantile classification method (as we have done in Appendix C), meaning all classes have the same number of data elements, in Figure C.2, we see that the general shapes of, at least, the cropland validation map are there, in small variations around the same value but still there. For the pasture area distribution however, when looking at the data with a quantile visualisation, the result seem to be very random for $\gamma = 0.714$ (Figure C.2b), with high values of pasture area dotted around the map lacking a clustering shape one might have expected. As γ increases however (Figures C.2d and C.2f) this randomness in the spatial distribution lessens and we see more of a clustering shape, which we can begin to glimpse in our pasture results for $\gamma = 0.995$, in Figure 4.4f.

Yet, as mentioned, the results generated with the certainty parameter prior preprocessing method are almost completely uniform around the same value. The pixels that deviates from this value are the pixels along the border of the area of investigation, and since they have a smaller area by nature than the pixels in the middle of the investigation area it is sensible that they are classified into the lower classes. Considering that we do not expect an almost uniform distribution of the land use area in the area of investigation, it stands to reason that the certainty parameter prior preprocessing method does not do a very good job in deaggregating the agricultural statistics.

Looking at the histograms in figure 4.5, we see more clearly the lack of variance in the results obtained using the certainty parameter method, the data being distributed as enormous spikes for lower values of the certainty parameter. For $\gamma = 0.995$, the pasture

area distribution (Figure 4.5f) are still distributed as a spike, but the farmland area distribution (Figure 4.5e) have a high enough variance to be divided into meaningful classes. However, it seems not plausible that almost all pixels should have 3000 hectares of farmland or more, even in pixels where we can see in the census map should contain primarily rain forest.

Our results are very similar to the certainty parameter priors (Figure 4.2), which leads us to believe that it is not the fault of the cross entropy minimisation that the results are almost uniform - the cross entropy minimisation's algorithms job is to fit the prior and the result and make them as similar as possible, which it has succeeded in doing. The reason for the uniformity are more likely to be due to the fact that the certainty parameter prior preprocessing method does not take into account what classes the ESA-CCI farmland- and pasture classes are most likely to be confused with and misclassified as, and vice versa. Instead, this method says that the ESA-CCI farmland- and pasture classes are equally likely to be misclassified as all other ESA-CCI classes. Looking at the comparison between our results and the true image, we observe that the cluster in the top middle of our farmland results in Figure 4.6a seems to correspond to the cluster in the top middle of the pasture true image in Figure 4.6d. This might indicate that there is some difficulty in distinguishing cropland from pastures in the satellite data, something which has been identified as struggle in remote sensing in general [27]. Thus, assuming a uniform likelihood of misclassifying classes in the ESA-CCI data set does not seem to be a good simplification for this specific problem.

Moving on to the confusion matric prior preprocessing method, we begin to note that the results obtained are more similar to the census data for this method than the results obtained from the certainty parameter method. Areas with large cropland clusterings in the census data are captured in our results, and the general shape of the pasture census data with its cluster along the bottom half and right side of the region is there. However, clusters which appear in the cropland results (such as along the top middle) cannot be found in the census data, and clusters in the pasture census data (such as along the top middle) cannot be found in the pasture results.

These differences in the clustering of cropland and pasture area in our confusion matrix method results from the census data could be due to the aggregation process of the census map - each census tract only show the dominating land use in it. There could be cropland/pasture in the region where the census map does not show it, if there is another land use that has a larger land cover percentage than cropland/pasture. It can also be due to a lack of accuracy of the confusion matrix generated by ESA-CCI. The values of the confusion matrix depends on the validation points considered when creating the confusion matrix. If not enough validation points are chosen that either are classified by ESA-CCI or truly is one of the classes, we might not have enough information to rescale the spatial data to our satisfaction.

Looking at the prior distribution obtained using the confusion matrix method in Figure 4.3, we see that they seem very close to the results obtained after running the cross entropy minimisation. The cluster in the top middle of the true image of the pasture distribution is not found in the prior either, which leads us to understand why the cross entropy minimisation results does not include it either. We see that the top middle

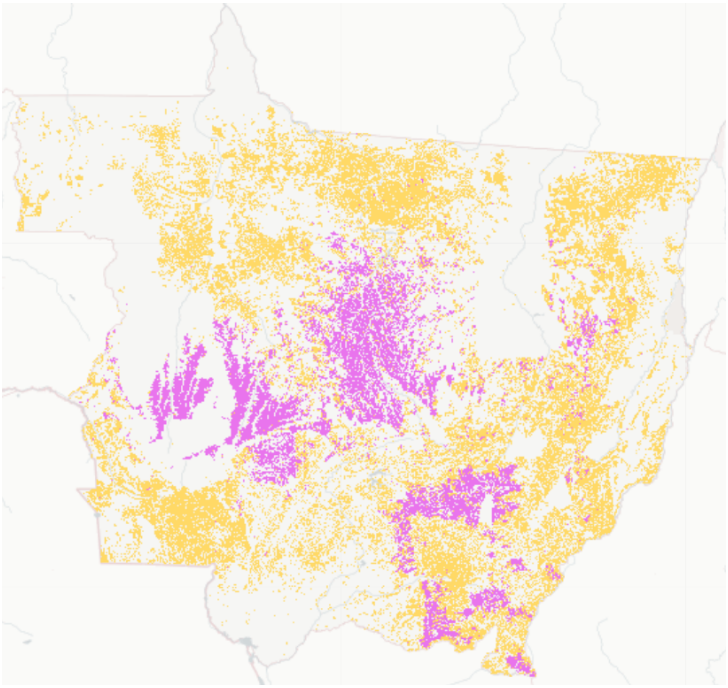


Figure 5.1: *Spatial distribution of cropland (yellow) and pasture (pink) land use in Mato Grosso for the year 2006 according to the MapBiomas classification algorithm. Image taken from online map viewing tool on MapBiomas webpage [28].*

cluster in the cropland results in Figure 4.7a are located about the same place where the pasture top middle cluster in the true image in Figure 4.7d, which might indicate that even though we took the confusion matrix into consideration, there are still some difficulty in distinguishing farmland from cropland.

During the work with this thesis, we tested a third method to handle the problem with zero-values in the prior; creating a confusion matrix between the satellite derived classification spatial data set from MapBiomas [21]. They offer a 30 m×30 m resolution and updates its maps over Brazil's land usage yearly. Since one of our criteria were to have a global land use map coverage as our prior, this data set was not chosen to become our prior, but since the product caught the general shape of the true distribution of cropland and pastures in Mato Grosso (the spatial distribution of the MapBiomas product can be seen in Figure 5.1), our idea were to rescale the ESA-CCI spatial dataset with the help of the MapBiomas spatial data set. Due to the difference in pixel size, the MapBiomas pixels were aggregated up into the same pixel size as ESA-CCI (300 m). A confusion matrix between the cropland and pasture classes were made, and a rescaling similar to the one described in 3.3.2 were performed. This did however not have a satisfying result; the prior of the cropland distribution and the prior of the pasture distribution became very similar to one another - making no real distinction between cropland and pasture. This method was therefore dropped from the thesis.

In conclusion, the performance of the cross entropy minimisation algorithm is dependent on the quality of the prior used with it. We have observed that the certainty parameter prior preprocessing method is not a great method to use in order to reach a believeable

prior. It is not until one assumes the satellite data is very accurately classified, 99.5% accuracy or even higher, where we see results having a similar shape to the census data and somewhat believable pixel area values. However, if no data about the most likely classes for each class to be confused with exists, it might be an adequate approximation to make in order to avoid having a zero-valued prior. The confusion matrix prior preprocessing method generates a much more believable distribution of pixel values, as well as aligning relatively well with the shape of the census data. It does however not eliminate the difficulty in distinguishing cropland from pasture entirely, which might be due to the lack of validation points in generating the confusion matrix.

Generally, it seems like the cross entropy minimisation algorithm performs well. Its purpose is to fit the optimisation variables as closely as possible to the respective prior values - limiting the cross-entropy - while also taking certain conditions into account. Our model is relatively simple in the sense that it takes regularisation of the optimisation variables for each land use, and the total land use area in each grid pixel into account. In the case of more precise statistics, such as suitability of area for crops or area yield statistics for different production methods, more conditions could be imposed on the cross entropy minimisation without loss of generality.

5.2 The optimisation process

During our optimisation process, we use the deterministic interior point method in order to find minimum of our optimising function. We can however not be sure to find the global optimum in this way since, depending on the initial values of the optimisation variables σ_0, we can get stuck in a local minimum and thus reach different results with each run of the algorithm. There are several ways in which this could be handled. In our investigations of a working method, we tried among other things implementing two stochastic optimisation algorithms: Particle Swarm Optimisation method and Genetic Algorithm.

A good in depth explanation of the particle swarm optimisation method can be found in chapter 5 of *Biologically inspired optimization methods: an introduction*[29], but explained in simple terms it keeps a set of candidate solutions (a *swarm* of *particles*) which move around in the space of feasible points for the optimisation problem. The particles movements are guided by their own best known solution as well as the swarm's best known position according to a given formula. This process is continued until either the termination condition has been met, or the maximum number of iterations has been conducted. This method does not guarantee finding the global optimum, but due to its stochastic nature, and that the set of candidate solutions affect each other in the determining of the candidates movements aids to reduce the risk of getting caught in a local optimum. However, when testing this method for several different parameter values, the method was both slow to compute and had a hard time finding the optimum despite allowing a large number of function evaluations. This might be due to the large number of optimisation variables, the number of constraints, or the nature of the constraints. Piotrowski, Napiorkowski, and Piotrowska [30] has made investigations in the number of particles needed for different dimensional problems and found that 70-100 particles work just as well for 10-dimensional as 100-dimensional problems; our problem is however 4806-dimensional

for our 20 km grid (2403 gird pixels and 2 land uses), so it is possible that it contributed to the difficulty. The particle swarm optimisation algorithm also works best for inequality constraints, and our problem has an equality constraint for each land use introduced. It is possible to convert an equality constraint into an inequality constraint by introducing a small tolerance term [31]. This was however tested without improvement of the optimisation result. It is nevertheless possible that the right parameter selection was not found in our investigations for example, or that the method - given more time - would reach the solution. It was for the purposes of this thesis however deemed to be too time consuming for the number of optimisations that we wanted to conducted, so the particle swarm optimisation method was not chosen in the end.

The genetic algorithm is explained in depth in chapter 3 of *Biologically inspired optimization methods: an introduction*[29]. Briefly, it is a method inspired by evolution and natural selection, in which a set of candidate solutions (*population* of *individuals*) are evaluated, and the individuals with the lowest optimisation value (the best *fitness*) are more likely to be randomly selected to mate and create new, similar individuals in the next iteration (*generation*). All new offspring also has a chance of randomly mutating, which brings diversity to the candidate solutions search space. This process is continued until the termination condition (for example a fixed number of iterations has been reached, or no improvement in the highest ranking fitness value over a number of generations has been made) has been reached. The genetic algorithm does not guarantee finding the optimum either, since it does not know how to prioritise long-term fitness over short-term fitness. It is possible to increase the chance of finding the global optimum by increasing the mutation rate or by introducing a parent-selection technique that aims to diversify the population, but there is no guarantee this will work either. Another problem is that fitness function evaluations van be very costly and time consuming, especially for high-dimensional problems. The optimal population size for the generic algorithm varies from problem to problem, but it has been observed that the larger the population, the more accurate the algorithm becomes. But, an increase in population size also increases the number of generations needed in order to converge - likely due to the overall increase in probability that a mutation will occur [32]. When we experimented with the genetic algorithm for the cross entropy minimisation algorithm, we observed for a population of 1000 individuals and varying selection rates and mutation rates an initially rapid but small decrease in best fitness value which then plateaued to never become smaller until the maximum number of generations had been reached. The time estimated for one run of the cross entropy minimisation algorithm with the aid of the generational algorithm were decided to be too large, and this method was thus discarded for this work.

5.3 Future work

It would be interesting to investigate what more information could do for the accuracy of the algorithm - a mixture of satellite imagery and additional spatial information and data sets might improve the quality of the prior estimation of the spatial distribution, which in turn might generate results that more accurately estimates the actual spatial distribution. In our investigations, we deliberately kept our information sources to a small number of global sources in order for the method to work in regions all over the world. However, if one's aim is to disaggregate statistics in a region with many spatial data sets

freely available it could be interesting to see if the optimisation improves, especially in regards of separating cropland and pasture classifications from one another.

We have included cropland and pasture land uses in our investigations, but another example of a large driving factor of deforestation is forest plantations. From the ESA-CCI data set alone, it is difficult to distinguish between wild forest and planted forest from the tree cover classifications. It can be conceived that with some spatial information of where forest plantations are located, or with some method estimating its placement from other relevant spatial information (such as roads or paper mills), forest plantations could be included in our cross entropy minimisation algorithm.

At the moment we are using a deterministic method of finding optimum of our cross entropy cost function, a method which has no guarantee of finding the global minimum. It would be interesting to see if the performance of the cross entropy minimisation algorithm improves with another optimisation algorithm, perhaps with the use of a stochastic algorithm with a different set up than the ones investigated in this work.

Bibliography

[1] P. G. Curtis, C. M. Slay, N. L. Harris, A. Tyukavina, and M. C. Hansen, "Classifying drivers of global forest loss", *Science*, vol. 361, no. 6407, pp. 1108–1111, 2018.

[2] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. Stehman, S. J. Goetz, T. R. Loveland, *et al.*, "High-resolution global maps of 21st-century forest cover change", *Science*, vol. 342, no. 6160, pp. 850–853, 2013.

[3] P. A. Matson, W. J. Parton, A. G. Power, and M. J. Swift, "Agricultural intensification and ecosystem properties", *Science*, vol. 277, no. 5325, pp. 504–509, 1997.

[4] E. M. Bennett, S. R. Carpenter, and N. F. Caraco, "Human impact on erodable phosphorus and eutrophication: A global perspective: Increasing accumulation of phosphorus in soil threatens rivers, lakes, and coastal oceans with eutrophication", *BioScience*, vol. 51, no. 3, pp. 227–234, 2001.

[5] J. A. Patz, P. Daszak, G. M. Tabor, A. A. Aguirre, M. Pearl, J. Epstein, N. D. Wolfe, A. M. Kilpatrick, J. Foufopoulos, D. Molyneux, *et al.*, "Unhealthy landscapes: Policy recommendations on land use change and infectious disease emergence", *Environmental health perspectives*, vol. 112, no. 10, pp. 1092–1098, 2004.

[6] P. Snyder, C. Delire, and J. Foley, "Evaluating the influence of different vegetation biomes on the global climate", *Climate Dynamics*, vol. 23, no. 3-4, pp. 279–302, 2004.

[7] L. Tracewski, S. H. Butchart, M. Di Marco, G. F. Ficetola, C. Rondinini, A. Symes, H. Wheatley, A. E. Beresford, and G. M. Buchanan, "Toward quantification of the impact of 21st-century deforestation on the extinction risk of terrestrial vertebrates", *Conservation Biology*, vol. 30, no. 5, pp. 1070–1079, 2016.

[8] I. Turner, "Species loss in fragments of tropical rain forest: A review of the evidence", *Journal of applied Ecology*, pp. 200–209, 1996.

[9] S. L. Pimm and P. Raven, "Extinction by numbers", *Nature*, vol. 403, no. 6772, pp. 843–845, 2000.

[10] A. P. Dobson, S. L. Pimm, L. Hannah, L. Kaufman, J. A. Ahumada, A. W. Ando, A. Bernstein, J. Busch, P. Daszak, J. Engelmann, *et al.*, "Ecology and economics for pandemic prevention", *Science*, vol. 369, no. 6502, pp. 379–381, 2020.

[11] F. Pendrill, U. M. Persson, J. Godar, and T. Kastner, "Deforestation displaced: Trade in forest-risk commodities and the prospects for a global forest transition", *Environmental Research Letters*, vol. 14, no. 5, p. 055 003, 2019.

[12] F. Pendrill, U. M. Persson, J. Godar, T. Kastner, D. Moran, S. Schmidt, and R. Wood, "Agricultural and forestry trade drives large share of tropical deforestation emissions", *Global Environmental Change*, vol. 56, pp. 1–10, 2019.

[13] L. You and S. Wood, "An entropy approach to spatial disaggregation of agricultural production", *Agricultural Systems*, vol. 90, no. 1-3, pp. 329–347, 2006.

[14] L. You, S. Wood, U. Wood-Sichra, and W. Wu, "Generating global crop distribution maps: From census to grid", *Agricultural Systems*, vol. 127, pp. 53–60, 2014.

[15] G. F. Watch. (2020). Brazil - mato grosso, Global Forest Watch, [Online]. Available: https://gfw.global/3l72ICO (visited on 08/20/2020).

[16] IBGE. (2020). Mapa da utilização da terra no censo agropecuário, Instituto Brasileiro de Geografia e Estatística, [Online]. Available: https://www.ibge.gov.br/geociencias/informacoes-ambientais/cobertura-e-uso-da-terra/10867-cobertura-e-uso-da-terra.html?=&t=o-que-e (visited on 08/14/2020).

[17] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: John Wiley & Sons, Inc, 1991.

[18] Mathworks. (2020). Fmincon, Mathworks, [Online]. Available: https://se.mathworks.com/help/optim/ug/fmincon.html (visited on 08/13/2020).

[19] IBGE. (2020). Censo agropecuário, Instituto Brasileiro de Geografia e Estatística, [Online]. Available: https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/21814-2017-censo-agropecuario.html?=&t=o-que-e (visited on 08/14/2020).

[20] ——, (2020). Censo agropecuário, Tabela 854 - número de estabelecimentos agropecuários e área dos estabelecimentos por utilização das terras, condição do produtor em relação às terras, tempo em que o produtor dirige o estabelecimento, grupos de área total e associação à cooperativa e/ou à entidade de classe, Instituto Brasileiro de Geografia e Estatística, [Online]. Available: https://sidra.ibge.gov.br/Tabela/854 (visited on 08/14/2020).

[21] MapBiomas. (2020). Mapbiomas, MapBiomas, [Online]. Available: https://mapbiomas.org/ (visited on 08/20/2020).

[22] IBGE. (2020). Operação censitária, Instituto Brasileiro de Geografia e Estatística, [Online]. Available: https://censo2010.ibge.gov.br/materiais/guia-do-censo/operacao-censitaria.html (visited on 08/19/2020).

[23] ——, *Mapeamento de uso e cobertura da terra utilizando os dados numericos do censo agropecuario 2006*, Instituto Brasileiro de Geografia e Estatística, 2018.

[24] ESA, *Land cover cci product user guide version 2*, 2017. [Online]. Available: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.

[25] C. C. C. Service, *Product user guide and specification, Icdr land cover 2016 and 2017*, UCLouvain/ Pierre Defourny, 2019. [Online]. Available: https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-land-cover?tab=doc.

[26] ——, *Product quality assessment report, Icdr land cover 2016 and 2017*, UCLouvain/ Pierre Defourny, 2019. [Online]. Available: https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-land-cover?tab=doc.

[27] F. Pendrill and U. M. Persson, "Combining global land cover datasets to quantify agricultural expansion into forests in latin america: Limitations and challenges", *PloS one*, vol. 12, no. 7, e0181202, 2017.

[28] MapBiomas. (2020). Land cover/use, MapBiomas, [Online]. Available: https://plataforma.mapbiomas.org/map# (visited on 08/20/2020).

[29] M. Wahde, *Biologically inspired optimization methods: an introduction*. WIT press, 2008.

[30] A. P. Piotrowski, J. J. Napiorkowski, and A. E. Piotrowska, "Population size in particle swarm optimization", *Swarm and Evolutionary Computation*, p. 100 718, 2020.

[31] G. T. Pulido and C. A. C. Coello, "A constraint-handling mechanism for particle swarm optimization", in *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No. 04TH8753)*, Ieee, vol. 2, 2004, pp. 1396–1403.

[32] S. G. B. Rylander and B. Gotshall, "Optimal population size and the genetic algorithm", *Population*, vol. 100, no. 400, p. 900, 2002.

A

Calculation of entropy of random variable with zero-probability event.

The definition of entropy in Equation 2.3 can be rewritten into

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log(p(x)). \tag{A.1}$$

This rewriting makes it evident to see that terms in the sum in this equation is undefined for events $x \in \mathcal{X}$ for which $p(x) = 0$, since the value for that term in the sum would be "$0 \log(0) = 0 \cdot -\infty$". We can however investigate the limit of the term as $p(x)$ goes towards 0 (from the positive side). For simplicity in this calculation, we will substitute $p(x)$ with p, making the limit of interest

$$\lim_{p \to 0^+} p \log(p). \tag{A.2}$$

For functions $f(x)$ and $g(x)$ that are differentiable on an open interval I with a possible exception at a point ξ that is contained in I, with the following conditions are fulfilled:

- $\lim_{x \to \xi} f(x) = \lim_{x \to \xi} g(x) = \begin{cases} 0 \\ \pm\infty \end{cases}$,
  meaning that the limit of functions f, g when x goes to ξ are both either 0 or $\pm\infty$,
- $g'(x) \neq 0,\ \forall x \in I,\ x \neq \xi$,
- $\exists \lim_{x \to \xi} \frac{f'(x)}{g'(x)}$,

L'Hôpital's rule states that

$$\lim_{x \to \xi} \frac{f(x)}{g(x)} = \lim_{x \to \xi} \frac{f'(x)}{g'(x)}. \tag{A.3}$$

By observing that our limit in Equation A.2 can be rewritten into

$$\lim_{p \to 0^+} \frac{\log(p)}{\frac{1}{p}}. \tag{A.4}$$

we can see that all conditions for using L'Hôpital's rule are met;

- $\log(p) \xrightarrow[p \to 0^+]{} -\infty$,

- $\frac{1}{p} \xrightarrow[x \to 0^+]{} \infty$,
- $\frac{1}{p} \neq 0,\ \forall x \in I = (0, \infty)$,
- $\lim_{p \to 0^+} \frac{(\log(p))'}{\left(\frac{1}{p}\right)'}$ exists as we will show below.

Our limit in Equation A.4 can therefore be calculated as

$$\lim_{p \to 0} \frac{\log(p)}{\frac{1}{p}} = \{\text{L'Hôpital's rule}\} = \lim_{p \to 0} \frac{\frac{1}{p}}{-\frac{1}{p^2}} = \lim_{p \to 0} -\frac{p^2}{p} = \lim_{p \to 0} -p = 0. \tag{A.5}$$

Thus, we can conclude that terms in our definition of entropy in Equation 2.3 where the probability of the event occuring is 0 will add 0 to the value of the entropy.

B

Legends

B.1 Land Cover legend for ESA-CCI-dataset

Label		Value		Colour
Global	Regional	Glo.	Reg.	
No data		0		
Cropland, rainfed		10		
	Cropland, rainfed, herbaceous cover		11	
	Cropland, rainfed, tree or shrub cover		12	
Cropland, irrigated or post-flooding		20		
Mosaic cropland (> 50%) / natural vegetation (tree, shrub, herbaceous cover) (< 50%)		30		
Mosaic natural vegetation (tree, shrub, herbaceous cover) (> 50%) / cropland (< 50%)		40		
Tree cover, broadleaved, evergreen, closed to open (> 15%)		50		
Tree cover, broadleaved, deciduous, closed to open (> 15%)		60		
	Tree cover, broadleaved, deciduous, closed (> 40%)		61	
	Tree cover, broadleaved, deciduous, open (15 − 40%)		62	
Tree cover, needleleaved, evergreen, closed to open (> 15%)		70		
	Tree cover, needleleaved, evergreen, closed (> 40%)		71	

Continued on next page

Continued from previous page

Label		Value		Color
Global	**Regional**	**Glo.**	**Reg.**	
	Tree cover, needle-leaved, evergreen, open (15 – 40%)		72	
Tree cover, needleleaved, deciduous, closed to open (> 15%)		80		
	Tree cover, needle-leaved, deciduous, closed (> 40%)		81	
	Tree cover, needle-leaved, decidious, open (15 – 40%)		82	
Tree cover, mixed leaf type (broadleaved and needle-leaved)		90		
Mosaic tree and shrub (> 50%) / herbaceous cover (< 50%)		100		
Mosaic herbaceous cover (> 50%) / tree and shrub (< 50%)		110		
Shrubland		120		
	Evergreen shrubland		121	
	Deciduous shrubland		122	
Grassland		130		
Lichens and mosses		140		
Sparse vegetation (tree, shrub, herbaceous cover) (< 15%)		150		
	Sparse tree (< 15%)		151	
	Sparse shrub ($<15%)		152	
	Sparse herbaceous cover ($<15%)		153	
Tree cover, flooded, fresh or brackish water		160		
Tree cover, flooded, saline water		170		
Shrub or herbaceous cover, flooded, fresh / saline / brackish water		180		
Urban areas		190		
Bare areas		200		
	Consolidated bare areas	201		
	Unconsolidated bare areas	202		
Water bodies		210		
Permanent snow and ice		220		

B.2 Land Cover legend for IBGE

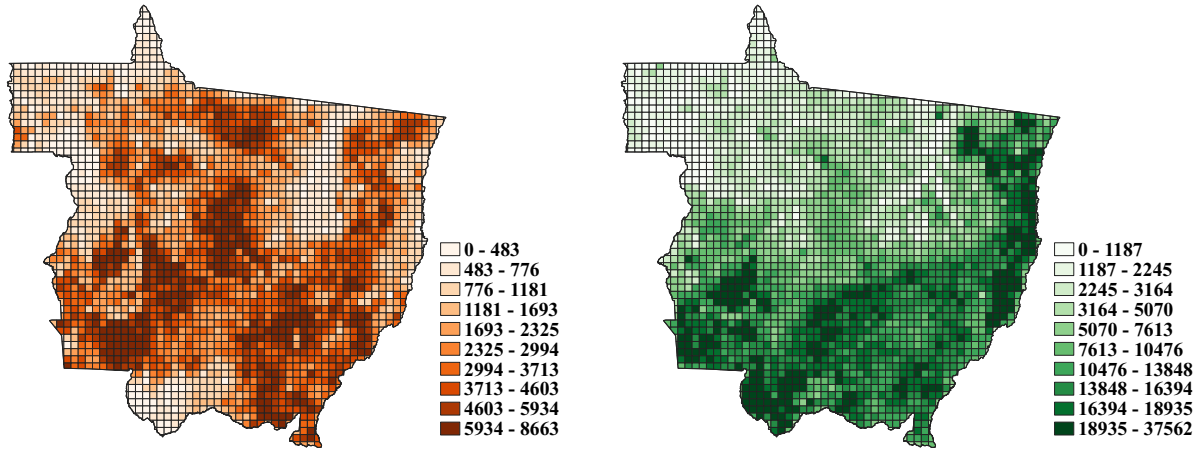
ID	Class	English translation	Colour
1	Lavouras permanentes	*Permanent cropland*	
2	Lavouras temporárias	*Temporary cropland*	
3	Pastagens naturais	*Natural pastures*	
4	Pastagens plantadas	*Planned pastures*	
5	Matas e/ou florestas naturais	*Natural woods and/or forests*	
7	Sistemas agroflorestais	*Agroforestry systems*	
9	Lavouras	*Cropland*	
10	Pastagens	*Pastures*	
11	Matas e/ou florestas	*Woods and/or forests*	
16	Outras coberturas e usos + Usos diversificados	*Other coverages and uses + Varied uses*	
22	Lavouras + Pastagens	*Cropland + Pastures*	
23	Lavouras + Matas e/ou florestas	*Agriculture + Woods and/or forests*	
24	Pastagens + Sistemas agroflorestais	*Pastures + Agroforestry systems*	
25	Pastagens + Outras coberturas e usos	*Pastures + Other coverages and uses*	
26	Pastagens + Lavouras	*Pastures + Cropland*	
27	Pastagens + Matas e/ou florestas	*Pastures + Woods and/or forests*	
29	Matas e/ou florestas + Outras coberturas e usos	*Woods and/or forests + Other coverages and uses*	
30	Matas e/ou florestas + Lavouras	*Woods and/or forests + Cropland*	
31	Matas e/ou florestas + Pastagens	*Woods and/or forests + Pastures*	
50	Usos diversificados	*Varied uses*	
51	Área com menos de 10% de ocupação por estabelecimentos agropecuários	*Area with less than 10% occupied by cropland and livestock establishments*	
52	Área entre 25% e 10% de ocupação por estabelecimentos agropecuários	*Area with between 25% and 10% occupied by cropland and livestock establishments*	
53	Área entre 50% e 25% de ocupação por estabelecimentos agropecuários	*Area with between 50% and 25% occupied by cropland and livestock establishments*	
	Área urbanizada	*Urban land*	

C

Additional results

C.1 Confusion matrix

Below, the results from the cross entropy minimisation algorithm run with the confusion matrix prior preprocessing method are presented. All results are generated by running the cross entropy minimisation algorithm 20 times for randomly generated initial conditions σ_0 order to increase the likelihood for us to find the global minimum. All legend entries are displayed in hectares.

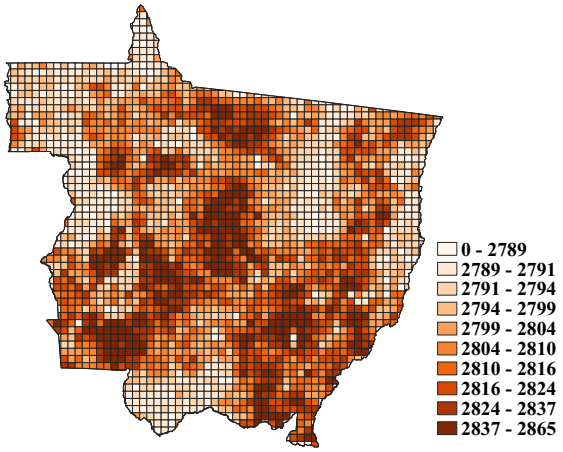


(a) *Farm area spatial distribution.* **(b)** *Pasture area spatial distribution*

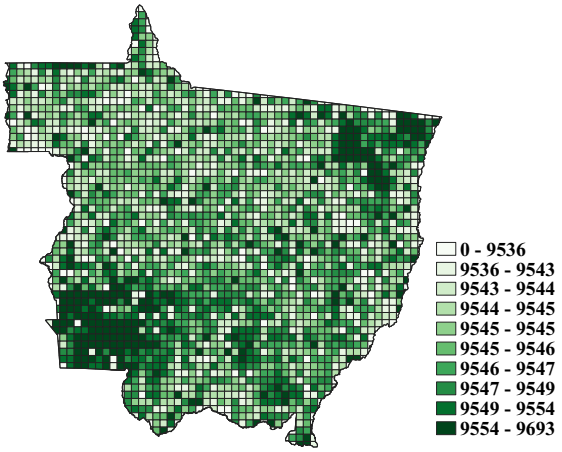
Figure C.1: *Land use area spatial distribution in Mato Grosso for the year 2006 obtained using the confusion matrix preprocessing method of the prior. Visualisation are done with quantile classification. All legend entries are displayed in hectares.*

C.2 Certainty parameter

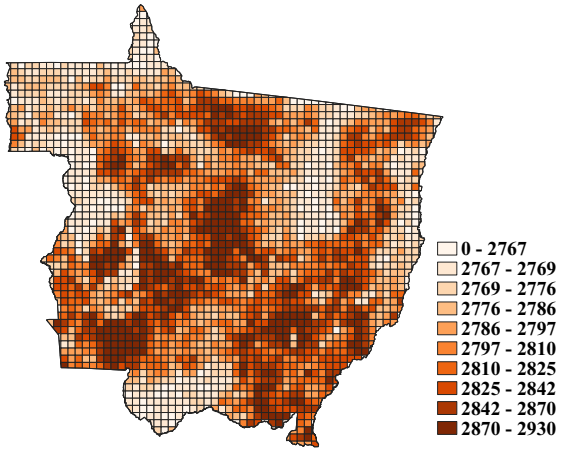
Below, the results from the cross entropy minimisation algorithm run with the certainty parameter prior preprocessing method for $\gamma = \{0.714, 0.850, 0.995\}$ are presented. All results are generated by running the cross entropy minimisation algorithm 20 times for randomly generated initial conditions for σ_0 to increase the likelihood for us to find the global minimum. All legend entries are displayed in hectares.



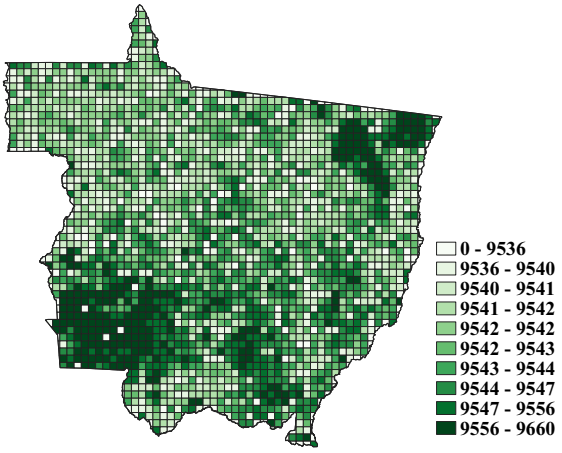
(a) *Farmland area distribution with* $\gamma = 0.714$.



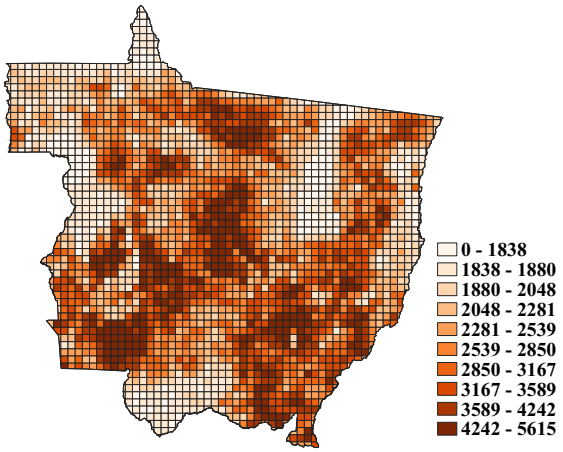
(b) *Pasture area distribution with* $\gamma = 0.714$.



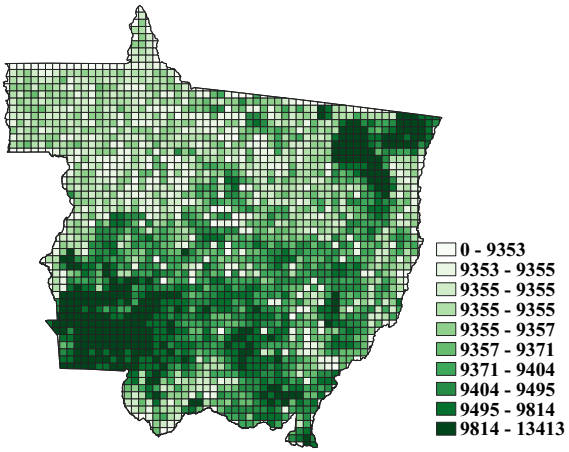
(c) *Farmland area distribution with* $\gamma = 0.850$.



(d) *Pasture area distribution with* $\gamma = 0.850$.



(e) *Farmland area distribution with* $\gamma = 0.995$.



(f) *Pasture area distribution with* $\gamma = 0.995$.

Figure C.2: *Land use area distribution obtained using the certainty parameter preprocessing method of the prior with* $\gamma = \{0.714,\ 0.850,\ 0.995\}$. *Visualisation are done with quantile classification. All legend entries are displayed in hectares.*

D

ESA-CCI's data set confusion matrix