Biobased Chemistry and Technology

MSc Thesis Biobased Chemistry and Technology

Dynamic Modelling and Control of Potato Proteins in Bulk Storage Facilities

Luc van Heeringen

1 November 2016

CONFIDENTIAL

WAGENINGEN UNIVERSITY
WAGENINGEN**UR**

OMNIVENT ✳

WAGENINGEN**UR**

Dynamic Modelling and Control of Potato Proteins in Bulk Storage Facilities

Name course	:	MSc Thesis Biobased Chemistry and Technology
Number	:	BCT-80436
Study load	:	36 ECTS
Date	:	01-11-2016
Student	:	Luc van Heeringen
Registration number	:	921208316080
Study programme	:	MBT
Report number	:	050BCT
Supervisor(s)	:	N. Grubben & K. Keesman
Examiners	:	K. Keesman & G. van Willigenburg
Group	:	Biobased Chemistry and Technology
Address	:	Bornse Weilanden 9
		6708 WG Wageningen
		The Netherlands

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Abstract

Potatoes (*Solanum tuberosum*) are an important agricultural product, as well as one of the main sources of natural starch. For the last several decades, the valorisation of potato protein, that is produced as a by-product in the starch industry, has been gaining interest as well. As the production of potatoes is a seasonally dependent process, and demand is high year-round, the storage of potatoes is of great importance. During long-term storage, the protein content of potatoes has been shown to be temperature-dependent, which is why research is needed on the optimal storage conditions for a high protein production.

In this research, a dynamic potato protein content model is developed, using Michaelis-Menten kinetics to describe the processes of protein synthesis and degradation during long-term storage. The single control variable that influences these processes is the storage temperature. The model includes the total protein concentration as well as the concentration of patatins and protease inhibitors, two major groups of proteins in potato tubers. It was validated using protein content measurements performed on potatoes of two different cultivars, after which the effect of the storage temperature on the protein content was assessed using model simulations.

In addition, an optimal storage temperature was calculated to maximise the total amount of protein during long-term storage. The costs of cooling or heating a storage facility were weighed against the additional profits that the temperature control provided to formulate an optimal storage temperature strategy.

Table of Contents

Acknowledgements

I would like to thank both my supervisors, Nik Grubben and dr. ir. Karel Keesman, for their assistance and guidance during this thesis project. I would also like to thank the director of Omnivent, ir. Luuk Salomons, for his support of the study.

Wageningen University & Research

1 November, 2016

1   Introduction

This section contains the background of the problem addressed in this thesis report, as well as the formulation of the problem. The goals and approach to reach these goals are also described here. Finally, the outline of this thesis is also mentioned.

1.1   Background

Potato (*Solanum tuberosum*) is one of the most important crops worldwide, accounting for about 45% of the worldwide tuber crop production (WCRTC, 2016). Potato is used as a food crop, but is also extensively used for the production of starch. During the starch production process, protein-rich waste water is produced (Løkra & Strætkvern, 2009). This protein solution (commonly called potato juice) has been the subject of many studies to determine its composition and functional properties (Holm & Eriksen, 1980; Kapoor *et al.,* 1975; Ralet & Guéguen, 2000). It was discovered that potato juice consists of three major fractions of proteins; the patatins, the protease inhibitors, and other proteins of high molecular weight (Pots, 1999).

Patatin is the name of a group of glycoproteins first characterized by Racusen & Foote (1980), with molecular masses of 40-45 kDa, thought to be the main storage proteins in potato tubers (Shewry, 2003). The protease inhibitors are a group of proteins with molecular masses of 7-21 kDa, which are thought to have several positive dietary effects (Hill *et al.,* 1990; Kennedy, 1998). The quantities of these proteins vary among different cultivars, but a potato tuber in general consists of 35-40% patatin, 25-50% protease inhibitors, and the rest are high molecular weight proteins. Both protease inhibitors and patatin have been evaluated for their nutritional quality (Kapoor *et al.,* 1975) and their ability to form stable emulsions (Ralet & Guéguen, 2000). After the costs of a possible thermal treatment were considered, Ralet & Guéguen (2000) concluded that the fraction containing patatins showed the most promise for industrial use. The rest of the proteins with high molecular mass are of relative insignificance for industrial use.

It has been found earlier that both total soluble protein and patatin levels fluctuate during long-term storage of potato tubers, depending on cultivar and **storage temperature** (Brierly *et al.,* 1996; Nowak, 1977; Brasil *et al.,* 1993). Storage of potato over long periods of time is useful because of the seasonality of tuber harvest. In Western Europe tubers are harvested in late summer, but are expected to be on the market during the whole year. Over a storage period of more than 25 weeks, roughly three periods can be distinguished. The post-harvest stage (first 5-10 weeks), followed by the dormancy stage (week 10-25), after which emergence will occur (Brierly *et al.,* 1996). The dormancy period can be extended by applying chemicals that inhibit sprouting. The potatoes that were used in this thesis were all chemically treated with Chlorpropham to inhibit sprouting.

Throughout the entire storage period the level of patatin decreases, but the total soluble protein level shows a more complicated pattern. Especially when comparing different potato cultivars, the dynamics and fluctuations of the protein content are not always the same. For instance, some studies show an increase in protein content (Brierly *et al.,* 1997) while others show a decrease (Kumar & Knowles, 1993).

These fluctuations in the total soluble protein and the differences in dynamics might mean that there is an optimum in terms of storage temperature at which the protein levels, and especially that of patatin, are highest.

1.2   Problem Formulation

An optimal storage temperature strategy is needed to be able to supply potatoes with high protein levels for industrial extraction. Also, protein-rich potatoes can be used for the frying industry because high levels of protein are associated with low levels of free amino acids (Brierly *et al.,* 1996). Low levels of amino acids are favourable for the frying industry because they are involved in the Maillard browning reaction (Khanbari & Thompson, 1993). Thus, depending on the use of the potato tuber, different optimal storage conditions can be found. To be able to evaluate the specific effects of these conditions more data has to be collected, and a mathematical model to be able to predict and control protein levels has to be constructed.

Modelling the exact reaction mechanisms involved in protein turnover is too complicated, because protein levels depend on a balance between synthesis and degradation. These processes in turn depend on gene expression and regulation, which are extremely hard to model. For controller design however, we assume that a simple dynamic protein model that depends on the temperature in the storage facility suffices.

1.3   Research Goals

The goal of this research is to construct a process-based mathematical model of appropriate complexity which describes the dynamics of protein turnover of potato tubers during long-term storage. The dynamic model can be used to design an optimal strategy for the storage temperature, that maximises the amount of protein inside the potato tubers at the end of the storage period.

To reach this, the following research question can be stated:

- **What is an appropriate protein model for the control of protein levels in potato tubers during long-term storage?**

Several sub questions relating to this can be formulated to provide some background information:

- How can the parameters of the system be estimated?
- Which processes influence the protein content in potato tubers during long-term storage?
- What is the best storage temperature to maximise the protein content during long-term storage?

1.4   Approach

First, the properties of potato proteins will be investigated. Also, the processes and reactions that influence protein content during long-term storage will be discussed. Finally, the kinetics of these reactions will be investigated to be able to formulate a mathematical model that describes the rate of change of the protein content.

To check the validity of the proposed model and estimate any unknown parameters, protein measurements on potato tubers of different cultivars stored at different conditions will be performed. Tubers stored from October 2015 till June 2016 will be analysed for their protein content to observe the dynamics of protein turnover during storage.

Finally, the controllability and observability of the system model will be assessed, and an optimal control strategy for the total protein level will be designed. Using optimal control algorithms, an

optimal storage temperature profile can be calculated that results in the highest protein content of the potatoes.

1.5   Outline Thesis

Chapter 2 will contain a literature review of potato proteins and their applications. The effects of long-term storage will also be discussed here, as well as the modelling possibilities and methods.

In chapter 3, the experiment setup will be described. The total soluble protein of potato tubers during storage will be measured, as well as the concentrations of several specific fractions. The total amino acid content will also be investigated. The results of these experiments are presented in chapter 4.

The mathematical modelling and the assumptions that were made for this are discussed in chapter 5. First, a basic protein turnover model will be presented, after which the model will be expanded with inhibition terms. The control that will be applied to the protein content in the model will be outlined in chapter 6.

Finally, chapters 7 and 8 contain the discussion points and the conclusions of this master thesis project.

2   Theory

The potato is one of the world's major crops, being produced and consumed all over the world. In 2013, potato tubers had the fifth highest crop production after sugar cane, maize, rice, and wheat (FAO, 2015). Potatoes are a very versatile crop, being mainly used for human consumption and starch production. However recently, the production of potato proteins as a by-product of the starch production process has become increasingly popular (Løkra & Strætkvern, 2009). This increases the need for the understanding of the metabolism of a potato, which influences both the starch and protein contents in the cells. In this section, some general background on the cellular processes occurring in potatoes during storage will be provided.

The two main processes that will be discussed are catabolism and anabolism (Figure 1). Catabolism can be explained as the breakdown of large molecules to smaller molecules, resulting in the production of energy. For instance, the process of glycolysis turns a glucose molecule into two pyruvate molecules, while simultaneously producing two adenosine triphosphate (ATP) molecules. ATP is a high-energy containing molecule that cells use to fuel other reactions. The pyruvate molecules can then be either used to construct other molecules or they can be broken down further to carbon dioxide (CO_2), resulting in the production of more ATP.
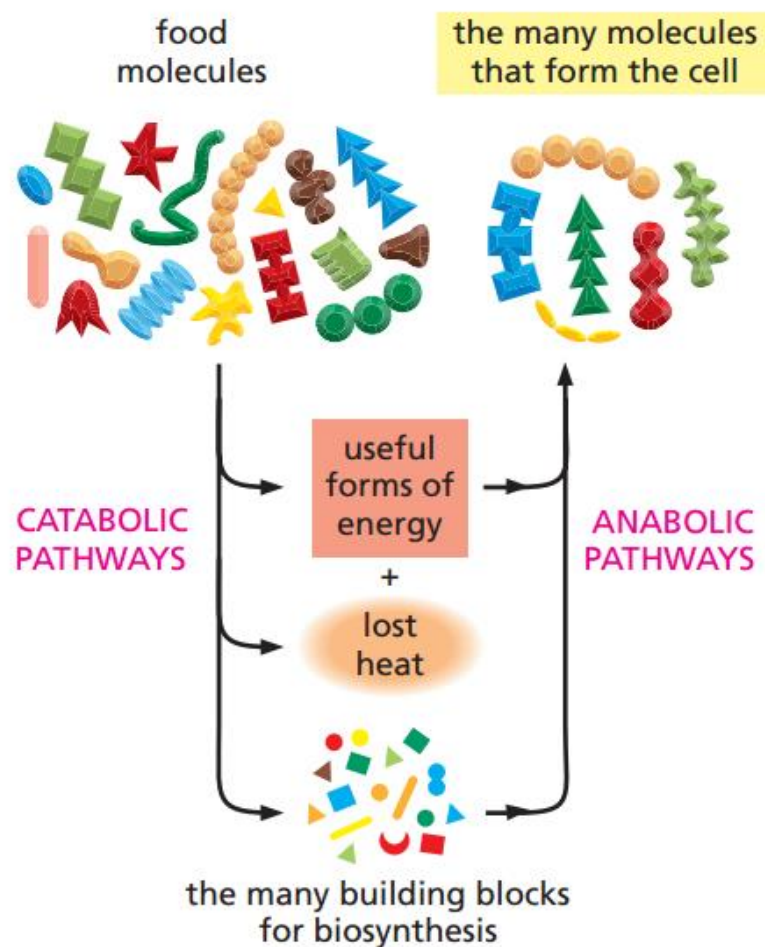

Figure 1. Catabolism and anabolism in a typical cell. Taken from Albert *et al.* (2010).

The counteracting process, anabolism, is the synthesis of larger, more complex molecules from smaller components by consuming energy. An important example of this is the synthesis of starch in plants, also showing the origin of starch as an energy storage molecule. Plants use the energy from the sun to form sugars during the day, and use the stored energy at night to grow. Potatoes and other tuber-forming plants also store energy in the form of starch in their tubers to survive periods of drought or the winter season.

Inside the cells of potato tubers, the three main examples of anabolic and catabolic processes are the interconversion of starch and sugars, the respiration of these sugars to provide energy, and the degradation and synthesis of proteins. This last process is called protein turnover.

2.1   Protein Turnover

To be able to study the protein levels inside an organism, it is vital to have some further understanding of the process of protein turnover. Turnover has sometimes been falsely specified as simply the sum of protein synthesis and protein degradation, but this would mean that there is no turnover when protein levels are constant. However, even though protein content may be constant, there is always continuous breakdown and synthesis of proteins (Figure 2). These two processes are often a necessity for an organism to maintain the concentrations of different enzymes. For example, in plants, when nutrients are not readily available in the soil, the only way to provide new amino acids is to break down existing protein. Protein breakdown to provide new amino acids is especially present in potato tubers during storage, which have no nutritional intake at all due to the lack of uptake through roots.
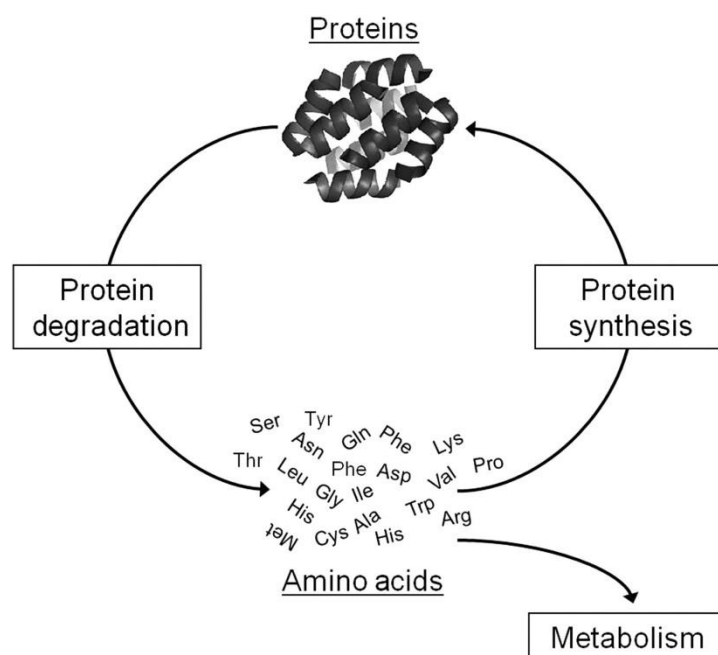

Figure 2. Schematic overview of protein turnover. Adapted from Baskin & Taegtmeyer (2011).

Another important aspect of protein turnover is that it can regulate enzyme activity (Glasziou, 1969; Marcus, 1971; Filner *et al.*, 2016). Instead of inducing a higher activity in an enzyme, simply raising the concentration of the enzyme will often be a solution that is less complicated and more energy-efficient for an organism. Inducing a higher activity would require conformational changes of the enzyme, which normally needs the activity of another different enzyme. Increasing the enzyme level

will also result in a quicker response (for instance to a virus or a similar intruder) when compared to a mechanism like inhibition, which relies first on synthesis of the inhibitor and the subsequent binding of the inhibitor to the target protein. In this case, direct degradation of the protein results in a faster response (Filner *et al.,* 2016).

2.2   Potato Proteins

Potato juice contains approximately 25-30% protein. These proteins have been classified by Pots (1999) into three groups: Patatins, protease inhibitors, and other proteins. Patatin is the name for a group of glycoproteins with similar molecular masses of around 40-45 kDa. It is thought to be the major storage protein in tubers because of its high accumulation in tuber tissue. Protease inhibitors are a more diverse group of proteins that have molecular masses of 7-21 kDa. Both these groups are of commercial interest because of their properties, which will be discussed below. The remaining proteins are mostly high molecular weight proteins such as starch synthetase, which will not be discussed in this thesis.

2.2.1   Patatin

Patatin was first named and fully characterised by Racusen & Foote (1980), although the isolation and partial characterisation was performed in the early 70's. They found that it is a glycoprotein that contains mannose as its main sugar. While first thought to be a single protein, a separation based on isoelectric focusing shows that there are several homogeneous proteins with similar molecular weights of roughly 40 to 45 kDa.

Next to its (presumed) function as a storage protein, it also exhibits enzymatic activity. Racusen (1986) found that the protein had lipid acyl hydrolase (LAH) activity acting on a large range of substrates, as well as exhibiting esterase activity. The function of patatin in potato tubers is therefore probably more than just storage of nitrogen. Supporting this are the several indications that patatin plays a role in the defence mechanism of tubers. Strickland *et al.* (1995) found that including patatin in the diet of corn rootworm, a major pest in North American corn fields, inhibited larval growth by 50%. They concluded that patatin can provide defence against some insect pests by negatively affecting the lipid metabolism. Another indication for a role in plant defence was discovered by Dhondt *et al.* (2000). In tobacco leaves infected with the tobacco mosaic virus, three patatin-like proteins were induced upon infection. Following this was an accumulation of lipid-derived defence signals, suggesting the involvement of the patatin-like proteins in the synthesis of defence molecules.

Patatin also has properties that make it interesting for industrial use. First of all, because of its specific LAH activity towards mono-acyl glycerols (Andrews *et al.*, 1988), the enzyme is very suitable for the industrial production of these compounds from glycerol and fatty acids. Furthermore, its ability to form stable emulsions was investigated by Ralet & Guéguen (2000), as well as its resistance to thermal processing. They found that, under mildly acidic conditions, patatin allowed the forming of stable emulsions with good resistance to coalescence. Two decades earlier, Holm & Eriksen (1980) had already found that potato proteins were superior to soy proteins regarding emulsification properties.

Finally, the nutritional value of patatin has been studied extensively as well. The biological value of patatin, which is a measure of the amount of protein that is incorporated in the body after ingestion, was found to be higher than that of peas, bread, and wheat (Schuphan, 1959). Also, the essential amino acid composition of patatin was shown to be comparable or even better than the Food and

Agriculture Organization's reference protein (Kapoor *et al.*, 1975), which shows that patatin can be used as a protein supplement in food products or animal feed.

2.2.2 Protease Inhibitors

First characterized in potato by Suh *et al.* (1990), these compounds form a group of proteins with molecular masses of about 16-25 kDa. There are several different types, characterized either by the type of protease they inhibit or the mechanism of action they use. Two different mechanisms are shown in Figure 3. In potato tubers, the type that is most abundant is the serine protease inhibitor, which inhibits the target protease by altering its active site (Gettins, 2002).
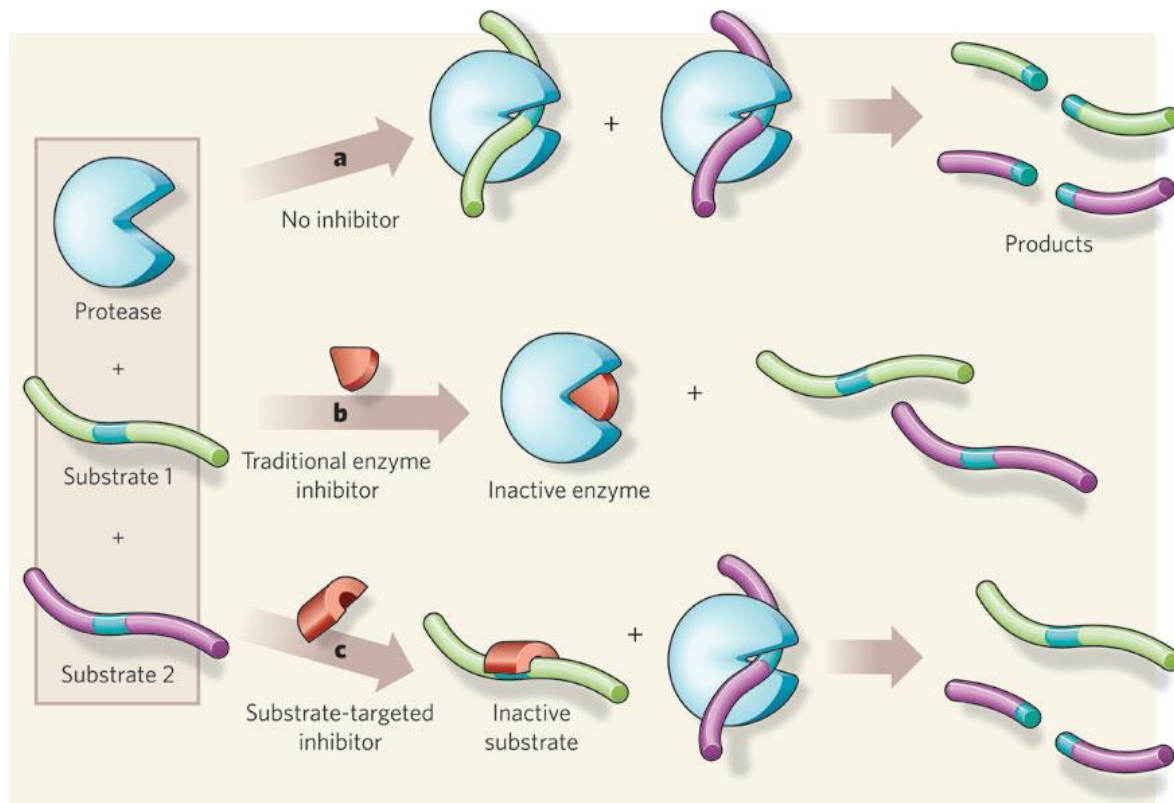
Figure 3. Mechanisms of action of protease inhibitors. Taken from http://e-gaho.blogspot.nl/2012/02/digestive-enzyme-therapy-cure-illnesses.html.

Protease inhibitors are thought to be part of the plant defence mechanism, neutralizing proteolytic enzymes from invading microorganisms and insects (Jongsma, 1995). During the last 25 years, this group of proteins has also gained attention because of possible positive dietary effects. Hill *et al*. (1990) found that adding protease inhibitor II from potato reduces the energy intake of the rest of the meal. Adding it could help people with eating disorders reduce their food intake. Another possible application for protease inhibitors was reviewed by Kennedy (1998). She states numerous studies that show the anticarcinogenic effect of protease inhibitors both *in vitro* and *in vivo*. Protease inhibitors have now been well established as chemopreventive agents against cancer.

The emulsification properties of protease inhibitors were investigated as well (Ralet & Guéguen, 2000). It was found that the use of protease inhibitors leads to very stable emulsions over a large pH range, as well as mildly salt conditions. This corresponds with the situation in actual food emulsions, showing that these compounds can be used in the food industry as creaming and emulsifying agents.

2.3   Storage

Several studies have already looked into the effects of long-term storage on the compositional parameters of potato tubers. The effect of different storage temperatures on protein content has been investigated by Brasil *et al*. (1993), who found that a higher storage temperature is associated with higher soluble protein levels. They hypothesize that a lower temperature slows down the metabolic activity of the potatoes, and higher temperature could be increasing the activity of enzymes involved in the regulation of protein levels. This is confirmed to some extent by Brierly *et al.* (1996), who analysed the proteinase activities of tubers stored at either 5°C or 10°C. They observed low enzyme activity during the first 20 weeks of storage, after which the activity increased greatly. For a higher storage temperature, this increase was double compared to the low-temperature stored tubers. The same research also confirms the higher protein levels at higher storage temperatures during the years 1992-1993. However, during the 1989-1990 season, the results differ greatly. That may be partly caused by a phenomenon that influences many of the conclusions drawn in most articles between 1950 and 2000, which is the lack of data points. Most of these studies have less than 10 measurements over a period of 30-40 weeks, which is often too little to draw conclusions from.

Another thing that becomes clear after comparing several datasets is that both the protein content and the dynamics during storage are very much dependent on the cultivar that is being analysed. For instance, Pots *et al.* (1999) compared the amount of soluble protein of the cultivars Bintje, Desiree, and Elkana over a period of 50 weeks. While the soluble protein of the first two cultivars decreased steadily over the whole period, it increased halfway through for Elkana potatoes. Consequently, there are several articles indicating an increase in protein content during storage (Brierly *et al.,* 1996; Brierly *et al.,* 1997), whereas others show a decline in protein content (Kumar & Knowles, 1993; Nowak, 1977), and again other articles that show such little change that they conclude that there are no protein dynamics at all (Blenkinsop *et al,* 2002; Mazza, 1983). These differences are also shown in Table 1, including their respective observed concentration ranges. It can be seen that the ranges differ between studies/cultivars, which supports the observation that the cultivar matters for the protein dynamics. Furthermore, the cultivars investigated in this report (Agria & Miss Malina) have not been analysed regarding their protein content at all, which makes formulating a hypothesis increasingly difficult.

Table 1. Protein dynamics and observed total protein concentration ranges found in literature. Ranges for constant protein dynamics are differences between cultivars, other values are changes over time. *This article shows an initial decrease in protein content, but an overall increase after the entire storage period.

Article	Protein Dynamics	Concentration Range	Cultivar Used
Blenkinsop *et al.* (2002)	Constant	4.8 - 10.9 mg/g FW	Novachip/Monona
Brierly *et al.* (1996)	Increasing	1.0 – 14.0[*] mg/g FW	Pentland Dell/Record
Brierly *et al.* (1997)	Increasing	1.5 – 13.5 mg/g FW	Pentland Dell
Kumar & Knowles (1993)	Decreasing	3.75 – 5.75 mg/g DW	Russet Burbank
Mazza (1983)	Constant	8 – 11.5 % of DW	Russet Burbank/Norchip
Nowak (1977)	Decreasing	9 – 24 mg/g DW	Baca/Bem

Another aspect of protein metabolism during storage is the amount of free amino acids. As proteins are composed of chains of amino acids, their abundance is directly linked to the synthesis rate of proteins. Matsuura-Endo *et al.* (2006) investigated the effect of storage temperature on the amount

of free amino acids in several potato tuber varieties, and found that lower temperatures can increase the amount of amino acids. This effect was also found by Davids *et al.* (2004) when comparing storage at 0°C, 12°C, and 24°C. As with the protein content, the effects of harvest year and cultivar have been investigated for the amino acid content as well. Davies (1977) made an extensive report on the amino acid compositions of 31 varieties of England and Ireland, indicating the big differences between different cultivars, whereas Viklund *et al.* (2008) investigated the impact of harvest year and cultivars to conclude that the amino acid content can vary significantly between different seasons and clones.

2.4   Modelling and Control

The behaviour of the total protein content and of each individual fraction will be displayed in the form of a state-space dynamic model, using a set of ordinary first-order differential, algebraic equations. The kinetics of the processes described by these differential equations will be described here. In order to investigate the properties of the model, the nonlinear dynamic model will be linearized. The protein content will then be regulated by applying optimal control theory. The relevant theory for these procedures will be explained in the following sections.

2.4.1   Model Structure

A general model structure for the rate of change of compounds is expressed as the sum of a rate of synthesis, a rate of degradation, and a reaction rate. This can be shown as follows, where X defines the compound and r defines the rate of each term:

$$\frac{dX}{dt} = r_s - r_d + r_r \tag{2.1}$$

In our case the compounds are the total protein content, the protease inhibitor content, and the patatin content. In all three cases, the compounds do not react and thus there is no reaction rate. This leaves the two processes of synthesis and degradation.

Both protein synthesis and degradation have been the subject of many studies investigating either their mechanics or their kinetics. While some studies have modelled the rate of protein synthesis by describing the entire process of gene expression and subsequent protein synthesis by ribosomes (Antoun *et al.*, 2006; von Heijne *et al.*, 1987), others have tried approaching synthesis by using Michaelis-Menten kinetics for the entire process (Lancelot *et al.*, 1986; Danfær, 1991). Several studies have also found that protein degradation can also be described by Michaelis-Menten kinetics (Hersch *et al.*, 2004; Grilly *et al.*, 2007; Gérard *et al.*, 2009). The general form of a Michaelis-Menten reaction rate is as follows:

$$r_r = V_{MAX} * \frac{[X]}{K_M + [X]} \tag{2.2}$$

Here, r symbolises the reaction rate in s^{-1}, and V_{MAX} is the maximum reaction rate when substrate concentrations are not limiting. The substrate concentration is $[X]$, and K_M is the Michaelis constant, both in $mg * g^{-1}$ fresh weight.

However, there are also some indications that protein turnover is a temperature-dependent process (Strnadova *et al.*, 1986). This can be incorporated in each term by combining the Michaelis-Menten kinetic model with the Arrhenius model, as has been done earlier by Davidson *et al.* (2012). This changes the parameter V_{MAX} according to the Arrhenius equation:

$$V_{MAX} = A * e^{-\frac{E_a}{RT}}, \tag{2.3}$$

where A is the pre-exponential factor (s^{-1}), E_a is the activation energy ($J * mol^{-1}$), R is the universal gas constant ($J * K^{-1} * mol^{-1}$), and T is the temperature in Kelvin. Each rate in the system will be described by a combination of these two kinetic models. The formulation of the model is shown in chapter 5 of this report.

2.4.2   Linearization

The complicated nature of dynamic processes, such as protein synthesis, protein degradation and using Michaelis-Menten kinetics, make the resulting set of equations nonlinear. A system with non-linear equations is defined as a system in which the output is not directly proportional to the input. To be able to fully analyse all aspects of the model, such as controllability and observability, it is needed to perform a linear approximation of the non-linear model.

Identification of non-linear dynamics can be done by determining the Jacobian matrices of the vector functions $f(x,u,p,t)$ and $g(x,u,p,t)$. Here, $\frac{dx(t)}{dt} = f(x,u,p,t)$ and $y(t) = g(x,u,p,t)$, which define the states and outputs of a dynamic state-space model. The dependence on the parameters and time will hereafter not be mentioned explicitly in these equations, but remain present. The Jacobian matrices contain the partial derivatives of f and g with respect to x and u, resulting in four matrices. The linearity test consists of checking whether these Jacobian matrices depend on the variables x or u. When any of the four Jacobian matrices does depend on x or u the system is defined as non-linear and linear approximation should be applied.

The most common function approximation method is called the Taylor series expansion. For an infinitely differentiable function $f(x)$, the Taylor series can be represented as follows:

$$f(x) = f(\overline{x}) + \frac{f'(\overline{x})}{1!}(x - \overline{x}) + \frac{f''(\overline{x})}{2!}(x - \overline{x})^2 + \cdots \tag{2.4}$$

In this equation, $\overline{x}$ represents the time point around which the linearization takes place. If the deviations $(x - \overline{x})$ from this linearization point are small, we can approximate $f(x)$ by neglecting the higher-order terms. Neglecting the higher-order terms results in the following simplification:

$$f(x) = f(\overline{x}) + \frac{f'(\overline{x})}{1!}(x - \overline{x}) \tag{2.5}$$

Simplifying a curve like this is the same as constructing a tangent line through x_0 (or $\overline{x}$) as in Figure 4. When the deviation stays small, this tangent line is an accurate estimate of the actual function.
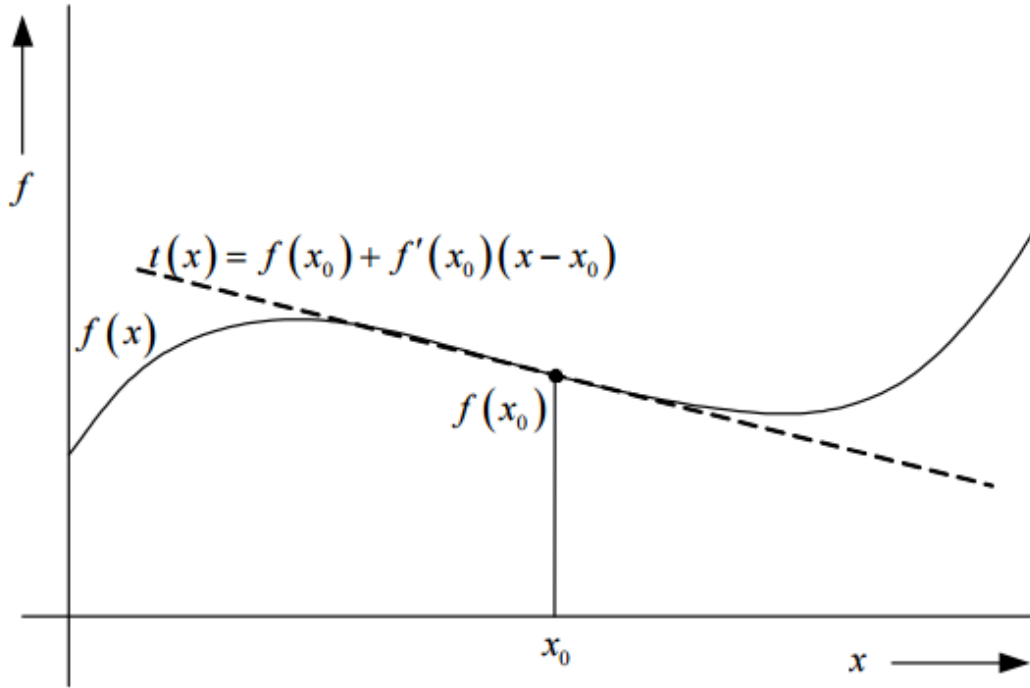
Figure 4. Approximation of $f(x)$ near$x = x_0$ by tangent line $t(x)$. Taken from Van Willigenburg & Van Ooteghem (2015).

This method can be expanded when the assessed function depends on multiple variables, such as a non-linear model depending on inputs (u) as well as states (x). When taking the derivative of a function dependent on multiple variables, we can take the partial derivatives of the function to the individual variables. The Taylor series around the linearization point $(\overline{x}, \overline{u})$ then turns into the following expression (neglecting higher-order terms):

$$f(x,u) \approx f(\overline{x}, \overline{u}) + \frac{\partial f}{\partial x}\bigg]_{x=\overline{x}, u=\overline{u}} (x - \overline{x}) + \frac{\partial f}{\partial u}\bigg]_{x=\overline{x}, u=\overline{u}} (u - \overline{u}) \qquad (2.6)$$

The input at the linearization point is represented by $\overline{u}$. The partial derivatives of f to x and u are the same as the Jacobian matrices mentioned earlier. The partial derivatives of g to x and u can be used in a similar way to approach $g(x,u)$ around the point $(\overline{x}, \overline{u})$:

$$g(x,u) \approx g(\overline{x}, \overline{u}) + \frac{\partial g}{\partial x}\bigg]_{x=\overline{x}, u=\overline{u}} (x - \overline{x}) + \frac{\partial g}{\partial u}\bigg]_{x=\overline{x}, u=\overline{u}} (u - \overline{u}) \qquad (2.7)$$

In order to use this information to construct a linearized model deviation variables have to be defined, describing the deviation from the linearization point. Each of these deviation variables has to remain small to be able to neglect the higher order terms of the Taylor series expansion.

$$\Delta x = x - \overline{x}$$

$$\Delta u = u - \overline{u}$$

$$\Delta y = y - \overline{y}$$

$$\overline{y} = g(\overline{x}, \overline{u})$$

Furthermore, it can be stated that:

$$\frac{d\Delta x}{dt} = \frac{d(x-\overline{x})}{dt} = \frac{dx}{dt} - \frac{d\overline{x}}{dt} = f(x,u) - f(\overline{x},\overline{u}) = f(\Delta x, \Delta u) \tag{2.8}$$

If equation 2.8 is combined with equations 2.6 and 2.7, the linearized state and output equations around the linearization point $(\overline{x}, \overline{u})$ are obtained:

$$\frac{d\Delta x}{dt} \approx \frac{\partial f}{\partial x}\bigg]_{x=\overline{x},u=\overline{u}} \Delta x + \frac{\partial f}{\partial u}\bigg]_{x=\overline{x},u=\overline{u}} \Delta u = A\Delta x + B\Delta u \tag{2.9}$$

$$\Delta y \approx \frac{\partial g}{\partial x}\bigg]_{x=\overline{x},u=\overline{u}} \Delta x + \frac{\partial g}{\partial u}\bigg]_{x=\overline{x},u=\overline{u}} \Delta u = C\Delta x + D\Delta u \tag{2.10}$$

As stated above, for a linear (or linearized) model, the Jacobian matrices are named $A, B\ C$, and D, where $A = \frac{\partial f}{\partial x}$, $B = \frac{\partial f}{\partial u}$, $C = \frac{\partial g}{\partial x}$, and $D = \frac{\partial g}{\partial u}$. If $f(x,u)$ has m equations, and the number of states equals n, the Jacobian matrix A is an $m \times n$ matrix:

$$A = \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{2.11}$$

The A matrix is of special importance to the system because several characteristics of the system behaviour depend on it, such as the stability of the system as well as controllability and observability.

System stability is defined by the eigenvalues of the A matrix. When looking at the area around a linearization point or a steady state value, the eigenvalues of A give information about the system behaviour in that area. For instance, if the eigenvalues all have a negative real part, the system is stable around the chosen point. This means that small state deviations tend to converge to zero. On the other hand, if any eigenvalue has a positive real part the system is unstable and small state deviations do not go to zero.

2.4.3   Controllability and Observability

Controllability of the system is a crucial characteristic for the design of controllers for dynamic systems. The concept of controllability can be defined as follows. A system is controllable if it is possible to reach any state $x(t)$ from initial state $x(0)$ within a finite timespan, using an unconstrained control vector (Ogata, 2010). However, this definition is quite difficult to test in practice, which is why a simpler test has been introduced. This test defines the controllability matrix, $\mathcal{R}$, as follows:

$$\mathcal{R} = \begin{bmatrix} B & AB & A^2B & \cdots & A^{n-1}B \end{bmatrix} \tag{2.12}$$

The controllability matrix has $n \times nm$ dimensions, where n is the amount of states and m is the amount of inputs. The controllability of the system is thus dependent on the Jacobian matrices A and B. The actual test to check the controllability is done by checking the row rank of matrix $\mathcal{R}$. If $\mathcal{R}$ has full row rank (rank($\mathcal{R}$)=n), the system is controllable. Full row rank means that all rows of the matrix are linearly independent.

Another crucial characteristic for a system is its observability. A system is observable if it is possible to determine the state $x(t)$ at time t from the observation of the corresponding output within a finite timespan (Ogata, 2010). As with the official controllability definition, this is difficult to assess in practice, which is why another simpler test is introduced here. This test uses the observability matrix, $\mathcal{O}$, to determine the observability.

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix} \tag{2.13}$$

The observability matrix has $pn \times n$ dimensions, where n is the amount of states and p is the number of outputs. The observability of the system is dependent on the Jacobian matrices A and C. The actual test to check the observability is similar to the controllability test, which is checking the column rank of matrix $\mathcal{O}$. If $\mathcal{O}$ has full column rank (rank($\mathcal{O}$)=n), the system is observable.

2.4.4 Optimal Control

The concept of optimal control is centred around supplying an optimal control input u^* to your model that results in the desired (optimal) system behaviour x^* (Figure 5). The behaviour of the desired system depends on the aim of the controller. The aim consists of several control objectives, which can be translated to a quantitative cost function (J) that depends on the control $u(t)$. The goal of optimal control is to compute an optimal control input $u^*(t)$ that minimizes this quantitative cost function, resulting in the optimal system dynamics.
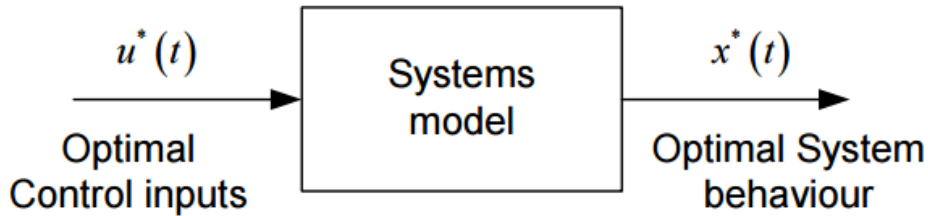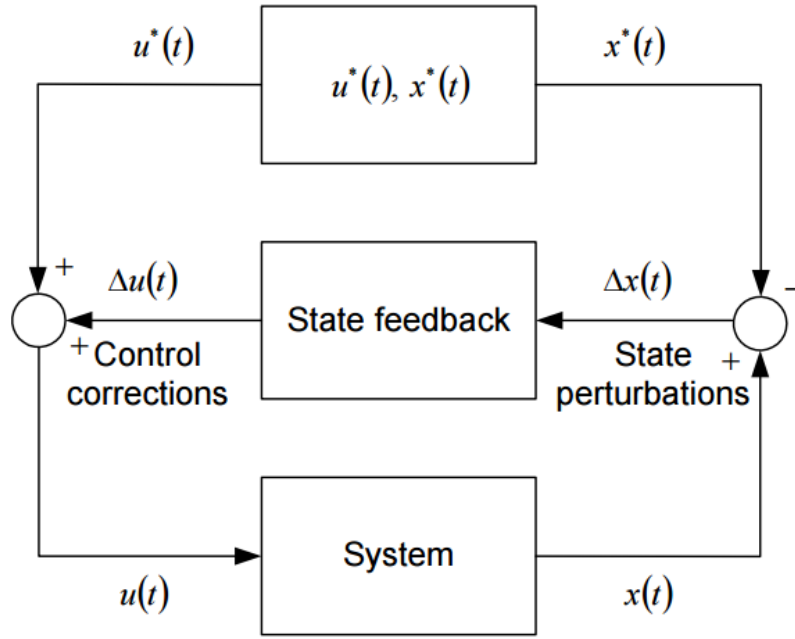

Figure 5. Schematic overview of optimal control. Taken from Willigenburg (2015).

The optimal input will only result in the optimal system behaviour if the systems model is completely accurate. In practice this will almost never happen, which means that the system behaviour contains deviations ($\Delta x(t)$) from the optimal behaviour. By performing measurements of the states of the system, these deviations can be calculated and the input to the system can be adjusted accordingly. This method is called feedback, of which two types exist. The first requires all state variables to be measured and the data is used to calculate control corrections $\Delta u(t)$ from the state deviations $\Delta x(t)$. This is called state feedback. The second type, output feedback, uses any outputs that are available, which is convenient if it is impossible to measure all states of a system. An entire optimal control scheme with state feedback can then be shown as in Figure 6.

Figure 6. Schematic overview of optimal control with state feedback. The optimal input $u^*(t)$ and optimal state $x^*(t)$ can be calculated offline in advance. The bottom block corresponds with the scheme shown in Figure 5. Taken from Willigenburg (2015).

An advantage of optimal control over other control methods is that the calculations of the optimal input and states can be done offline, instead of when the experiment has started. This saves computation time when the control system is running. Another advantage is that the use of a cost function allows the aim of the control to be quantitatively formulated.

The cost function $J(u(t))$ is divided into two parts, the terminal costs and the running costs. The terminal costs are often a function of the states at the final time of the experiment/simulation, and are depicted as $\phi(x(t_f))$. The running costs are mostly integrations over time, and depend on the states, control inputs, or external inputs ($d(t)$). They are often depicted as $\int_{t_0}^{t_f} L(x, u, d)$. Both terms are preceded by a weight factor α. The entire cost function that has to be minimized is then written as:

$$J(u(t)) = \alpha_1 \phi\left(x(t_f)\right) + \alpha_2 \int_{t_0}^{t_f} L(x, u, d)\, dt \qquad (2.14)$$

The actual computation of $u^*(t)$ in this thesis is done in Matlab, using an optimal control software package called PROPT (Rutquist & Edvall, 2008). PROPT is capable of solving highly complex optimal control problems, using a collocation method with Gauss or Chebyshev points. The exact mechanism of this program will not be discussed.

3   Experiment Setup

This chapter contains the materials and methods of the experiment that was performed in this thesis. Tubers from two different cultivars were assessed for their protein and amino acid content. Table 2 contains a small overview of the methods used for each determination.

Table 2. Overview of methods used in the determination of different compounds in potato tubers. PI=Protease Inhibitors.

Determination	Method	Original Protocol
Total protein content	Coomassie Blue dye-binding assay	Bradford (1996)
Patatin and PI fractions	Gel-filtration column	Brierly *et al*. (1996)
Amino acid content	Reverse-phase HPLC with *o*-phthalaldehyde derivation	Hanczkó *et al*. (2007)

3.1   Tuber Storage and Preparation

Potato (*Solanum tuberosum*) tubers of two different cultivars (Agria and Miss Malina) were obtained from storage facilities at local farmers. All tubers were stored at 6-15 °C and 90-95% relative humidity. Each week, 5 tubers from each cultivar were taken from storage, peeled and chopped, and pieces of several tubers with a combined weight of 200 grams were juiced. This resulted in 200 ml potato juice, after which 300 ml water was used to wash the food processor of any leftover proteins or sugars. Of the resulting potato juice, 2 ml was frozen at -20 °C and stored until protein determination. This procedure resulted in approximately 0.4 ml potato juice per gram fresh weight of the tubers, which was used for the further calculations of protein levels.

3.2   Protein Determination

Total soluble protein of the stored potato tubers was determined by an adaptation of the Coomassie Blue dye binding assay of Bradford (1976) with bovine serum albumin (Sigma-Aldrich) as a standard. The samples were diluted twice in Tris-HCl pH 7.0 containing 0.1 mM dithiothreitol (hereafter referred to as Tris buffer), so the final concentration was in the range of the protein standards. Then, the diluted protein samples were filtered through a 0.45 μm pore size membrane. The assay was done by adding 1.5 ml of Bradford Reagent (Sigma-Aldrich) to 0.05 ml of sample. After 25 minutes, the absorbance at 595 nm was measured using a spectrophotometer and compared to the absorbance of the protein standard. The change in colour with increasing protein concentrations is shown in Figure 7.
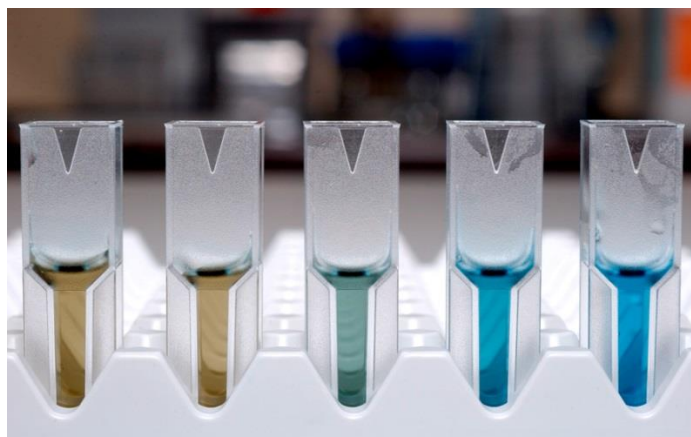
Figure 7. Colour progression of Bradford assay with increasing protein concentrations from left to right. Taken from Neuroscience Core Facility (2016).

The fractions of patatin and protease inhibitors were determined by gel filtration as done before by Brierley *et al.* (1996). Protein samples were diluted five times, again in Tris buffer, and filtered through a 0.2 µm pore size membrane. A Biosep SEC-s2000 gel-filtration column (Phenomenex) was used to separate protein fractions, the patatin fraction was identified by comparison with a 45 kDa glycoprotein standard (Phenomenex), and the area of the patatin peak was compared with the total peak area to give the proportion of patatin. Figure 8 shows an example chromatogram, with protein number 4 being the same size as patatin. The protease inhibitor fraction was identified by comparison with a 15 kDa protein standard, and the two peaks around this molecular weight were taken as being 15-21 kDa. The area of these peaks was then compared with the total peak area to give the proportion of protease inhibitors.
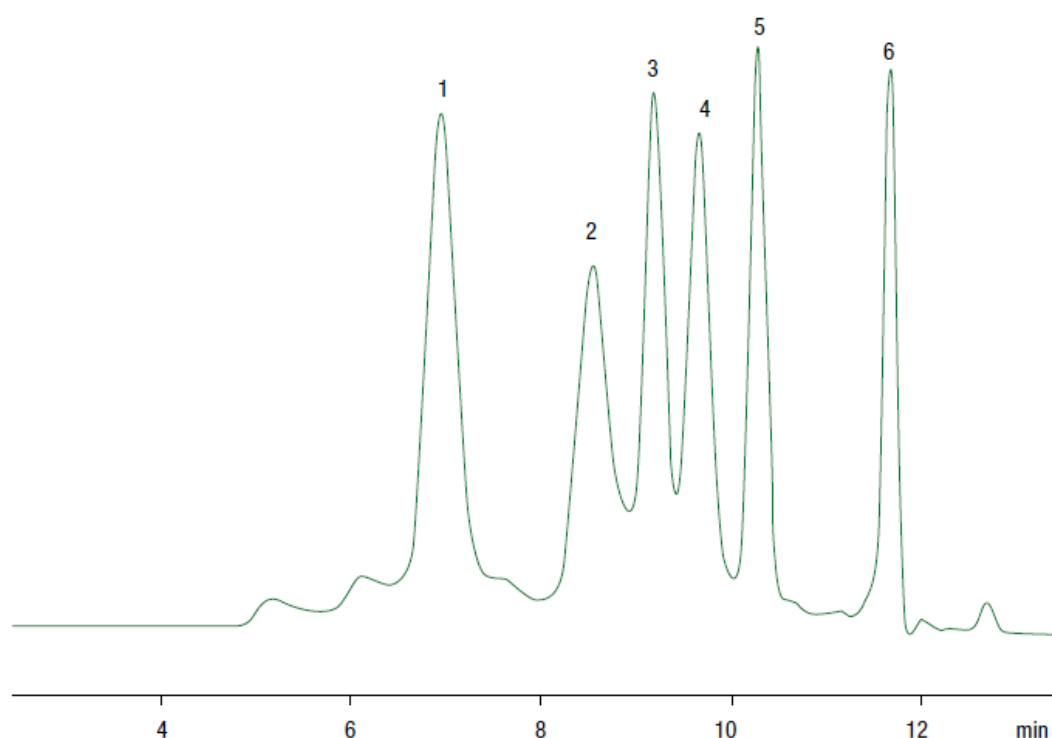
Figure 8. Example chromatogram of protein mixture separated by gel-filtration. Proteins: 1. Thyroglobulin 669 kDa, 2. IgG 156 kDa, 3. BSA 66 kDa, 4. Ovalbumin 45 kDa, 5. Myoglobin 16.9 kDa, 6. Uridine 244 Da. Adapted from Phenomenex (2016).

3.3   Amino Acid Determination

The individual amino acid concentrations were determined using reverse-phase HPLC (Thermo Fisher, Dionex). Each sample was diluted 40x in Tris buffer, and 0.1 ml norleucine (0.4 mM) was added as an internal standard to 0.5 ml protein sample. Finally, 0.4 ml methanol was added to precipitate proteins. After centrifugation, the supernatant was filtered through a 0.2 µm pore size membrane. The samples were then derivatised according to the method of Hanczkó *et al.* (2007), using *o*-phthalaldehyde-ethanethiol. Separation of the amino acids was done by gradient elution using a 2.1x150 mm Acquity UPLC BEH C18 column and a 2.1x5 mm guard column.

4   Experiment Results

This chapter will contain the results of the experiment described in chapter 3. First, the result of the protein determinations will be shown, after which the amino acid determination results will be assessed.

4.1   Protein Content

Different cultivars show different trends regarding both the total soluble protein as well as the individual fractions. The protein content of the Miss Malina (MM) cultivar seems to decline steadily, with a small peak after 10-12 weeks (Figure 10A). Both the patatin and the protease inhibitor fraction exhibited similar behaviour as the total protein (Figure 10B+C). This can be explained by the fact that the individual fractions stay fairly constant (Figure 9). The fraction of patatin seems to decrease by a small amount, from a range of 16-19% to 15-17%, but otherwise stays fairly constant. The fraction of protease inhibitors stays within a narrow range, between 32.5% and 37%.
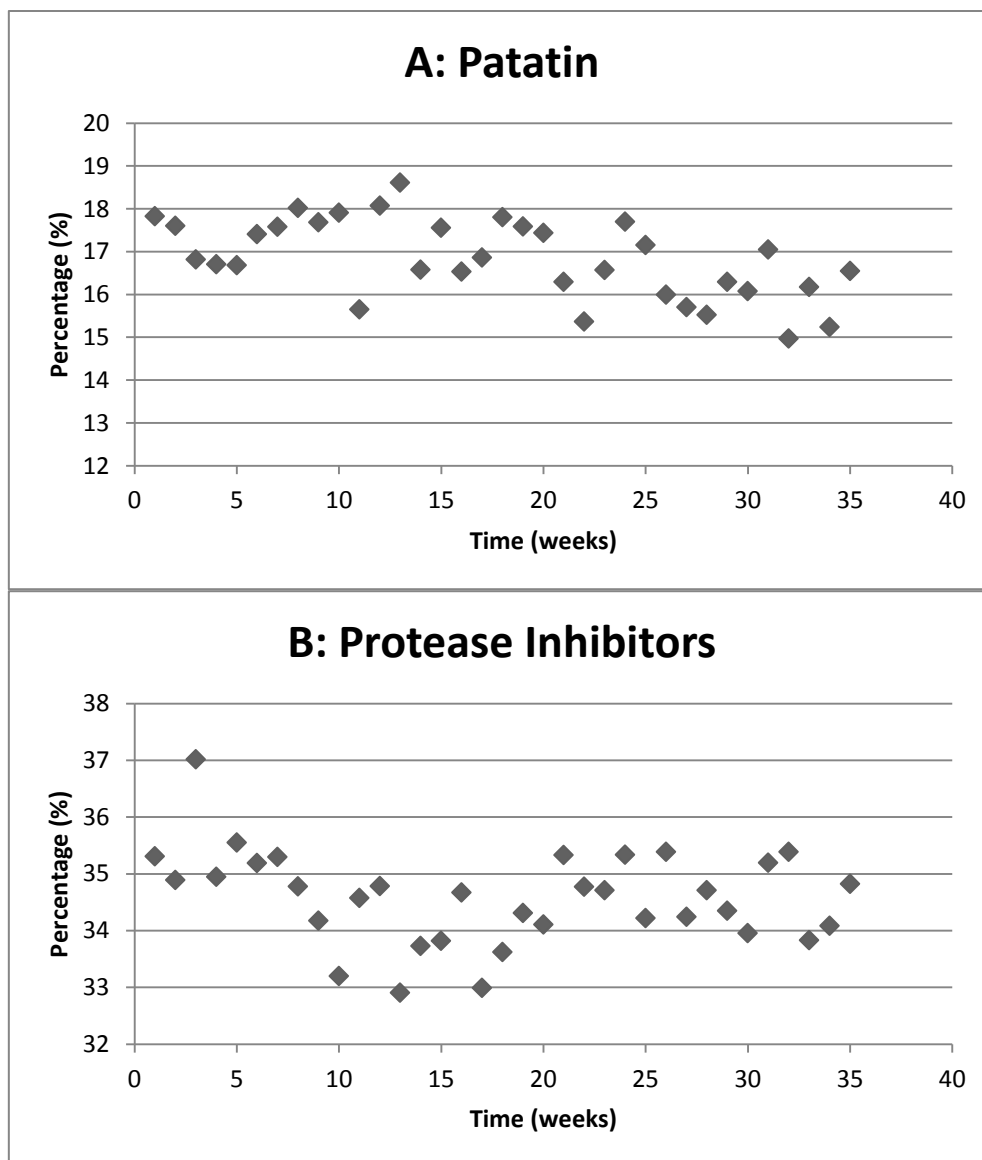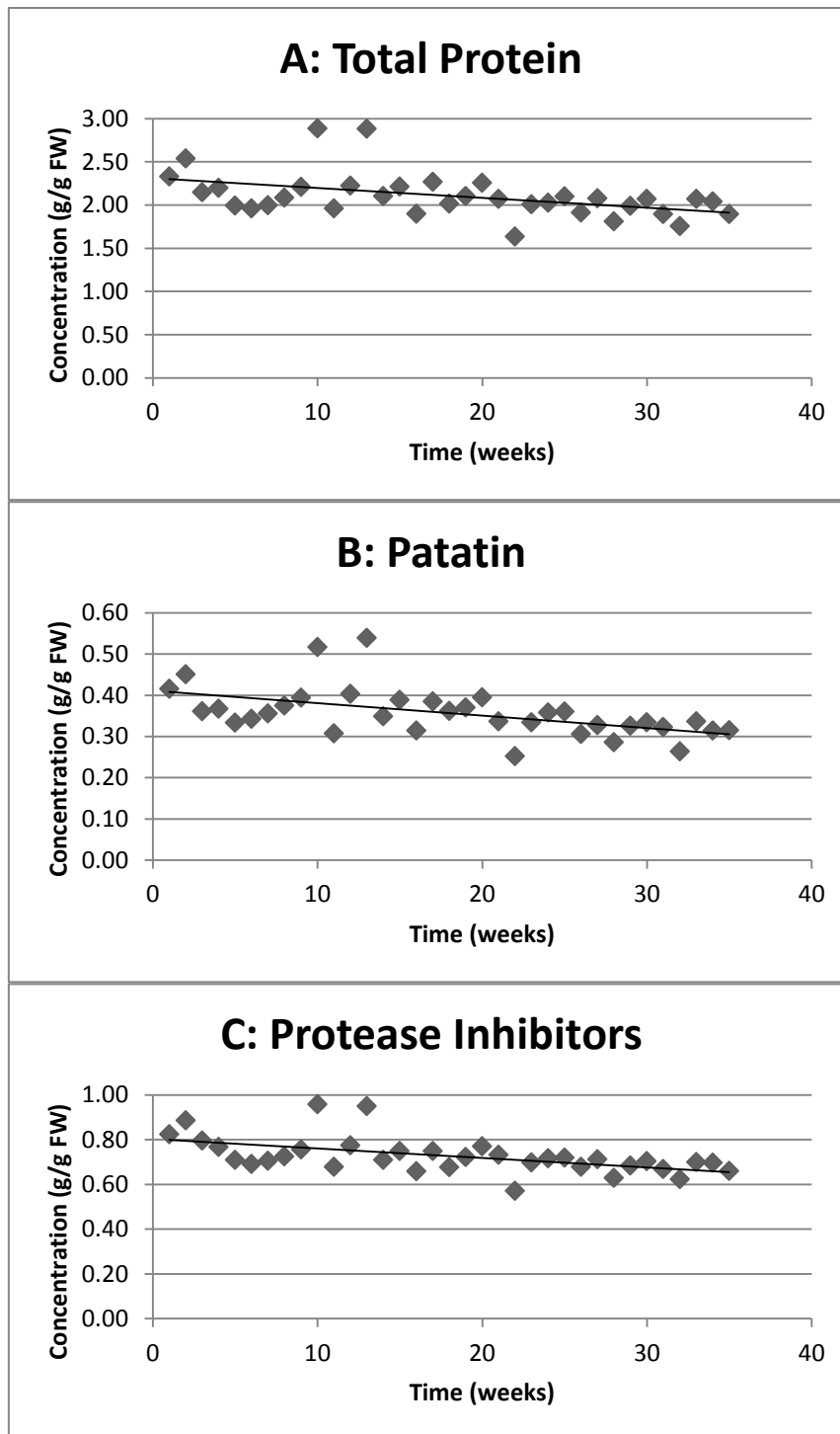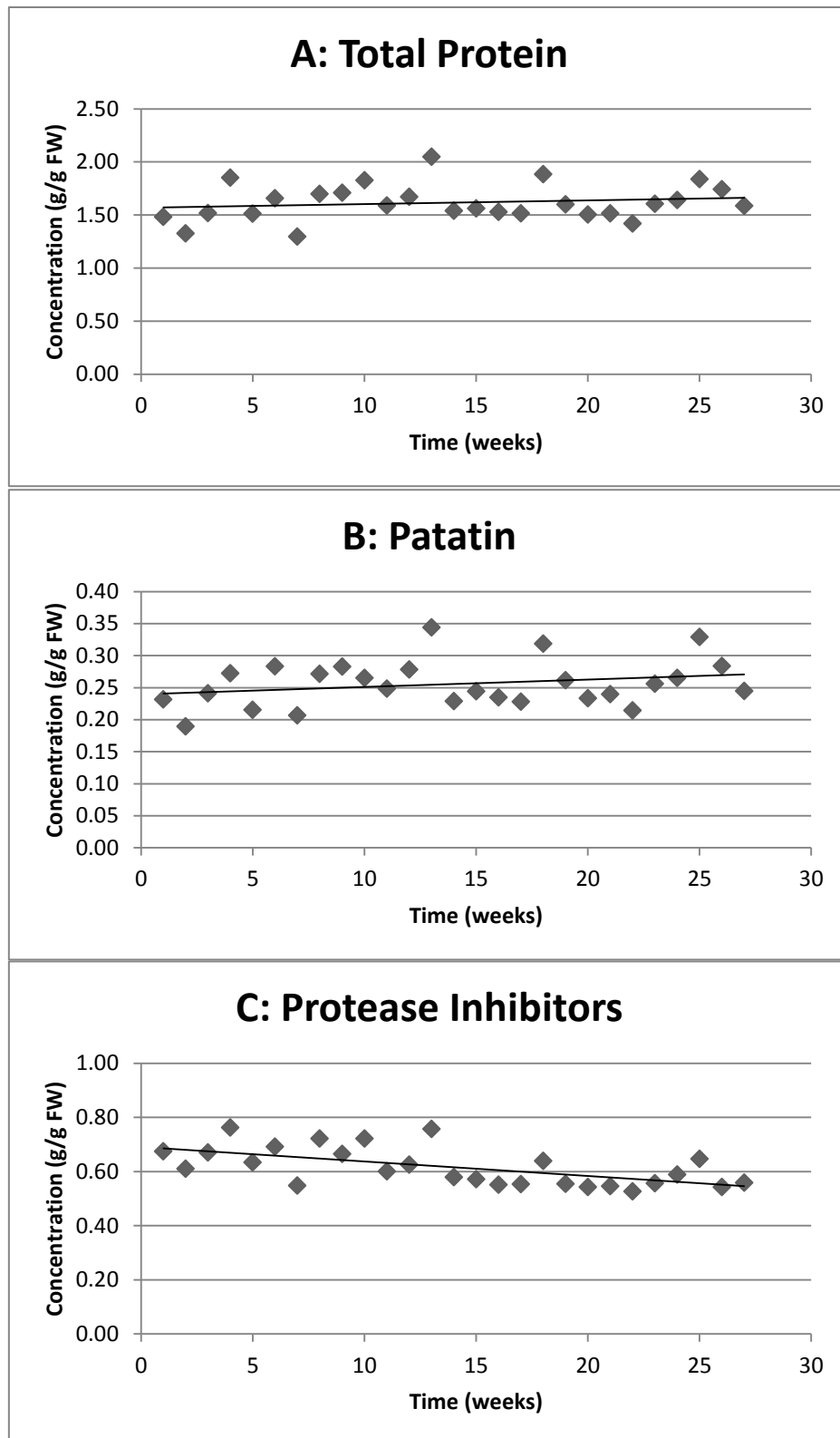

Figure 9. Percentages of total protein content of patatin (A) and protease inhibitors (B) of Miss Malina cultivar.

Figure 10. Protein dynamics of Miss Malina cultivar. Rough trend lines ($R^2 < 0.3$) drawn in black. Concentrations shown in gram protein per gram fresh potato weight.

On the other hand, the total protein level of the Agria cultivar does not seem to have a clear structure (Figure 11A). When attributing a rough linear trend to this data, the protein content seems to stay fairly constant over time. The patatin fraction shows a similar trend. However, the protease inhibitor content seems to drop over time (Figure 11C).

Figure 11. Protein dynamics of Agria cultivar. Rough trend lines ($R^2 < 0.4$) drawn in black. Concentrations shown in gram protein per gram fresh potato weight.

The decrease of protease inhibitor content becomes especially clear when plotting the protease inhibitors as a fraction of the total protein, showing a decline from approximately 47% to around 35% (Figure 12B). The reason for this decline of the fraction of protease inhibitors is unknown.

As with the Miss Malina cultivar, the patatin fraction stays reasonably constant in the Agria potatoes. A slight increase, from a range of 14-16% to 15-18%, can be observed (Figure 12A).
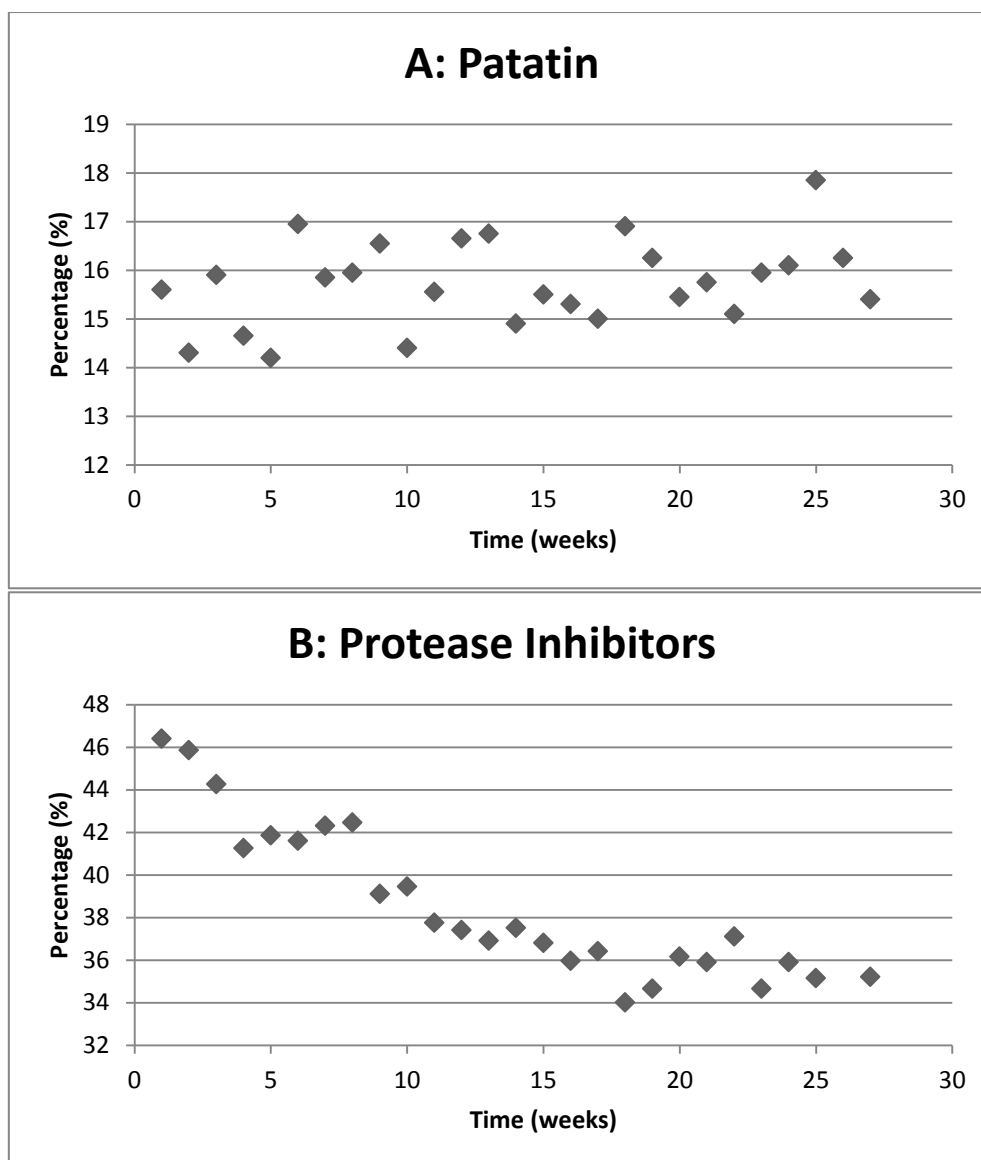

Figure 12. Percentages of total protein content of patatin (A) and protease inhibitors (B) of Agria cultivar.

4.2   Amino Acid Content

Figure 13 shows a chromatogram of the HPLC amino acid determination. The top diagram shows the chromatogram of a sample containing several amino acid standards, while the bottom diagram shows one of the potato juice samples. As can be seen, a single high peak of an unknown substance masks the presence of the other amino acids. This substance was present in all samples that were analysed. The origin of this substance is currently still unknown, and it is beyond the scope and timeframe of this research to find out what it is. Due to this, the amino acid data was regrettably not usable in this thesis.
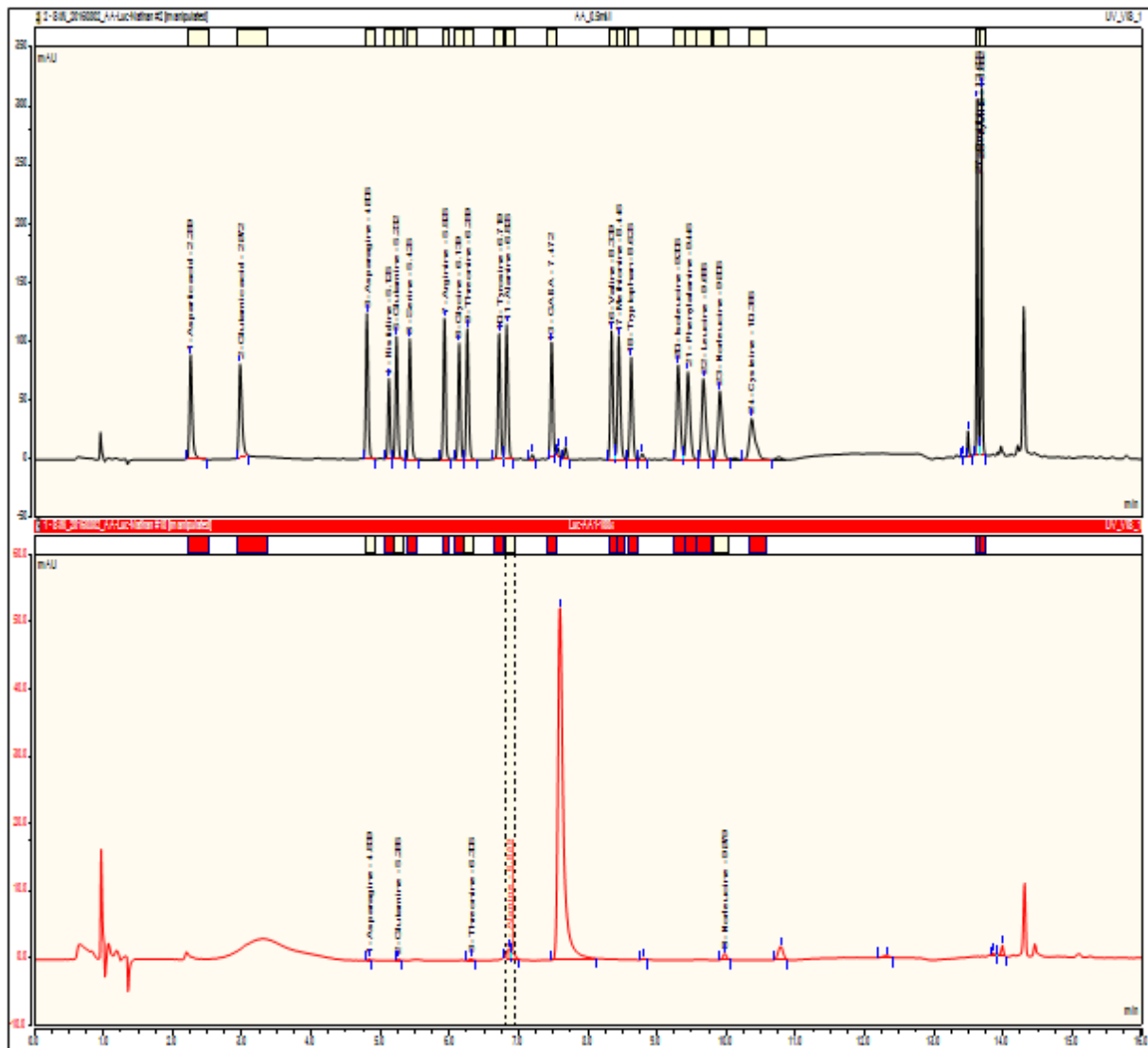
Figure 13. Chromatogram of amino acid determination. Top diagram shows amino acid standards; bottom diagram shows a Miss Malina sample.

4.3   Storage Temperature

The temperature inside the storage facility of both the Miss Malina and Agria potatoes was measured using temperature sensors. Figure 14 shows these measurements over the entire storage period. The temperature inside the storage facilities fluctuates between 6°C and 16°C, showing a clear increase when approaching summer for the Miss Malina cultivar. The Agria tubers were harvested in April, before the increase in temperature.
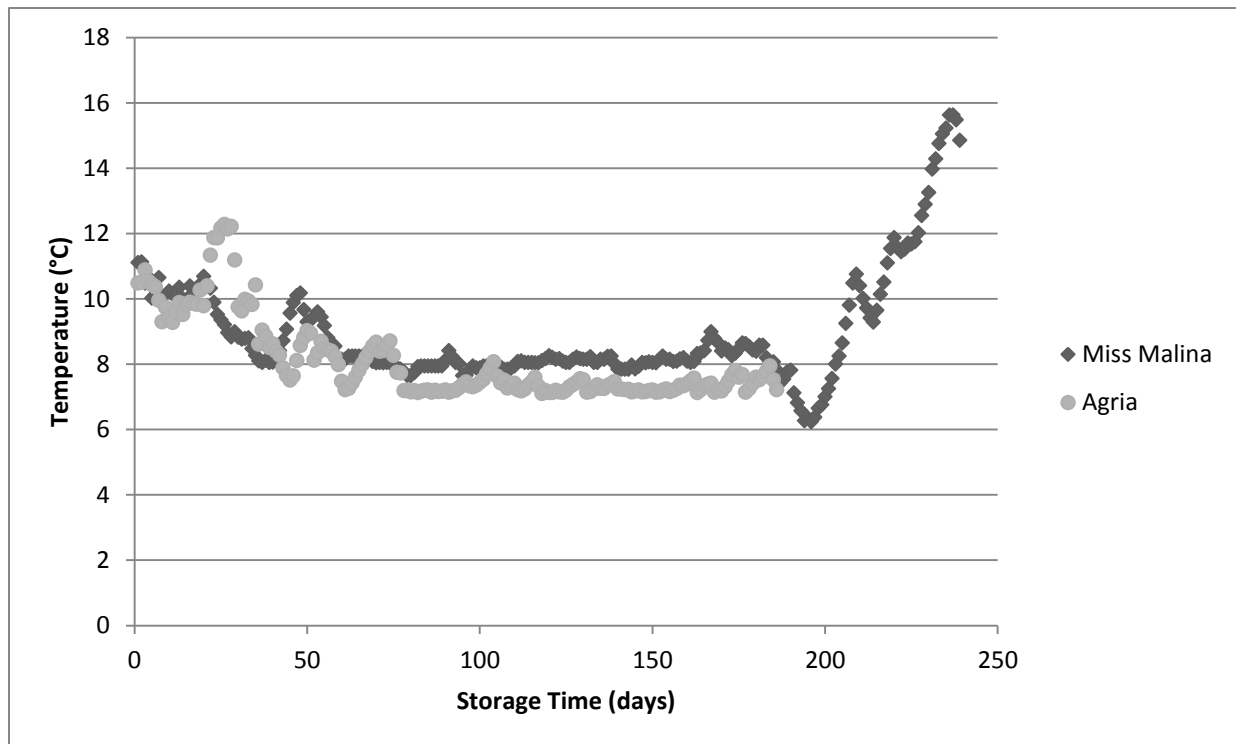
Figure 14. Storage temperature of Agria and Miss Malina potato tubers over 2015-2016. Starting date at $t = 0$ is 16 October 2015.

5   Modelling Protein Dynamics

Modelling of the protein dynamics leads to a complete protein turnover model, which was tested using the data obtained in the experiments. This section contains the steps that were made during the modelling, as well as any assumptions.

5.1   General Assumptions

Several assumptions were made for the initial model that will be introduced in what follows. Additional assumptions that were made for the more complicated models will be explained in subsequent paragraphs.

First of all, a uniform distribution of protein throughout the potato tuber was assumed. The concentration of each compound is therefore not dependent on the location in the tuber. Also, the potato is considered a closed system that does not interact with the environment. This means that no excretion or uptake of protein and amino acids takes place. Consequently, the rates of change for each compound are the sums of their synthesis and degradation rates and do not depend on migration of compounds. The temperature inside the potato is also assumed to be evenly distributed, and is the same as the temperature inside the storage facility.

Furthermore, patatin is not synthesized during storage (Paiva *et al.*, 1983). The rate of change of patatin is thus only dependent on its degradation. Both protein synthesis rate and degradation rate are assumed to follow Michaelis-Menten enzyme kinetics and are temperature-dependent, as described in section 2.4.1.

To summarise the assumptions that were made are:

- Uniform distribution of both temperature and molecules in each potato tuber
- No uptake or excretion of protein or amino acids (no migration)
- No new synthesis of patatin
- Michaelis-Menten kinetics for both synthesis and degradation
- The maximum reaction rates V_{MAX} are dependent upon temperature according to the Arrhenius equation

5.2   Initial Model

Taking into account all assumptions mentioned in section 5.1, the initial model was constructed. For each component an ordinary differential equation was formulated, containing a rate of synthesis and a rate of degradation following the kinetics discussed earlier. Together these form the rate of change of the component over time.

$$\frac{d[TP](t)}{dt} = Y_{TP/AA} * A_{PS}e^{-\frac{E_{PS}}{RT}} * \frac{[AA]}{K_{PS}+[AA]} - A_{PD}e^{-\frac{E_{PD}}{RT}} * \frac{[TP]}{K_{PD}+[TP]} \tag{5.3}$$

$$\frac{d[Pat](t)}{dt} = -A_{PD}e^{-\frac{E_{PD}}{RT}} * \frac{[Pat]}{K_{PD}+[Pat]} \tag{5.4}$$

$$\frac{d[AA](t)}{dt} = Y_{AA/TP} * A_{PD}e^{-\frac{E_{PD}}{RT}} * \frac{[TP]}{K_{PD}+[TP]} - A_{PS}e^{-\frac{E_{PS}}{RT}} * \frac{[AA]}{K_{PS}+[AA]} \tag{5.5}$$

Here, each positive term stands for the synthesis rate of a compound (protein or amino acids), and each negative term stands for the degradation rate of the compound. This model assumes that both protein synthesis and its degradation follow first-order kinetics, as has been seen in photosynthetic algae (Lancelot *et al.,* 1986).

The total soluble protein concentration is shown as $[TP]$, the amino acid content is shown as $[AA]$, and patatin concentration is $[Pat]$. The suffixes PS and PD mean protein synthesis and protein degradation, respectively. A yield factor $Y_{AA/TP}$ is added to reflect the amount of amino acids that are formed during the degradation of a gram of protein. On the other hand, the yield factor $Y_{TP/AA}$ reflects how many proteins can be made per gram of amino acids. It was assumed that these factors are inversely correlated as follows: $Y_{TP/AA} = \frac{1}{Y_{AA/TP}}$.

When validating the initial model with the obtained data, it seemed to be able to describe the data fairly well. However, because of the lack of data of the amino acids, the model was simplified by removing the amino acids' differential equation. To do this, the assumption has to be made that new protein synthesis is a zero-order process, such that: $Y_{TP/AA} * \frac{[AA]}{K_{PS}+[AA]} = 1$. Eliminating this term in turn meant that evidence had to be found that synthesis can in fact be independent of the amino acid concentration, which was found to be fairly common for total protein synthesis in animals (Schimke & Doyle, 1970; Hargrove, 1993) and for individual proteins in plants (Huffaker, 1974; Jones, 1984). Furthermore, Trewavas (1972) found that protein synthesis in common duckweed was usually a zero-order process, but first-order in exponentially growing tissues. Taking all this into account, it was decided to leave out the amino acids from the model, and perhaps later see whether the amino acids state could be controlled and observed when it was added. The removal of the amino acids resulted in the following model:

$$\frac{d[TP](t)}{dt} = A_{PS}e^{-\frac{E_{PS}}{RT}} - A_{PD}e^{-\frac{E_{PD}}{RT}} * \frac{[TP]}{K_{PD}+[TP]} \tag{5.6}$$

$$\frac{d[Pat](t)}{dt} = -A_{PD}e^{-\frac{E_{PD}}{RT}} * \frac{[Pat]}{K_{PD}+[Pat]} \tag{5.7}$$

Here, only two states are left, TP and Pat.

5.3   Inhibition Model

The model can be extended by including the inhibition of protein degradation by protease inhibitors (PI), which are common molecules in potato tubers (see section 2.1.2 of this report). This will add an extra state, and will slightly change the degradation terms of the model:

$$\frac{d[TP](t)}{dt} = A_{PS}e^{-\frac{E_{PS}}{RT}} - A_{PD}e^{-\frac{E_{PD}}{RT}} * \frac{[TP]}{K_{PD}*\left(1+\frac{[PI]}{K_i}\right)+[TP]} \tag{5.8}$$

$$\frac{d[PI](t)}{dt} = A_{PI}e^{-\frac{E_{PI}}{RT}} - A_{PD}e^{-\frac{E_{PD}}{RT}} * \frac{[PI]}{K_{PD}*\left(1+\frac{[PI]}{K_i}\right)+[PI]} \tag{5.9}$$

$$\frac{d[Pat](t)}{dt} = -A_{PD}e^{-\frac{E_{PD}}{RT}} * \frac{[Pat]}{K_{PD}*\left(1+\frac{[PI]}{K_i}\right)+[Pat]} \tag{5.10}$$

With K_i being the inhibitor's dissociation constant.

An assumption that was made here is that each compound degrades at the same rate. This means that the degradation terms in each differential equation are the same. On the other hand, the synthesis of individual compounds is assumed to occur at a different rate than the total protein synthesis, and as was mentioned earlier, patatin is not newly synthesized during storage at all. Therefore, the parameters A_{PI} and E_{PI} are added to define protease inhibitor synthesis. The term $\left(1 + \frac{[PI]}{K_i}\right)$ appears in the degradation terms to incorporate the competitive inhibition of proteases by the protease inhibitors, which is used before with Michaelis-Menten models with inhibition (Nxumalo *et al.*, 1998). As competitive inhibition is the most common among the mechanisms of protease inhibitors in plants (Ryan, 1990), it was assumed as the mechanism of action here.

Finally, writing the model in a state-space form by defining all states, inputs, and parameters results in the following model:

$$\frac{dx_1(t)}{dt} = p_1 e^{-\frac{p_6}{p_9 u_1}} - p_2 e^{-\frac{p_7}{p_9 u_1}} * \frac{x_1}{p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_1} \tag{5.11}$$

$$\frac{dx_2(t)}{dt} = p_3 e^{-\frac{p_8}{p_9 u_1}} - p_2 e^{-\frac{p_7}{p_9 u_1}} * \frac{x_2}{p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_2} \tag{5.12}$$

$$\frac{dx_3(t)}{dt} = -p_2 e^{-\frac{p_7}{p_9 u_1}} * \frac{x_3}{p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_3} \tag{5.13}$$

These three differential equations together form the state vector equation $f(x, u, p) = \frac{dx}{dt}$ with $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, $u = u_1$, and $p = \begin{bmatrix} p_1 \\ \dots \\ p_9 \end{bmatrix}$. The original notation, the state-space notation, a short description, and the units of each variable and parameter are shown in Table 3.

Table 3. State-space notation of variables and parameters.

Original notation	System notation	Variable description
$[TP]$	x_1	Total protein concentration [mg/g fresh weight]
$[PI]$	x_2	Protease inhibitor concentration [mg/g fresh weight]
$[Pat]$	x_3	Patatin concentration [mg/g fresh weight]
T	u_1	Storage temperature [Kelvin]
A_{PS}	p_1	Arrhenius constant for total protein synthesis [day^{-1}]
A_{PD}	p_2	Arrhenius constant for protein degradation [day^{-1}]
A_{PI}	p_3	Arrhenius constant for protease inhibitor synthesis [day^{-1}]
K_{PD}	p_4	Michaelis constant for protein degradation [mg/g fresh weight]
K_i	p_5	Inhibitor's dissociation constant [mg/g fresh weight]
E_{PS}	p_6	Activation energy of total protein synthesis [$J * mol^{-1}$]
E_{PD}	P_7	Activation energy of protein degradation [$J * mol^{-1}$]
E_{PI}	p_8	Activation energy of protease inhibitor synthesis [$J * mol^{-1}$]
R	p_9	Universal gas constant [$8.31446\, J * K^{-1} * mol^{-1}$]

All three states were measured, so each state is also an output. The output equation is then simply:

$$y = g(x, u, p) = x \tag{5.14}$$

5.4 Linearized Model

As the model is not linear ($\frac{\partial f}{\partial x}$ is dependent on $x(t)$, see below), the linearization procedure described in section 2.4 was applied. Taking the partial derivatives of $f(x, u, p)$ to x and u in $\overline{x}$ and $\overline{u}$ results in the following A and B matrices:

$$A(:,1) = \left[\begin{array}{c} \dfrac{x_1 p_2 e^{-\frac{p_7}{p_9 u_1}}}{\left(p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_1\right)^2} - \dfrac{p_2 e^{-\frac{p_7}{p_9 u_1}}}{p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_1} \\ 0 \\ 0 \end{array} \right]_{x=\overline{x}}$$

$$A(:,2) = \left[\begin{array}{c} \dfrac{x_1 p_2 p_4 e^{-\frac{p_7}{p_9 u_1}}}{p_5 \left(p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_1\right)^2} \\ \dfrac{\left(\frac{p_4}{p_5} + 1\right) * x_2 p_2 e^{-\frac{p_7}{p_9 u_1}}}{\left(p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_2\right)^2} - \dfrac{p_2 e^{-\frac{p_7}{p_9 u_1}}}{p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_2} \\ \dfrac{x_3 p_2 p_4 e^{-\frac{p_7}{p_9 u_1}}}{p_5 \left(p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_3\right)^2} \end{array} \right]_{x=\overline{x}}$$

$$A(:,3) = \begin{bmatrix} 0 \\ 0 \\ \dfrac{x_3 p_2 e^{-\frac{p_7}{p_9 u_1}}}{\left(p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_3\right)^2} - \dfrac{p_2 e^{-\frac{p_7}{p_9 u_1}}}{p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_3} \end{bmatrix}_{x=\overline{x}}$$

Due to its large size, the A matrix has been split up in three separate columns, where $A(:,j)$ denotes the jth column of A.

$$B = \begin{bmatrix} \dfrac{p_2 p_6 e^{-\frac{p_6}{p_9 u_1}}}{(p_9 u_1)^2} - \dfrac{x_1 p_2 p_7 e^{-\frac{p_7}{p_9 u_1}}}{(p_9 u_1)^2 * \left(p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_1\right)} \\ \dfrac{p_3 p_8 e^{-\frac{p_8}{p_9 u_1}}}{(p_9 u_1)^2} - \dfrac{x_2 p_2 p_7 e^{-\frac{p_7}{p_9 u_1}}}{(p_9 u_1)^2 * \left(p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_2\right)} \\ - \dfrac{x_3 p_2 p_7 e^{-\frac{p_7}{p_9 u_1}}}{(p_9 u_1)^2 * \left(p_4 * \left(1 + \frac{x_2}{p_5}\right) + x_3\right)} \end{bmatrix}_{u=\overline{u}}$$

When taking the partial derivatives of $g(x, u, p)$ to x and u we obtain the C and D matrices:

$$C = \frac{\partial g}{\partial x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$D = \frac{\partial g}{\partial u} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

These matrices define the linearized model:

$$\frac{d\Delta x(t)}{dt} = A(t) * \Delta x(t) + B(t) * \Delta u(t) \tag{5.16}$$

$$\Delta y(t) = C * \Delta x(t) + D * \Delta u(t) \tag{5.17}$$

Where $\Delta x(t) = x(t) - \overline{x}(t)$, $\Delta u(t) = u(t) - \overline{u}(t)$, and $\Delta y(t) = y(t) - \overline{y}(t)$.

Because there is not a clear equilibrium or setpoint, the linearization cannot be done around a single point in state space. Therefore, the Jacobian matrices have to be calculated around a reference trajectory $(\overline{x}(t), \overline{u}(t))$, leading to time-varying matrices $A(t)$ and $B(t)$. The system is then simulated (using the current matrices) until the next time point, after which the linearization is repeated.

5.5   Model Validation

Both the non-linear and linearized model were validated using the data that was obtained from the protein measurements. A non-linear least squares method was used to calculate the unknown parameters of the system, after which the sensitivities of the system to each parameter were

assessed as well. The effect of different storage temperatures on the system was investigated by simulating the model response to different inputs.

5.5.1   Parameter Estimation

A parameter estimation procedure was performed using the Matlab function *lsqnonlin*. This function finds the minimum of the sum of squares of a function $f(p)$ by changing the values of the parameters p. The function $f(p)$ should then be written such that it returns the residuals between the simulated values and the measurements, after which *lsqnonlin* calculates the least-squares. As our model contains eight parameters, it is best to perform this procedure for a few parameters at a time. This improves the accuracy of the parameter estimation. Therefore, two parameters were estimated at a time while holding the other parameters constant. This was repeated several times to converge on the right solution. The resulting parameter values for both the non-linear case and the linearized case are shown in Table 4.

Table 4. Estimated parameter values for non-linear and linearized cases. Units for each value are the same as in Table 3.

Parameter	Non-linear (Miss Malina)	Linearized (Miss Malina)	Non-linear (Agria)	Linearized (Agria)
A_{PS}	$6.5171e+7$	$9.124e+6$	$1.2712e+8$	$2.6247e+7$
A_{PD}	$1.9894e+7$	$4.637e+6$	$2.8857e+7$	$3.077e+6$
A_{PI}	$9.4350e+6$	$9.285e+6$	$6.4700e+6$	$7.632e+6$
K_{PD}	$5.7012e+1$	$5.7756e+1$	$4.4434e+1$	$5.5725e+1$
K_i	1.1017	$6.610e-2$	$6.6123e-4$	$9.860e-2$
E_{PS}	$6.0612e+4$	$6.3742e+4$	$5.9737e+4$	$6.3946e+4$
E_{PD}	$4.4886e+4$	$4.4973e+4$	$3.3179e+4$	$4.6337e+4$
E_{PI}	$5.9582e+4$	$6.7394e+4$	$1.4623e+5$	$8.0575e+4$

The estimated parameter values are all in the same order of magnitude in each case, except for K_i. If this value becomes smaller, the influence of protease inhibitors on protein degradation becomes larger. Apparently, the inhibition of proteases in the non-linear model is stronger in the Agria cultivar than in the Miss Malina potatoes. However, when the system is linearized this difference is not present anymore.

The identifiability of the model parameters was then assessed. A model is called identifiable if it is possible to find the true values of its parameters by comparing the outcome of different parameter sets with observed data values. If multiple sets of parameters result in the same model output, the model is called unidentifiable. Each parameter set then provides a local minimum instead of a global minimum in the distribution of the sum of squares that was calculated earlier with *lsqnonlin*. Often global minima are difficult to find. Two conditions have to be fulfilled in order to arrive at a minimum:

$$\frac{\partial f(p)}{\partial p} = 0, \quad and \quad \frac{\partial^2 f(p)}{\partial p_i \partial p_j} > 0 \qquad (5.18)$$

The first derivatives define either a minimum or maximum, and the second derivatives define whether it is a minimum or maximum. The $p \times p$ matrix containing the first derivative is denoted by ϕ. This matrix can give some information about which parameters can be fixed, because small values correspond with little changes in the parameter estimates. The second condition of equation 5.18

can be assessed by a singular value decomposition (SVD) of ϕ. A short explanation of SVD is given below. Firstly, the matrix ϕ is decomposed into three matrices as follows:

$$\phi = USV'$$ (5.19)

The matrix of the most importance here is matrix S, which contains the singular values. If all these singular values are not equal to zero, the matrix $(\phi'\phi)$, as used in the least-squares algorithm, is invertible, and the system parameters are identifiable. The function of the matrices U and V will not be discussed here. Table 5 shows the singular values for all four scenarios.

Table 5. Singular values for non-linear and linearized cases. SV=singular values.

Non-linear SV (Miss Malina)	Linearized SV (Miss Malina)	Non-linear SV (Agria)	Linearized SV (Agria)
$6.2469e-1$	$1.2156e+1$	$6.2335e+2$	2.3623
$7.7704e-2$	$1.331e-1$	$1.0394e-1$	$1.1632e-1$
$1.7686e-2$	$2.8277e-2$	$2.2983e-4$	$3.3343e-3$
$3.1191e-3$	$8.8684e-3$	$2.1338e-5$	$1.2440e-4$
$2.2897e-4$	$3.5859e-4$	$2.9153e-6$	$6.8697e-5$
$2.4067e-5$	$4.5592e-5$	$1.0032e-8$	$4.4567e-5$
$5.5737e-6$	$1.0163e-5$	$1.339e-21$	$7.8151e-8$
$1.1989e-6$	$1.9616e-8$	0	$3.8738e-8$

Although only the third column of Table 5 contains a zero, the other columns contain values that are close to zero. This means that one or more of the parameters are not identifiable. When calculating the standard deviations of the parameter estimates (Table 6), this becomes even clearer. All standard deviations are several orders of magnitude larger than the corresponding estimated values, indicating very inaccurate estimates.

Table 6. Standard deviations for parameter estimates in Table 4. SD=Standard deviation. Units for each value are the same as in Table 3.

Parameter	Non-linear SD (Miss Malina)	Linearized SD (Miss Malina)	Linearized SD (Agria)
A_{PS}	$1.6030e+11$	$4.4236e+10$	$2.5552e+10$
A_{PD}	$3.8878e+10$	$4.5079e+11$	$4.8984e+10$
A_{PI}	$3.9906e+10$	$3.6512e+10$	$5.8386e+12$
K_{PD}	$9.1722e+4$	$5.6156e+6$	$9.8022e+5$
K_i	$1.4390e+2$	5.5636	$8.0741e+1$
E_{PS}	$5.7488e+6$	$1.3204e+7$	$2.4125e+6$
E_{PD}	$2.1676e+6$	$1.9294e+6$	$4.4081e+6$
E_{PI}	$9.9067e+6$	$9.7601e+6$	$1.8488e+9$

The implications of the unidentifiability of the system (represented by equations 5.8-5.10) will be discussed in section 7.2. The simulations and controller design were still performed with the set of parameters of Table 4, non-linear and linearized, respectively.

5.5.2   Simulations

Using the measured temperature as an input, and the nonlinear parameters that were estimated in the previous section, the nonlinear model responses were simulated. Figures 15 and 16 show the simulation results compared with the measurements obtained in the experiment for, respectively, the Miss Malina and Agria cultivar. It can be seen that the model responses are smooth instead of following the fluctuations in the data and provide reasonable fits.
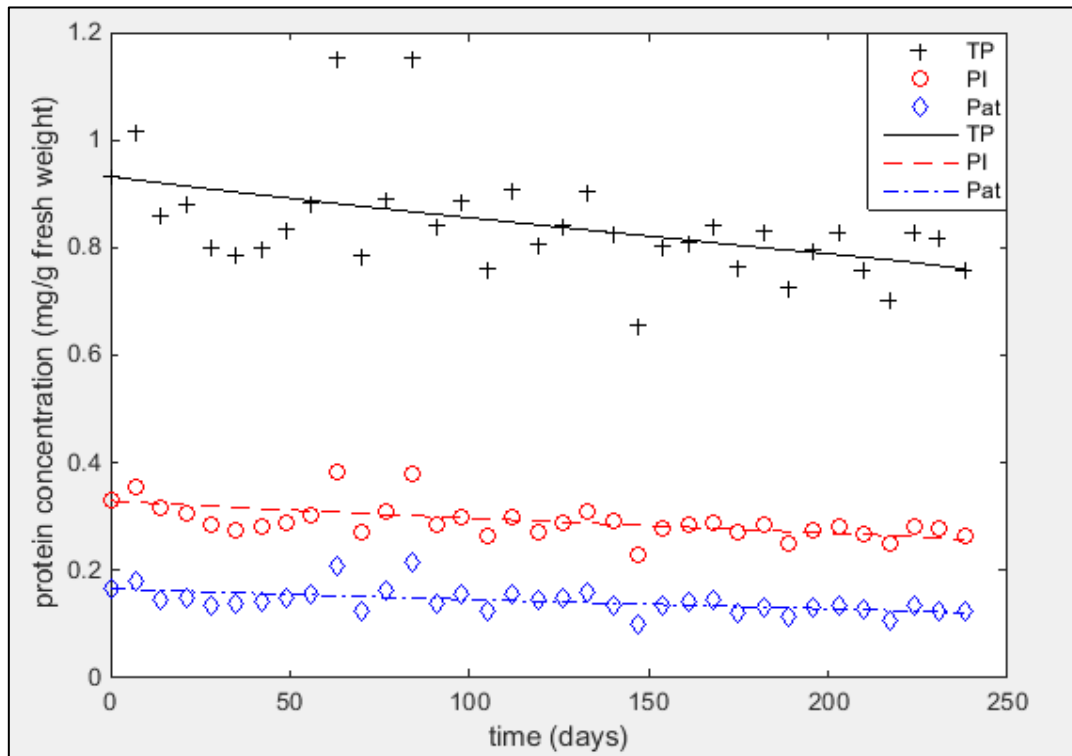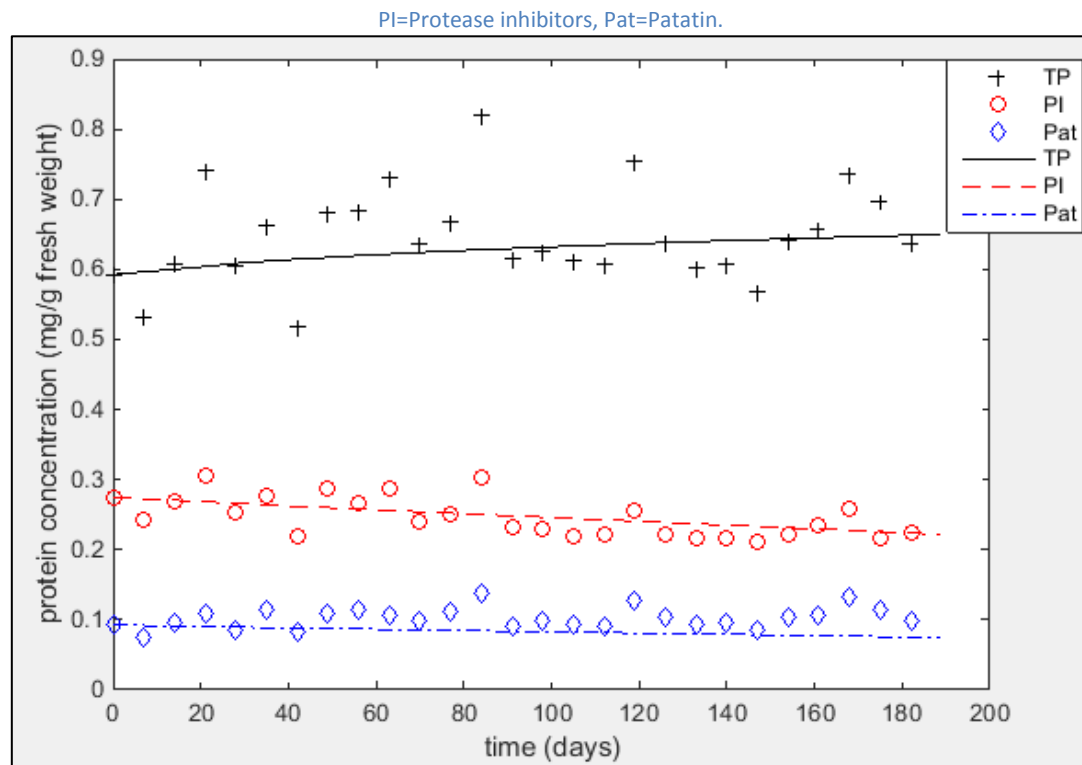

Figure 15. Simulated model response compared to experimental data for Miss Malina potatoes. Simulated responses are shown as (dashed or dot-dashed) lines, experimental data are shown as markers (+, o, or diamonds). TP=Total protein,

Figure 16. Simulated model response compared to experimental data for Agria potatoes. Simulated responses are shown as (dashed or dot-dashed) lines, experimental data are shown as markers (+, o, or diamonds). TP=Total protein, PI=Protease inhibitors, Pat=Patatin.

Also, the system response to different temperatures was simulated using the non-linear model. The real measured temperature was compared with a constant storage temperature of 0°C, 5°C, 15°C and 20°C. Figure 17 shows for each individual component its simulated response to different storage temperature for Miss Malina potatoes, and Figure 18 shows the same for the Agria cultivar.
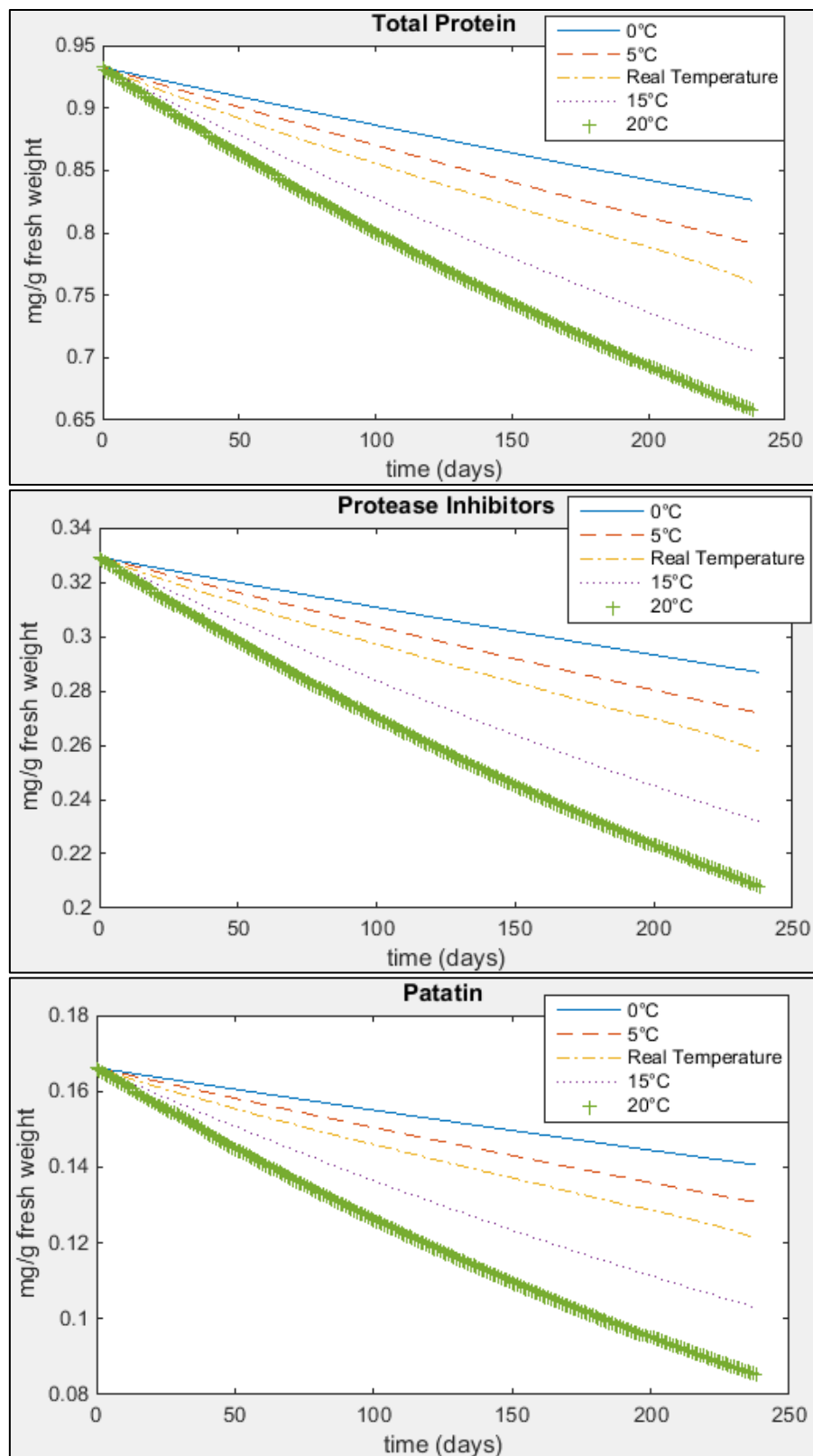
Figure 17. Simulated response of protein content in cultivar Miss Malina potatoes to different storage temperatures.
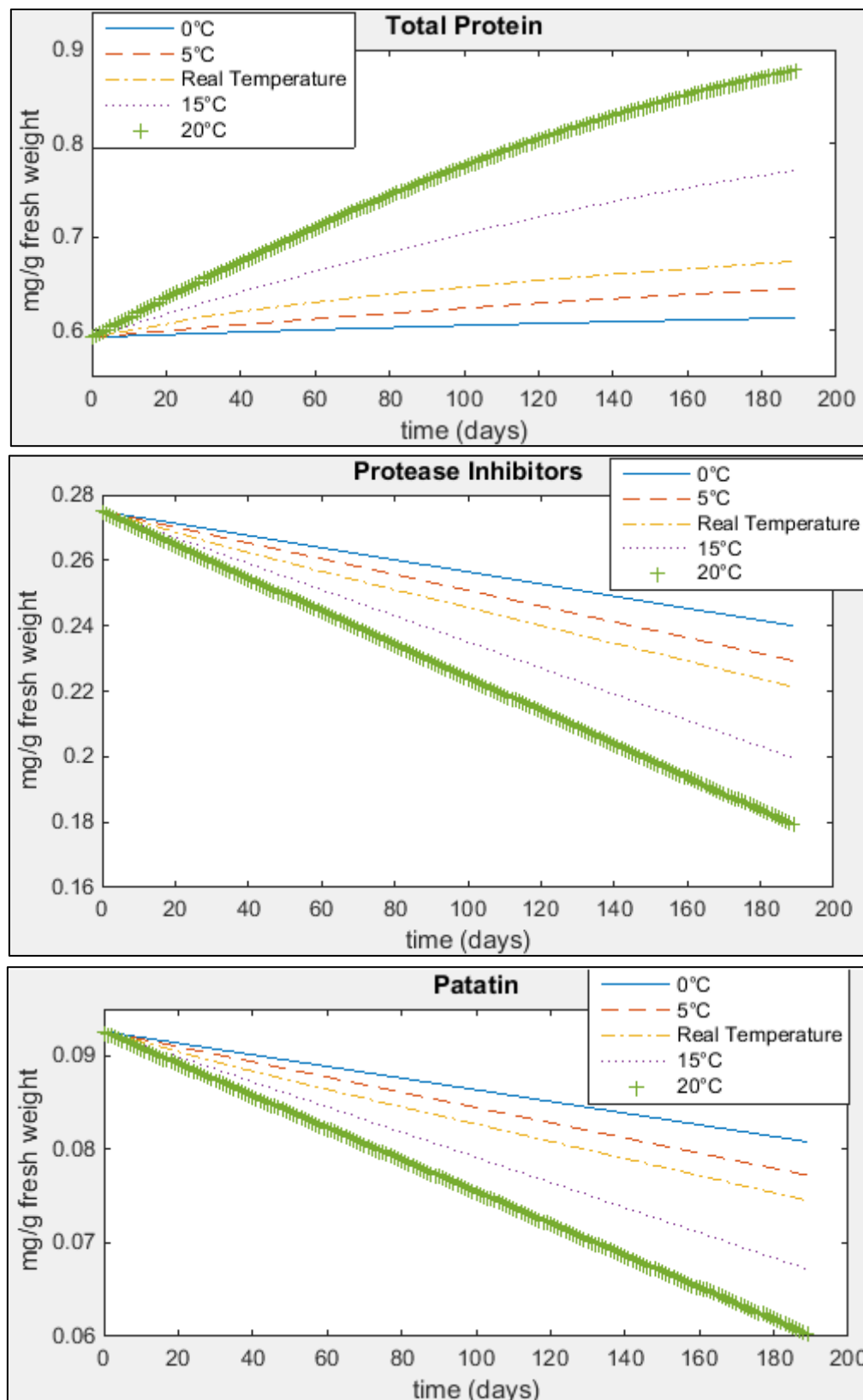
Figure 18. Simulated response of protein content in cultivar Agria potatoes to different storage temperatures.

For the Miss Malina cultivar, both the total protein content and the individual fractions behave similarly. Although increasing the temperature raises both the rate of protein synthesis and degradation, the influence of the temperature is greater on the degradation term than on the

synthesis term. From the model simulations it is concluded that this cultivar will therefore always have a net degradation of protein, and an increase in storage temperature will result in a lower protein content at the end of the storage period.

The potatoes of the Agria cultivar show a different model response to storage temperature. Especially the dynamics of the total protein content, which are the completely opposite of the dynamics of the Miss Malina cultivar. Increasing the storage temperature results in a higher protein content, but also in a lower protease inhibitor and patatin content. The influence of temperature on degradation is therefore still larger than the influence of temperature on protease inhibitor synthesis, but smaller than the influence of temperature on total protein synthesis. The Agria potatoes will always have a net total protein increase, but the individual fractions that were investigated here will drop over time. This can also be seen in Figure 12B, which shows a clear decrease in relative protease inhibitor content of the Agria potatoes.

6   Controlling Protein Content

Given the calibrated and validated protein model, a model-based controller can be designed. First, the controllability and observability of the system will be assessed, after which the assumptions and constraints for the optimal control problem will be formulated. The aim of the control will then be quantified in the cost function, and the optimal storage temperature control will be computed. Finally, an economic evaluation will be done to compare different strategies with respect to protein selling prices and energy prices.

6.1   Controllability and Observability

As mentioned in section 2.4.2, the controllability and observability of a linearized systems model is determined by the A, B, and C matrices that were calculated in section 5.4. If we recall equation 2.12, and we know that the number of states n is three and the number of inputs m is one, hence we get a 3×3 controllability matrix. The solutions shown in this section were calculated using the data of the Miss Malina cultivar, but were also confirmed for the Agria cultivar. The time-varying controllability matrix, as both A and B are time-varying matrices, is not shown here. For $t = 0$, the rank of the controllability matrix $\mathcal{R}$ is given by:

$$Rank(\mathcal{R}) = rank\left(\begin{bmatrix} -2.86e-6 & -8.74e-13 & 2.56e-15 \\ -1.22e-6 & 1.47e-11 & -1.78e-16 \\ -8.06e-7 & 2.13e-11 & -1.10e-15 \end{bmatrix}\right) = 3 \tag{6.1}$$

And thus $\mathcal{R}$ has full rank. This means that all three states of our system are controllable at $t = 0$. At first it seems that the numerical estimate of the rank might be smaller than three because of the low values. However, because the order of magnitude of the A matrix is also around 10^{-6} the numerical rank of $\mathcal{R}$ is still relevant. The controllability of the system was also confirmed halfway through the simulation, and at the end of the simulation.

From Equations 5.11-5.14, we know that the number of states n is 3 and the number of outputs p is also 3, thus we get a 9×3 observability matrix. For $t = 0$, the rank of the observability matrix $\mathcal{O}$ is given by:

$$Rank(\mathcal{O}) = rank\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -7.20e-5 & 1.70e-4 & 0 \\ 0 & -1.21e-5 & 0 \\ 0 & 3.04e-5 & -7.23e-5 \\ 5.19e-9 & -1.43e-8 & 0 \\ 0 & 1.46e-10 & 0 \\ 0 & -2.57e-9 & 5.23e-9 \end{bmatrix}\right) = 3 \tag{6.2}$$

And thus $\mathcal{O}$ has full rank too. This means that all three states of our system are observable, which makes perfect sense because we actively observe all states as outputs. This was also confirmed halfway through the simulation, and at the end of the simulation.

To analyse the effect of adding an extra state to the linearized model in the form of the amino acids, these two matrices were constructed for this case as well. The controllability matrix then becomes a 4×4 matrix, and the observability matrix has the dimensions 12×4. They received the subscript

AA, to denote that the amino acids have been added as a state (according to equation 5.5, with protease inhibition added).

$$Rank(\mathcal{R}_{AA}) = rank\left(\begin{bmatrix} -4.51e-6 & 1.20e-10 & -6.19e-15 & 4.18e-19 \\ -1.22e-6 & 1.47e-11 & -1.78e-16 & 2.15e-21 \\ -8.06e-7 & 2.12e-11 & -1.10e-15 & 7.38e-20 \\ 2.03e-4 & -5.40e-9 & 2.78e-13 & -1.88e-17 \end{bmatrix}\right) = 3 \qquad (6.3)$$

And thus $\mathcal{R}_{AA}$ is rank deficient. A rank deficiency here means that the system is now not fully controllable[1]. This is caused by the interdependence of the protein content and the amino acid content.

The rank of the observability matrix of the four-dimensional linearized model is given by:

$$Rank(\mathcal{O}_{AA}) = rank\left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -7.21e-5 & 1.70e-4 & 0 & 8.16e-9 \\ 0 & -1.21e-5 & 0 & 0 \\ 0 & 3.04e-5 & -7.23e-5 & 0 \\ 0 & 0 & 0 & 0 \\ -3.77e-13 & 1.06e-12 & 0 & 4.27e-17 \\ 0 & -1.76e-15 & 0 & 0 \\ 0 & 1.90e-13 & -3.79e-13 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\right) = 4 \qquad (6.4)$$

And thus $\mathcal{O}_{AA}$ has full column rank. Apparently, even though we do not measure the amino acid content, we are still able to "observe" its value. The dependence mentioned above, which is a hindrance for the controllability of the system, allows the amino acid content to be calculated from the protein concentration. However, the lack of controllability means it is not possible to reach all states in state space in finite time. This may not be a problem if we do not want to control the amino acids concentration.

An important reason to exclude amino acids from the model was that we had no data available because of the difficulties with the measurements. However, from the observability analysis we see that all the states are actually observable, which means that we do not need this data. However, because the system with amino acids is not fully controllable, we will still refrain from using it for the further analysis.

6.2   Optimal Control

As mentioned before, the use of a cost function in optimal control provides us with a way to quantify the aim of the control strategy. However, this also forces us to specify very clearly what we want, because a small deviation can result in a big difference in the final optimal control inputs that are calculated. For most commercial processes however, the aim of the control is to maximise production or profit.

---

[1] The null space of $\mathcal{R}_{AA}$ is given by $[0 \quad 0 \quad 0 \quad -1]^T$ and thus the fourth state of the linearized model is uncontrollable, which is the amino acid concentration.

6.2.1   Assumptions and Constraints

The ventilation strategy we are investigating depends on three main devices: System fans, a humidifier, and a cooling unit (Mol, 2015). Together, the settings of these devices control the cooling process. As the processes involved with this equipment have already been described in much detail by Mol (2015), they will be lumped together in a term describing the heat exchange between temperature of the air in the storage facility, potatoes, and outdoor air temperature and is given by:

$$E_{req} = C_{pa}M_a * |T_{IN} - T_{OUT}| + C_{pp}M_{bulk} * |T_{IN} - T_{OUT}| \tag{6.5}$$

Here, E_{req} (in Joules) represents the required energy to reach the desired indoor temperature T_{IN} (in Kelvin). The indoor temperature is assumed to be homogeneously distributed through both the air and the bulk, and corresponds to the control input $u(t)$. The required energy consists of a term describing the heat exchange with the air inside the storage facility, and a term describing the heat exchange with the potato bulk. The parameters C_{pa} and C_{pp} are the specific heat capacities of air and potatoes (in $JK^{-1}kg^{-1}$), and the parameters M_a and M_{bulk} are the masses of the air and bulk (in kg). The mass of the air can be calculated from the dimensions of the storage facility (Table 7) and the density of air. As the mass of the potatoes is much higher than the mass of the air, E_{req} is dominated by the second term related to the potato bulk. Notice that for the calculation of E_{req} the absolute temperature difference is taken. The values of each parameter can be also found in Appendix A.

Table 7. Dimensions of storage facility. Data taken from Mol (2015).

	Size (m)		
Dimension	x	y	z
Total facility	21.1	4.6	6.0
Bulk	16.5	4.6	4.0

One of the assumptions that were made is that the ventilation system can reach any temperature within one day, and can maintain any humidity. As the humidity of the air has no known effect on the dynamics of the protein content, the only reason to use a humidifier is to prevent dehydration of the tubers. We can distinguish between two scenarios regarding dehydration. Either dehydration is no problem at all because farmers get paid per kg protein, or farmers want to prevent dehydration at all times because they get paid per kg potato. In the first case disregarding the costs of a humidifier can easily be done because it will not be activated, and in the second case the humidifier will have to maintain a high humidity to prevent evaporation in the bulk. In both cases the costs are assumed to be reasonably constant, and consequently not affected by the temperature. Similarly, the costs of the ventilation fans are assumed to be fairly constant as well, and are therefore also not taken into account.

The minimum temperature in the storage facility was set at 0°C, to prevent freezing damage of the potatoes. The maximum temperature (for potential heating of the facility) was set at 30°C, because higher temperatures than this are known to induce proteolytic genes (Garbarino *et al.* 1992). Furthermore, the concentrations of each component should not be negative, because it is impossible to have negative concentrations.

The outdoor temperature was obtained from weather data in 2015-2016 in the Netherlands (Appendix B), and for simplicity four seasonal periods were distinguished: Autumn, winter, spring, and summer, each with their own average temperature. For autumn this was 9.8°C, for winter 4.4°C, for spring 8.6°C, and for summer 16.1°C.

6.2.2   Cost Function

The cost function that will be used here mostly depends on the aim of the control, which is to maximise the production of protein. In our case this can be the individual protein fractions, or the total protein content. We will take the total protein content, because the entire protein fraction can be valorised. This would mean that the cost function $J(u(t))$ is only dependent on the final concentration of total protein. The solver will then try to minimize the value of the following equation:

$$J(u(t)) = -x_1(t_f) \tag{6.6}$$

Here, $u(t)$ depicts the temperature over time, and $x_1(t_f)$ depicts the total protein concentration at the final time t_f. The minus sign ensures we are looking for the highest value of $x_1(t_f)$, not the lowest.

However, in practice, temperature control comes at a price that is not included in the cost function (6.6). The energy consumption from a heat exchange system, a ventilation fan, or a humidifier can be a significant part of the costs. As mentioned above, the effect of the ventilation fan and humidifier are assumed to be constant and are thus omitted from the cost function. The costs of the heat exchange process will be represented by the required energy over the entire period multiplied with a certain energy price. On the other hand, the revenues from the sale of protein will be represented by the total amount of protein (in kg/kg potatoes) at the end of the storage period multiplied with the entire bulk. The final cost function can then be depicted as:

$$J(u(t)) = -\frac{x_1(t_f)}{1000} * M_{bulk} * P_{prot} + P_E * \int_{t_0}^{t_f}(E_{req})dt \tag{6.7}$$

In the first term, which represents the revenues of protein sales, the final protein concentration $\left(x_1(t_f)\right)$ in mg protein/g potato is first adjusted to have the units kg protein per kg potatoes. This is done by dividing it by 1000, and is then multiplied with the potato weight (M_{bulk}) in kg and the protein price (P_{prot}) in euro per kg to arrive at the income in euro. The second term represents the costs of the heat exchange process. It consists of the total energy consumption, which is the integral of E_{req} over the storage time, multiplied with the energy price (P_E) in euro per Joule to arrive at the costs in euro.

If we compare equation 6.7 with equation 2.14, we can see that both the energy price and the weighted protein price essentially function as the two weighting factors of the cost function. The effect of changing the values of these parameters will be assessed in section 6.3.
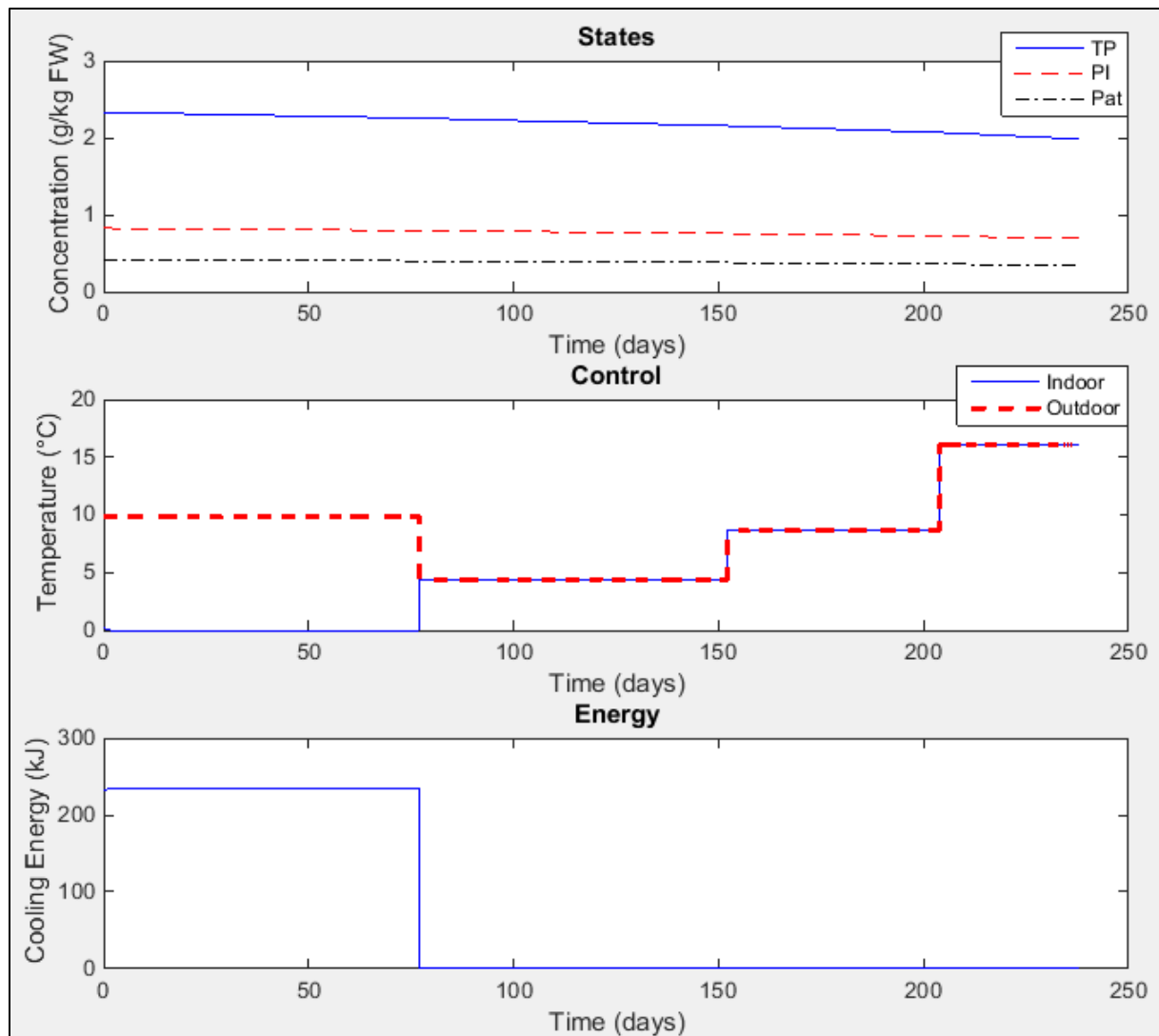
6.2.3 Temperature Control

The optimal temperature control trajectory was calculated using Tomlab solver in Matlab using the constraints, assumptions, and cost function mentioned in the previous sections. For the Miss Malina cultivar, Figure 19 shows the dynamics of the protein levels, as well as the temperature inside the facility and the associated energy consumption to reach this temperature.
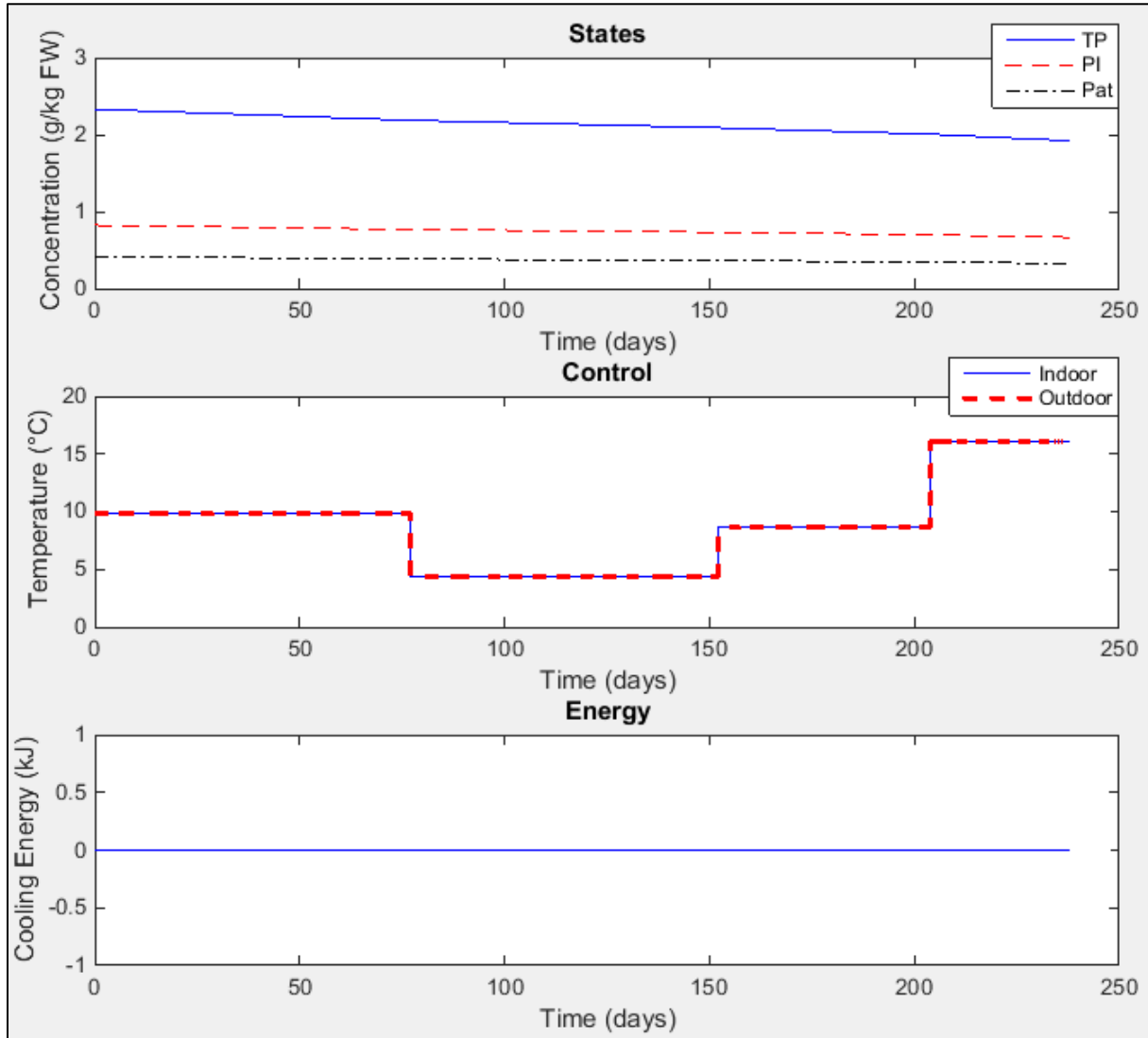
Figure 19. Optimal control scheme for storage temperature of Miss Malina potatoes. Protein selling price is €20/kg protein. TP=Total protein, PI=Protease inhibitors, Pat=Patatin.

It can be seen that the protein price of €20/kg protein is too low for cooling to be profitable. The storage temperature follows the outdoor temperature profile, and there is no energy consumption. The final total protein content is $1.9209 \left[\frac{g}{kg\ FW}\right]$. Cooling becomes profitable at a protein selling price of roughly €50/kg, and at about €80/kg the entire first period gets cooled maximally to 0°C (Figure 20). The final total protein concentration for this last scenario is $1.9828 \left[\frac{g}{kg\ FW}\right]$.

Figure 20. Optimal control scheme for storage temperature of Miss Malina potatoes. Protein selling price is €80/kg protein. TP=Total protein, PI=Protease inhibitors, Pat=Patatin.

Increasing the protein price even more results in the next periods being maximally cooled one by one. The protein price was increased to a maximum of €500/kg, which may not be realistic but shows the potential of cooling the entire storage period. This will be shown in further detail in the next section.

The option of heating was investigated for the Agria cultivar, because this was shown to increase its total protein content (Figure 18). For the Agria potatoes, Figure 21 shows the dynamics of the protein levels, as well as the temperature inside the facility and the associated energy consumption to reach this temperature. There are only three periods here, because harvesting of the Agria cultivar was done before summer. Again, at a protein price of €20/kg heating is not feasible.
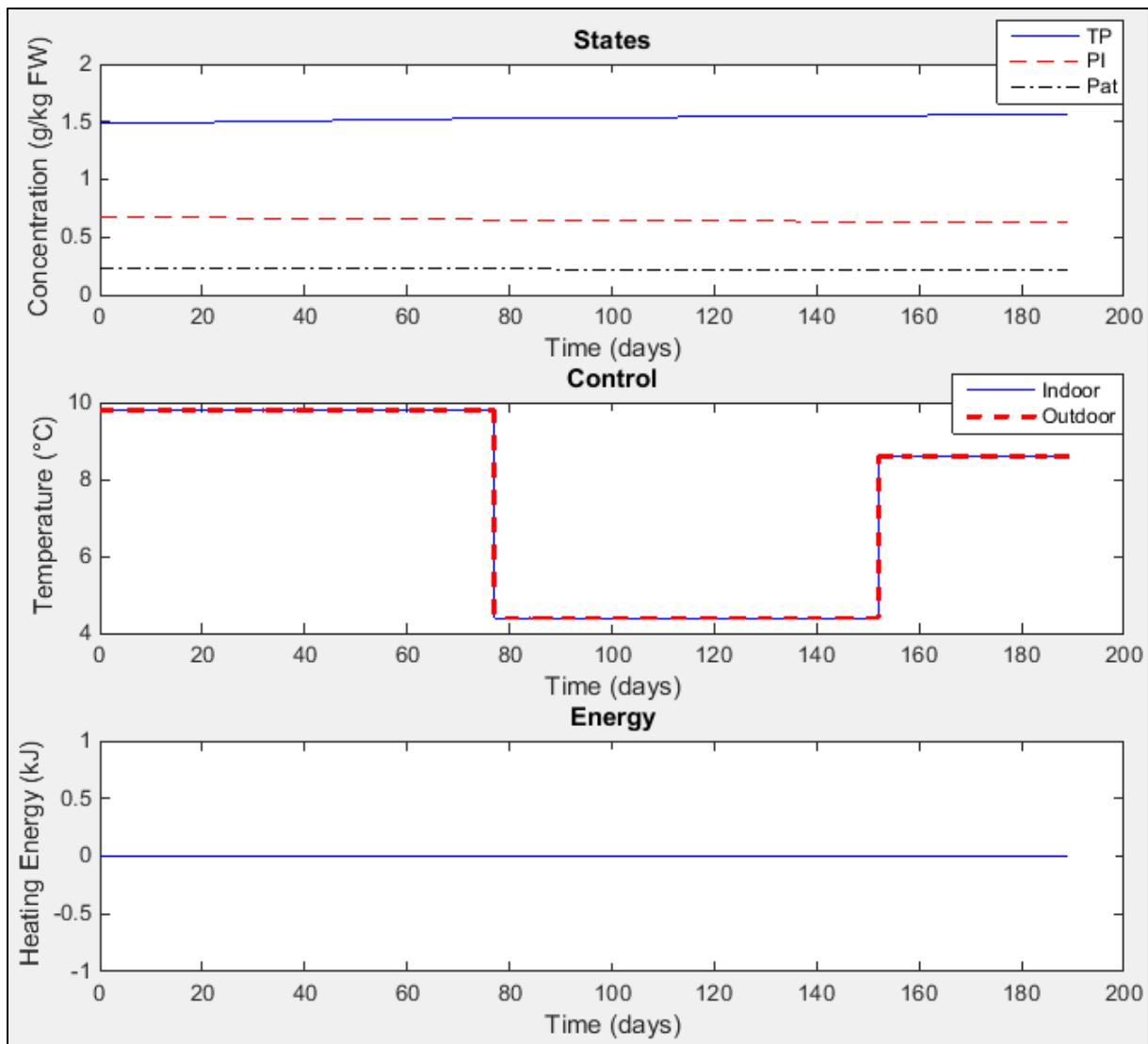
Figure 21. Optimal control scheme for storage temperature of Agria potatoes. Protein selling price is €20/kg protein. TP=Total protein, PI=Protease inhibitors, Pat=Patatin.

Heating becomes feasible at a protein selling price of around €90/kg, at which it is most profitable to heat the entire first period to the maximum temperature of 30°C (Figure 22). When comparing the heating of Agria with the cooling of Miss Malina, it can be seen that the influence of heating is much higher than that of cooling. This is probably caused by the higher temperature difference.

The increase in final total protein concentration also shows that heating has more influence than cooling, because cooling only raised the content by roughly $0.06\frac{g}{kg\ FW}$, and heating the entire first period raises it by more than $0.3\frac{g}{kg\ FW}$. However, if heating is stopped after the first period, the protein content slightly increases after that. This is probably caused by the fact that protein degradation is dependent on the amount of total protein and protease inhibitor.

When we look at equation 5.8, we can see that if $[TP]$ becomes large and $[PI]$ becomes small, the degradation term becomes larger. Therefore, the net protein content increase becomes smaller. As with the case of the Miss Malina cultivar, if the protein price is raised the other periods can be heated as well. This is investigated in more detail in the next section.
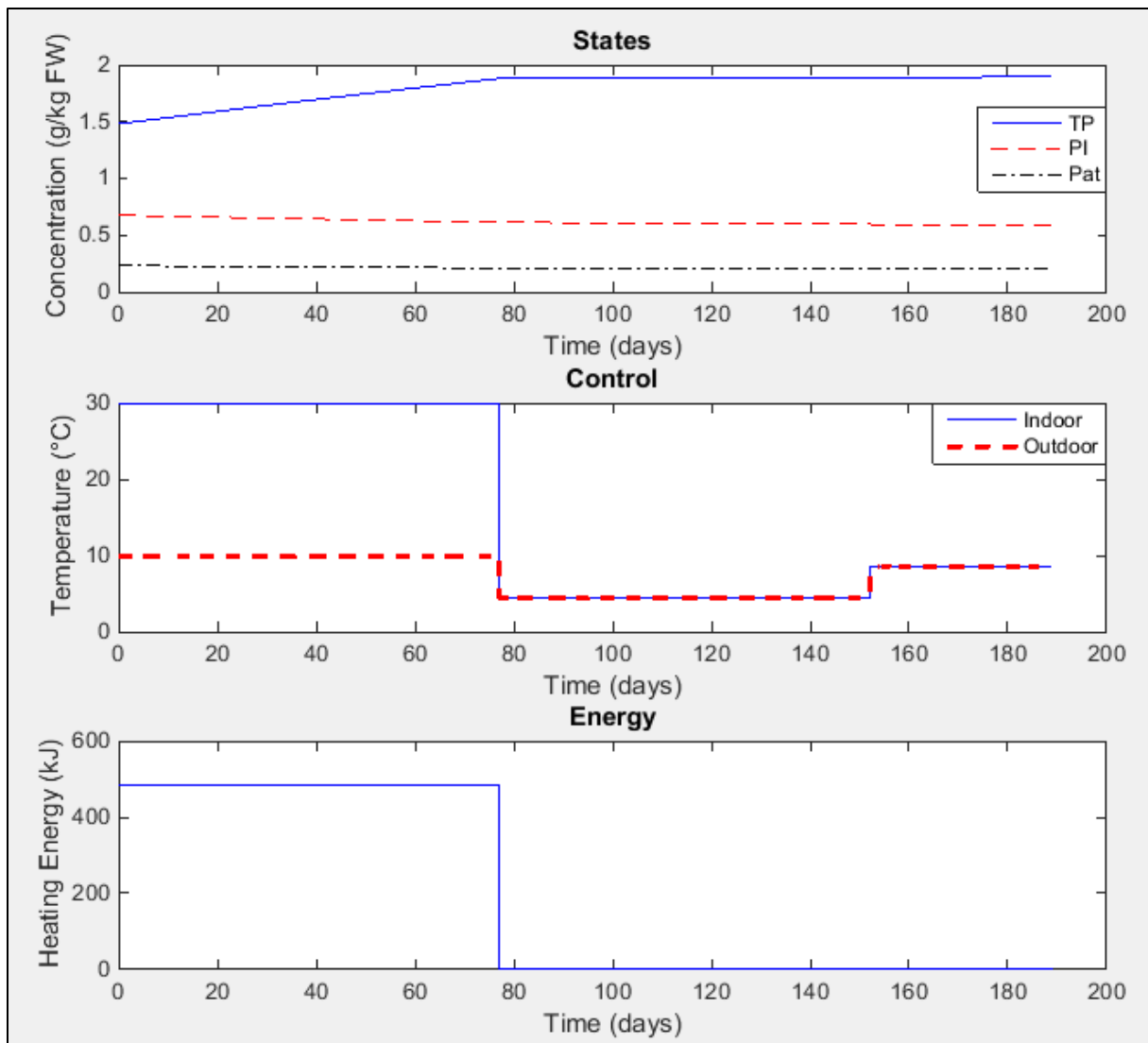
Figure 22. Optimal control scheme for storage temperature of Agria potatoes. Protein selling price is €90/kg protein. TP=Total protein, PI=Protease inhibitors, Pat=Patatin.

6.3   Economic Evaluation

In the following section, the optimal storage temperature control was calculated for a large range of protein selling prices and energy prices. The total energy consumption, the final total protein concentration, and the revenues compared to no heating or cooling were calculated for each scenario. In the case of heating the facility for an increase in protein content in Agria potatoes, the moisture loss that results from a higher temperature was also assessed.

6.3.1   Energy Consumption and Temperature

The total energy consumption at the end of the storage period as well as the average storage temperature were calculated as functions of the protein selling price and energy price. Figure 23 shows the energy consumption for both cultivars. It can be seen that the dynamics of the cooling process differ from those of the heating process. Apparently, an increase in protein selling price brings about a gradual increase in the amount of cooling that is feasible. It flattens out at high protein prices. On the other hand, for the Agria potatoes it seems to be most efficient to heat an entire period at once and then heat the entire next period when the protein price is sufficiently high.
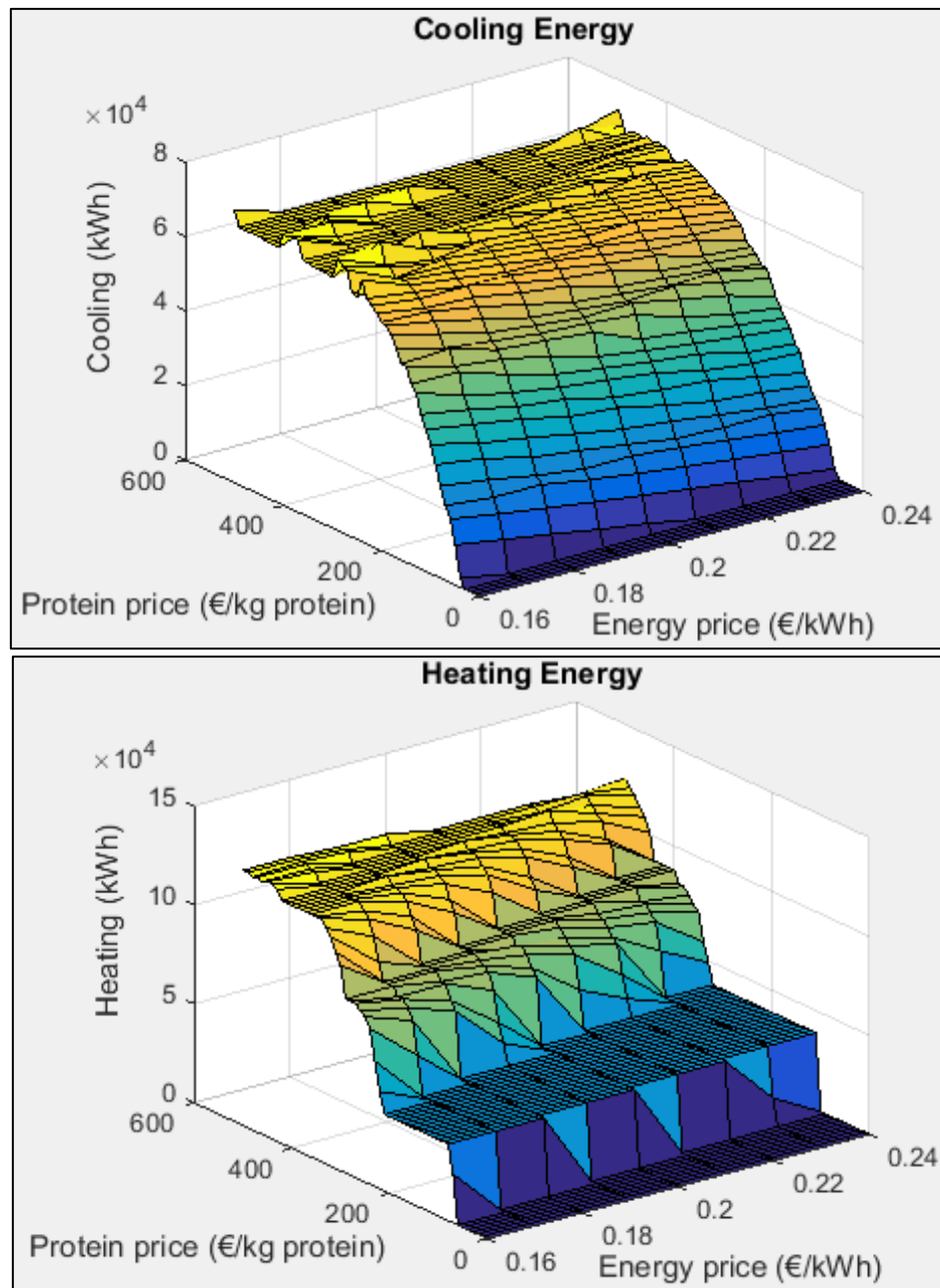

Figure 23. Total energy consumption of cooling or heating of storage facility. Top diagram corresponds with Miss Malina cultivar; the bottom diagram corresponds with Agria cultivar.
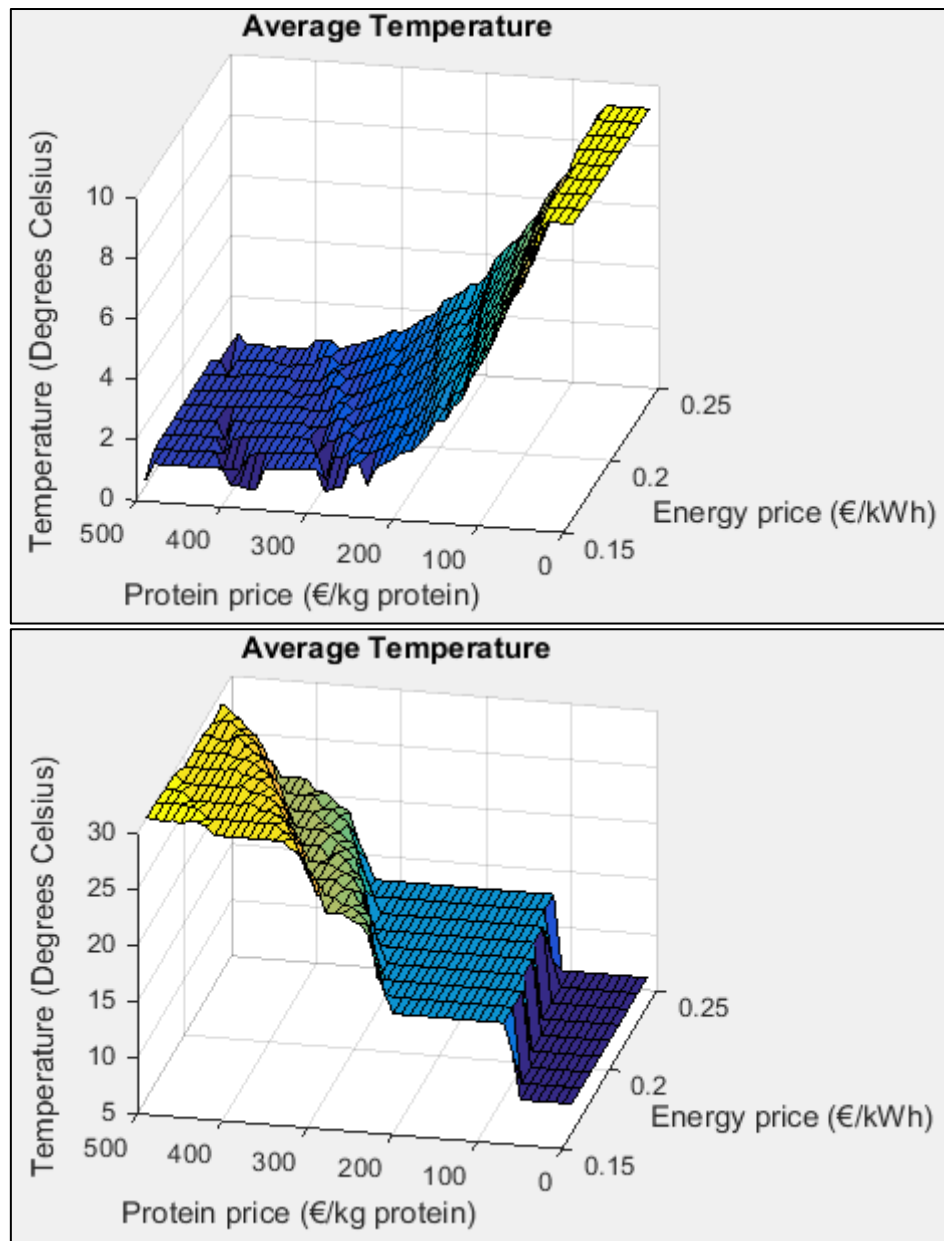
Figure 24. Average storage temperature as function of protein selling price and energy price. Top diagram corresponds with Miss Malina cultivar; the bottom diagram corresponds with Agria cultivar.

Figure 24 shows the average storage temperature as a function of the protein selling price and energy price. The average temperature profile follows the same dynamics as the energy consumption, which was to be expected. At sufficiently high protein selling prices coupled with lower energy prices, the average storage temperature of the Miss Malina cultivar approaches 0°C. For this case, the average storage temperature of the Agria cultivar approaches 30°C. This means that the energy consumption at those points is equal to the maximum possible total energy consumption.

Another observation that can be made is that the energy price seems to have more influence at higher protein selling prices. This is probably due to the simultaneous increase in energy consumption.

6.3.2   Total Protein Concentration

The total protein concentration at the end of the storage period was calculated as a function of the protein selling price and energy price. Figure 25 shows the final protein concentration for both cultivars. Again the same trends are shown in both cultivars, reflecting the influence of increasing or decreasing the storage temperature.
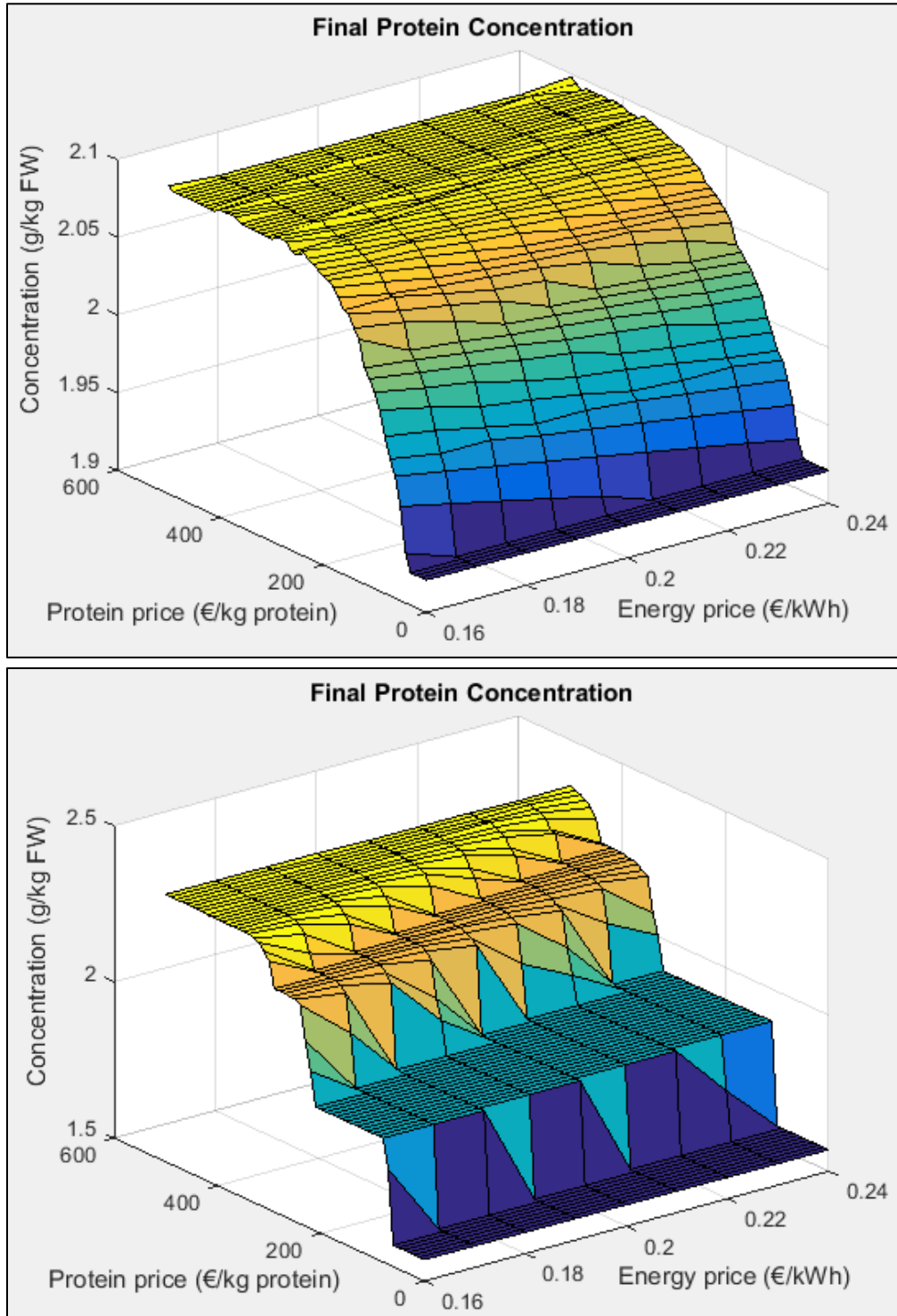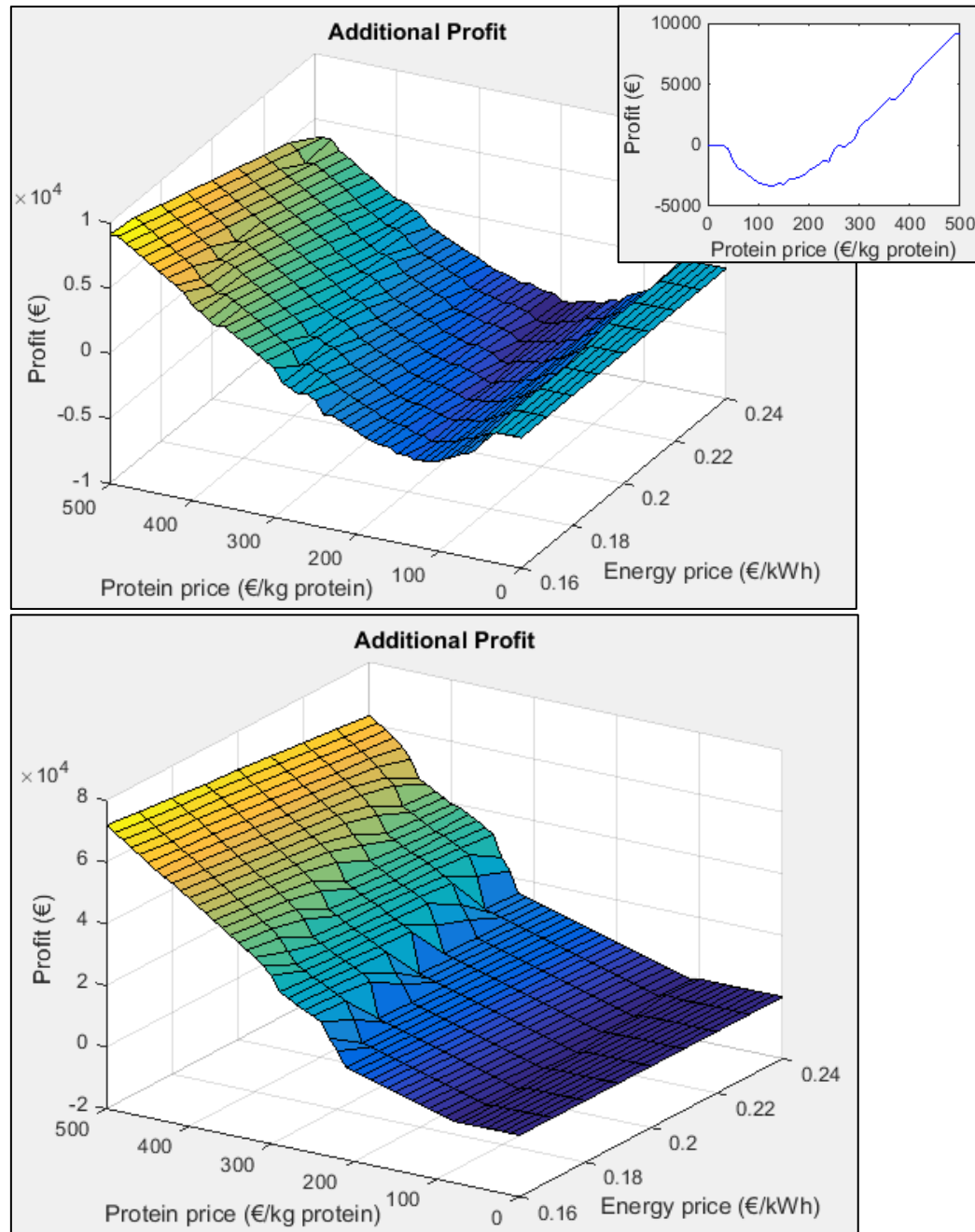

Figure 25. Total protein concentration at the end of the storage period. Top diagram corresponds with Miss Malina cultivar; bottom diagram corresponds with Agria cultivar.

Related to the amount of protein that is produced is the profit that is made from the selling of the proteins. First, the profit of potato storage without energy consumption was calculated by multiplying the protein concentration with the amount of potatoes and the protein selling price. This profit is independent of the energy price, and follows a linear relationship (Appendix E). Then the profit was derived including the processes of cooling or heating, and the difference between the two was calculated. Figure 26 shows the additional profit compared to the situation without energy consumption.

Figure 26. Additional profit caused by cooling or heating storage facility compared to outdoor temperature. Top diagram corresponds with Miss Malina cultivar; bottom diagram corresponds with Agria cultivar. Top-right graph shows additional profit as function of protein price at an energy price of €0.2/kWh.

Both diagrams in Figure 26 start at zero, because there is no energy consumption at low protein prices in any scenario. The bottom graph clearly shows the positive influence of heating for the Agria cultivar, with additional profits rising when the amount of heating rises.

However, the top diagram related to the Miss Malina potatoes actually shows a large range where cooling is less profitable. Even though the optimal control algorithm calculated the maximum profit at each point, there is still a whole range of values for which the costs of cooling outweigh the benefits. This will be further discussed in section 7.2.

6.3.3   Humidity

Although the costs of moisture losses of the potatoes were not taken into account in the cost function of the optimal control, the effect of an increase in storage temperature on evaporation was investigated as well. The moisture loss of the potatoes as a result of evaporation was calculated using an adaptation of the model introduced by Lukasse *et al.* (2007). This model is shown in Appendix D. Over the course of the day the humidity of the air surrounding the potatoes rises by the evaporation of water from the potatoes. It was assumed that by ventilating for 1-2 hours at the end of each day, the additional water content in the air due to evaporation is removed, and the process starts again with air having a relative humidity of 95%.

The amount of moisture that is removed from the potatoes per day is shown in Figure 27. This was calculated as the increase in moisture concentration in the air multiplied with the amount of air surrounding the bulk. It can be seen that the moisture losses show an exponential trend, with increasing evaporation rates at higher temperatures. This means that the costs of heating become exponentially higher with higher temperatures, which is an extra cost that has to be included in the cost function in future research.
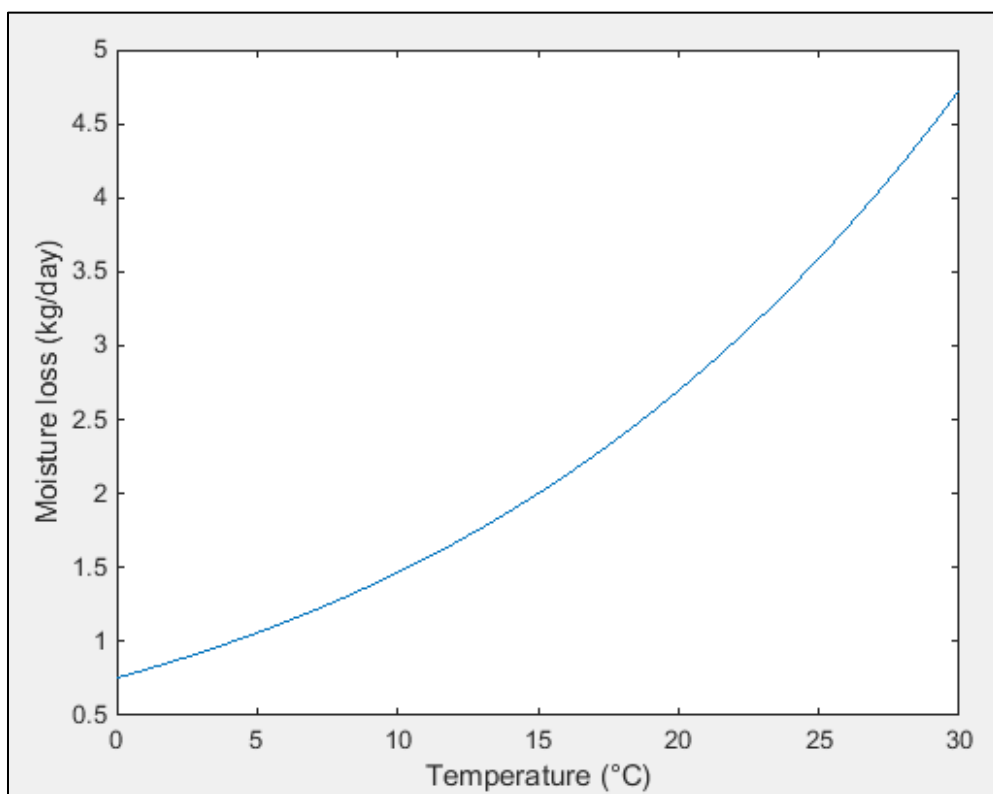

Figure 27. Effect of storage temperature on potato moisture loss. Moisture loss in kg water per day.

7   Discussion

This section will contain a discussion on the results of both the experimental procedures and the modelling procedures of this thesis. Issues regarding the experiment will first be discussed. Subsequently, the modelling shortages will be discussed.

7.1   Experiment

Some issues regarding the experimental data will be discussed here. For instance, the experiment was performed using five potatoes for each sample, which might be a too small number to be representable for a bulk of several tons of potato. This can also be seen in the standard deviations of the samples (Appendix C), which are quite high for some samples. The data also shows much fluctuation in the protein content, whereas data from other articles does not (Mazza, 1983; Brasil *et al.*, 1993; Pots *et al.*, 1999). On the other hand, most other articles have a significantly smaller amount of data points, which may reduce the fluctuations in the protein content. It is therefore difficult to say whether the data that was collected in this thesis project is inferior or superior to datasets found in literature.

Another setback regarding the experiment results was the failure to retrieve any data on the amino acids. While this could be worked around in the model by assuming zero-order protein synthesis, the completeness of the model suffered from it. It also might be possible that the parameters of the model could be estimated more accurately if more data was available. Finally, apart from the proteins the dynamics of the amino acids would have been interesting to observe because of their relation to frying colour (Khanbari & Thompson, 1993).

7.2   Modelling

As has been mentioned in section 5.5, the accuracy of the parameter estimates was quite low. This can be caused by a number of things, but most probably some parameters in the model are redundant. A solution to this can be to find parameter values in literature and fix them, or lump several parameters together in one parameter. Regarding the model used in this thesis, it might be concluded that instead of formulating the rate of change as the sum of two terms, the process can be described by a single term that depicts the net change. This would eliminate the different parameters associated with either synthesis or degradation and replace them with a single set describing the sum of these two processes.

A simple numerical experiment of this idea is to set all parameters except A_{PS} and A_{PD} to a fixed value. The sum of squares of the residuals was then calculated when simulating the model with different combinations for these two parameters. Figure 28 shows the three-dimensional graph depicting the results of these calculations, while Figure 29 shows a contour plot of the results. It can be seen that there is a linear relation between A_{PS} and A_{PD} that gives similar outcomes for the sum of squares. A linear relation is a strong indication that these two parameters can be lumped together in a single one, which can then be accurately estimated. Lumping parameters together can probably be done in a similar way with the parameters A_{PI} plus A_{PD}, E_{PS} plus E_{PD}, and E_{PI} plus E_{PD}. This would the reduce the amount of unknown parameters to six instead of eight, which may also improve the identifiability of the system.
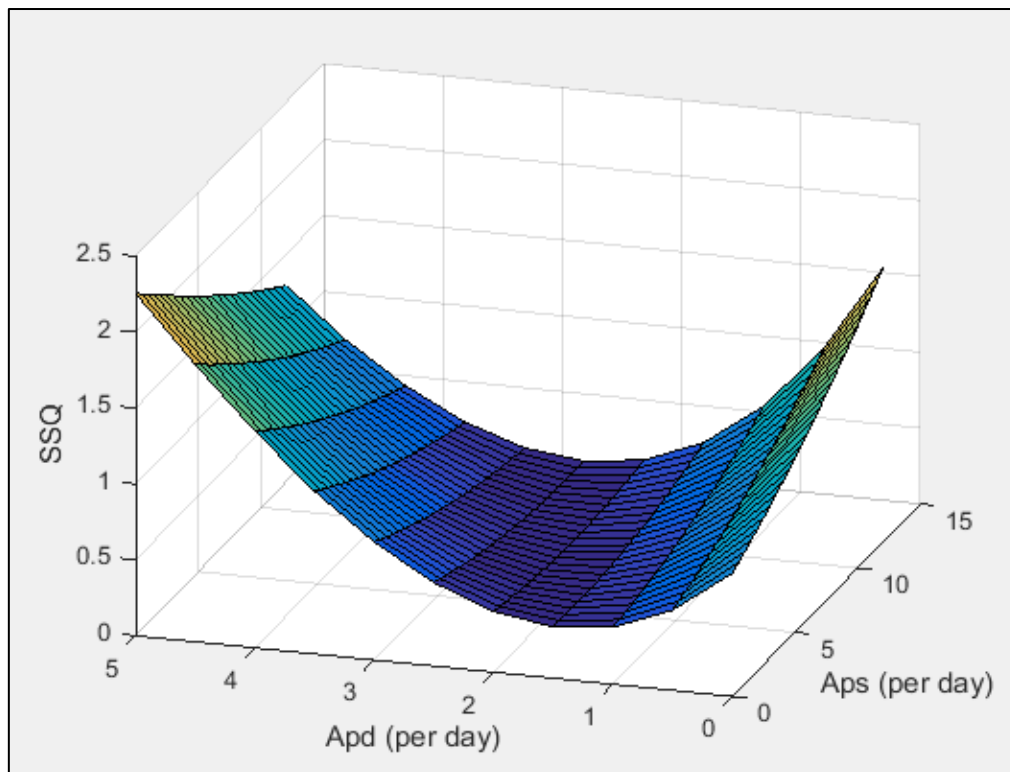
Figure 28. Sum of squares for different parameter combinations of Miss Malina cultivar parameter estimation. SSQ=sum of squares of residuals, Apd=Arrhenius constant protein degradation, Aps= Arrhenius constant protein synthesis.
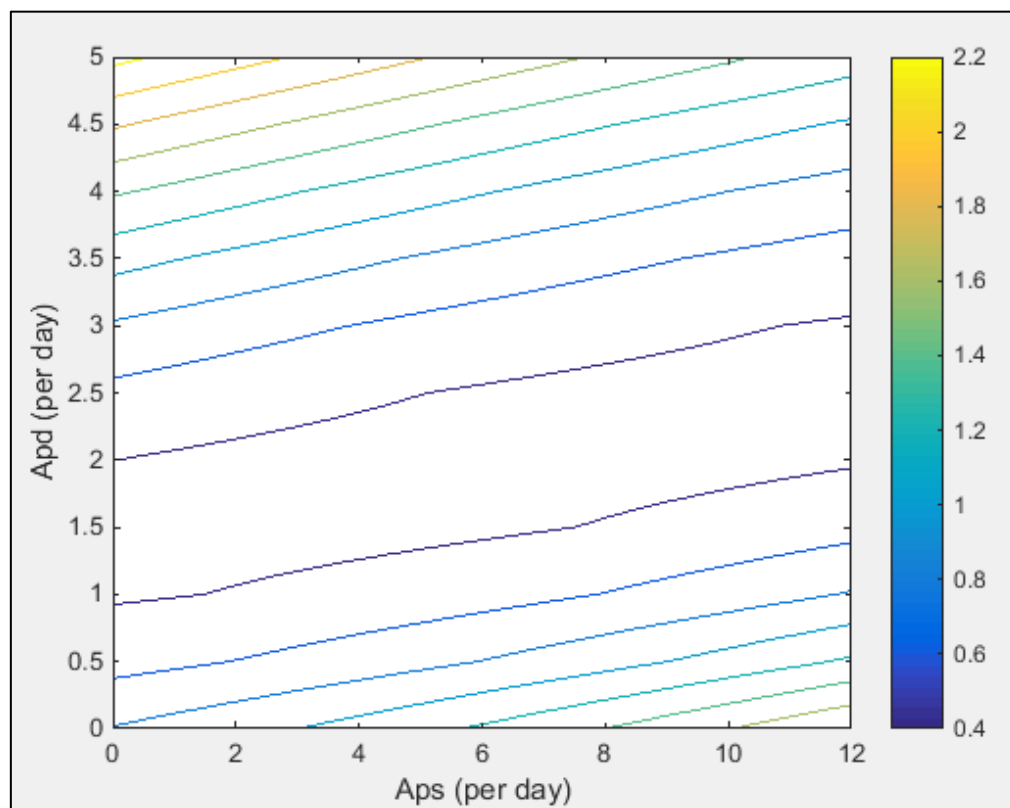
Figure 29. Contour plot of sum of squares for different parameter combinations of Miss Malina cultivar parameter estimation. Coloured bar represents sum of squares values, Apd=Arrhenius constant protein degradation, Aps= Arrhenius constant protein synthesis.

Another factor that has influenced the results of the simulation is the simplification that has been made regarding the processes of protein synthesis. Although multiple studies have shown that the rate of total protein synthesis can be approximated by Michaelis-Menten kinetics, the actual rate of synthesis for individual proteins depends on a lot of processes that are lumped together here.

In order for protein synthesis to occur genes have to be expressed first, resulting in RNA strands. These strands have to make their way out of the cell nucleus to ribosomes in the cytoplasm. The protein synthesis is then performed by reading the RNA strand and constructing a line of amino acids that together form the protein. Even by explaining this very complicated process like this means simplifying it greatly. Mathematical modelling of detailed protein synthesis has been done before on prokaryotic organisms (Drew, 2001), but it remains to be done on higher organisms such as plants or mammals.

7.3   Economy

Several economic aspects of potato storage were not taken into account in this study. The increase in moisture loss with an increasing storage temperature was calculated, but the effects of this on the costs were not assessed. Significant moisture losses can have a number of side effects, especially for potatoes that are not necessarily meant for the starch industry. If the potatoes have been severely dehydrated, the process of frying may behave differently. Furthermore, potato farmers are currently being paid a potato price per kg of potatoes. Any moisture loss is a direct loss of weight, and thus money.

Relating to moisture losses are the costs of a humidifier to counter the moisture loss during storage, as well as the costs of applying ventilation. These things have been modelled for a potato storage facility before (Mol, 2015), but only for a narrow range in storage temperature (5-8℃). The storage temperatures that are proposed in this study are more radical than this, and might influence the effect of the humidifier in different ways than found until now.

In current practice, heating of the storage facility is unheard of, especially relatively high temperatures such as 30℃. However, if farmers in the future were not paid per kilogram potatoes but are paid for the amount of starch or protein that their potatoes yield, higher temperatures may prove to be of importance. Future research should thus be conducted to obtain data on the influence of high storage temperatures on the starch and protein content. This effect is irrelevant for potatoes used for the food industry, because moisture losses will negatively affect the quality of potatoes used for consumption.

The costs of ventilation depend on the length of the storage period, the number of ventilation periods and the capacity of the ventilation fans (Mol, 2015). Not only do these aspects affect the humidity of the facility, they also affect the temperature. This effect should therefore also be incorporated in future research.

Finally, a note about the temperature control that was calculated for the Miss Malina cultivar. Apparently, cooling too much at a protein selling price that is relatively low results in less profit than without any cooling at all. For example, at an energy price of €20/kWh and a protein price of €200/kg protein, the profit without cooling is €90280,-. For the same prices, the process with cooling gives a profit of €86770,-. This is counterintuitive, because the optimal control objective was to maximise the net profit. Exactly the same cost function was used for the calculation of the Agria optimal

temperature, and there we can clearly see that the control provides more profit than without control (Figure 26). It is therefore unclear what caused the optimal control algorithm to supply a suboptimal choice, especially when considered that the initial setting was set to apply no temperature control.

8   Conclusions

In this research, the protein content of potatoes during storage was measured after conservation of the samples for several months at $-18℃$. A mathematical model was formulated to describe the dynamics of the protein content over time. Several simulations were performed to assess the influence of the control variable on the system outputs, and a control strategy was calculated. The results of these analyses will be used to answer the research questions that were formulated in the introduction of this report. Several sub questions related to the main objective will be answered first.

How can the parameters of the system be estimated?

During the parameter estimation procedure, it became apparent that the parameters of the system were not identifiable. This problem was mainly caused by the fact that some parameters of the model were redundant. Most likely, the two main processes of synthesis and degradation can be lumped together to obtain a net rate of change of the protein content. This would eliminate several parameters that are currently the cause of the unidentifiability of the system.

Another solution to the problem would be to conduct targeted research to find specific parameter values. Several studies have been performed on specifically protein synthesis rates (Simon & Azam, 1989; Drew, 2001) or protein degradation rates (Broderick, 1987; Ørskov & McDonald, 1979). Most of the studies on synthesis are performed in bacteria, while most studies on degradation are performed in mammals. Therefore, additional research has to be performed on the protein synthesis and degradation rates in plants to make it possible to fix several parameters in the model. This will improve the identifiability of the system, and the accuracy of the remaining parameter estimates.

Which processes influence the protein content in potato tubers during long-term storage?

The main process that defines protein content was identified as protein turnover. Protein turnover consists of the sum of the total protein synthesis and protein degradation. It was found that both of these parts can be modelled as following Michaelis-Menten kinetics coupled with a temperature dependence. This temperature dependence was expressed in terms of the Arrhenius equation, which relates the maximum reaction rate to the temperature with an exponential term. Temperature was therefore identified as the main control variable that affects the protein content.

Another influence on the protein degradation was identified as the content of protease inhibitors of the potato tubers. A higher protease inhibitor content is associated with a higher inhibition of protein degradation, and consequently a higher total protein content. The protease inhibitor content of potatoes is relatively high, at 35-50 percent of the total protein, which means that their contribution to protein degradation is probably relevant.

Furthermore, it was found that humidity and moisture content of the potatoes is relatively unimportant. No direct influence of storage humidity on protein content was found in literature, and it was therefore not taken into account during the modelling and control of the process.

What is the best storage temperature to maximise the protein content during long-term storage?

The ideal storage temperature for maximum protein production from potatoes was calculated using model simulations. It was found that the potato cultivar that is stored has a big influence on the effect of the storage temperature. For the Miss Malina cultivar, a low storage temperature resulted

in a higher protein content. This holds for the total protein as well as the concentrations of the individual fractions. On the other hand, the Agria potatoes were found to produce a higher total protein content at a higher storage temperature.

For both cultivars, an extreme temperature works better than an average temperature such as the temperatures that are currently used. However, these temperatures are often not favourable for other applications. For instance, the low storage temperature proposed for the Miss Malina potatoes will cause a phenomenon called cold-sweetening, which speeds up the degradation of starch to sugars (Sowokinos, 2001). The heightened sugar levels that this causes also negatively affect the potato quality for frying (Khanbari & Thomas, 1992).

Finally, the following main research question was stated:

What is an appropriate protein model for the control of protein levels in potato tubers during long-term storage?

The protein model was set up by formulating differential equations to describe the rates of change in protein content. The two major groups of protein in potato, patatins and protease inhibitors, were also described in a similar manner. It was found that both the synthesis and degradation could be modelled as following Michaelis-Menten kinetics coupled with the Arrhenius theory. Consequently, the only control parameter is temperature. Another consequence of modelling the process as was done is the following. It appears that once an individual potato (cultivar) has the tendency to increase or decrease, this trend can only be limited but not reversed. Effectively this means that once the protein begins to decline, it will not increase anymore, or vice versa. For the Miss Malina cultivar, this is confirmed by the data that was obtained. The data shows a steady decline in protein content, without significant increases. The Agria data on the other hand shows multiple increases and decreases, without any clear structure.

Data obtained over multiple years, with higher sample sizes, may provide additional proof that the model proposed in this thesis can be used to control protein levels in potato tubers during storage. Experiments with different, more extreme storage temperatures than currently used may also provide insight in the use of novel storage strategies. To conclude, the model applied in this thesis provides insight in the behaviour of the protein content of potatoes in storage, as well as its response to different storage temperatures. However, more research is needed to be able to accurately estimate the model parameters and validate the model for multiple storage seasons and potato cultivars.

9   References

Alberts, B., Bray, D., Hopkin, K., Johnson, A., Lewis, J., Raff, M., ... & Walter, P. (2013). *Essential cell biology*. Garland Science.

Andrews, D. L., Beames, B., Summers, M. D., & Park, W. D. (1988). Characterization of the lipid acyl hydrolase activity of the major potato (*Solanum tuberosum*) tuber protein, patatin, by cloning and abundant expression in a baculovirus vector. *Biochemical Journal*, *252*(1), 199–206.

Antoun, A., Pavlov, M. Y., Lovmar, M., & Ehrenberg, M. (2006). How initiation factors tune the rate of initiation of protein synthesis in bacteria. *The EMBO Journal*, *25*(11), 2539–50.

Baskin, K. K., & Taegtmeyer, H. (2011). Taking pressure off the heart: the ins and outs of atrophic remodelling. *Cardiovascular research*, *90*(2), 243-250.

Blenkinsop, R. W., Copp, L. J., Yada, R. Y., & Marangoni, A. G. (2002). Changes in compositional parameters of tubers of potato (*Solanum tuberosum*) during low-temperature storage and their relationship to chip processing quality. *Journal of Agricultural and Food Chemistry*, *50*(16), 4545–4553.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*(1–2), 248–254.

Brasil, P. J., César, A. B., & Marcio, H. P. (1993). Effects of different storage temperatures on protein quantities of potato tubers. *Revista Brasileira de Fisiologia Vegetal*, *5*(2), 167-170.

Brierley, E. R., Bonner, P. L. R., & Cobb, a. H. (1997). Aspects of amino acid metabolism in stored potato tubers (cv. Pentland dell). *Plant Science*, *127*(1), 17–24.

Brierley, E. R., Bonner, P. L. R., & Cobb, A. H. (1996). Factors Influencing the Free Amino Acid Content of Potato (*Solanum tuberosum*) Tubers during Prolonged Storage. *Journal of the Science of Food and Agriculture*, *70*, 515–525.

Broderick, G. A. (1987). Determination of protein degradation rates using a rumen in vitro system containing inhibitors of microbial nitrogen metabolism. *British journal of nutrition*, *58*(03), 463-475.

Davids, S. J., Yaylayan, V. A., & Turcotte, G. (2004). Use of unusual storage temperatures to improve the amino acid profile of potatoes for novel flavoring applications. *LWT - Food Science and Technology*, *37*(6), 619–626.

Davidson, E. A., Samanta, S., Caramori, S. S., & Savage, K. (2012). The Dual Arrhenius and Michaelis-Menten kinetics model for decomposition of soil organic matter at hourly to seasonal time scales. *Global Change Biology*, *18*(1), 371–384.

Davies, A. M. C. (1977). The free amino acids of tubers of potato varieties grown in England and Ireland. *Potato Research*, *20*(1), 9–21.

Dhondt, S., Geoffroy, P., Stelmach, B. A., Legrand, M., & Heitz, T. (2000). Soluble phospholipase A2 activity is induced before oxylipin accumulation in tobacco mosaic virus-infected tobacco leaves and is contributed by patatin-like enzymes. *Plant Journal*, *23*(4), 431–440.

Drew, D. A. (2001). A mathematical model for prokaryotic protein synthesis. *Bulletin of Mathematical Biology*, *63*(2), 329–51.

Dumas, A., Dijkstra, J., & France, J. (2008). Mathematical modelling in animal nutrition: a centenary review. *The Journal of Agricultural Science*, *146*(02), 123-142.

Filner, P., Wray, J. L., & Varner, J. E. (2016). Enzyme Induction in Higher Plants, *165*(3891), 358–367.

Garbarino, J. E., Rockhold, D. R., & Belknap, W. R. (1992). Expression of stress-responsive ubiquitin genes in potato tubers. *Plant Molecular Biology*, *20*(2), 235–244.

Gettins, P. G. W. (2002). Serpin structure, mechanism, and function. *Chemical Reviews*, *102*(12), 4751–4803.

Glasziou, K. T. (1969). Control of enzyme formation and inactivation in plants. *Annual review of plant physiology*, *20*(1), 63-88.

Grilly, C., Stricker, J., Pang, W. L., Bennett, M. R., & Hasty, J. (2007). A synthetic gene network for tuning protein degradation in Saccharomyces cerevisiae. *Molecular Systems Biology*, *3*(127), 127.

Hanczko, R., Jámbor, A., Perl, A., & Molnár-Perl, I. (2007). Advances in the o-phthalaldehyde derivatizations: comeback to the o-phthalaldehyde–ethanethiol reagent. *Journal of Chromatography A*, *1163*(1), 25-42.

Hargrove, J. L. (1993). Microcomputer-assisted kinetic modelling of mammalian gene expression. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *7*(12), 1163–70.

Heijne, von G., Blomberg, C., & Liljenström, H. (1987). Theoretical modelling of protein synthesis. *Journal of theoretical biology*, *125*(1), 1-14.

Hersch, G. L., Baker, T. a, & Sauer, R. T. (2004). SspB delivery of substrates for ClpXP proteolysis probed by the design of improved degradation tags. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(33), 12136–41.

Hill, A. J., & Peikin, S. R. (1990). Oral administration of proteinase inhibitor II from potatoes reduces energy intake in man. *Physiology and Behavior*, *48*(2), 241–246.

Holm, F., & Eriksen, S. (1980). Emulsifying properties of undenatured potato protein concentrate. *International Journal of Food Science & Technology*, *15*(1), 71-83.

Huffaker, R. C., & Peterson, L. W. (1974). Protein Turnover in Plants and Possible Means of Its Regulation. *Annual Review of Plant Physiology*, *25*, 363–392.

Jones, D. H. (1984). Phenylalanine ammonia-lyase: Regulation of its induction, and its role in plant development. *Phytochemistry*, *23*(7), 1349–1359.

Kapoor, A. C., Desborough, S. L., & Li, P. H. (1975). Potato tuber proteins and their nutritional quality. *Potato Research*, *18*(3), 469–478.

Kennedy, A. (1998). Chemopreventive Agents Protease Inhibitors. *Pharmacology & Therapeutics*, *78*(3), 167–209.

Khanbari, O. S., & Thompson, A. K. (1993). Effects of amino acids and glucose on the fry colour of potato crisps. *Potato Research*, *36*(4), 359–364.

Kumar, G. N. M., & Knowles, N. R. (1993). Age of potato seed-tubers influences protein synthesis during sprouting, (Dhindsa 1982), 262–270.

Lancelot, C., Mathot, S., & Owens, N. (1986). Modelling protein synthesis, a step to an accurate estimate of net primary production: *Phaeocystis pouchetii* colonies in Belgian coastal waters. *Marine Ecology Progress Series*, *32*, 193–202.

Løkra, S., & Strætkvern, K. O. (2009). Industrial Proteins from Potato Juice. A Review. *Food*, *3*(1), 88–95.

Lukasse, L. J. S., de Kramer-Cuppen, J. E., & van der Voort, A. J. (2007). A physical model to predict climate dynamics in ventilated bulk-storage of agricultural produce. *International Journal of Refrigeration*, *30*(1), 195–204.

Marcus, A. (1971). Enzyme induction in plants. *Annual Review of Plant Physiology*, *22*(1), 313-336.

Matsuura-Endo, C., Ohara-Takada, A., Chuda, Y., Ono, H., Yada, H., Yoshida, M., & Mori, M. (2006). Effects of storage temperature on the contents of sugars and free amino acids in tubers from different potato cultivars and acrylamide in chips. *Bioscience, Biotechnology, and Biochemistry*, *70*(5), 1173–80.

Mazza, G. (1983). Correlations between quality parameters of potatoes during growth and long-term storage. *American Potato Journal*, *60*(3), 145–159.

Mol, J. C. (2015). Dynamic simulation and control of potato bulk storage. MSc Thesis, Wageningen University & Research.

Neuroscience Core Facility | Biological Assays. (2016). Retrieved October 21, 2016, from http://sites.gsu.edu/neuroscience-core/training/assays/

Nowak, J. (1977). Biochemical changes in stored potato-tubers with different rest periods .1. influence of storage temperature and isopropyl phenylcarbamates (IPC and CIPC) on protein changes. *International Journal of Plant Physiology*, *81*(2), 113–124.

Nxumalo, F., Glover, N. R., & Tracey, A. S. (1998). Kinetics and molecular modelling studies of the inhibition of protein tyrosine phosphatases by N,N-dimethylhydroxylamine complexes of vanadium. *Journal of Biological Inorganic Chemistry*, *3*(5), 534–542.

Ogata, K. (2010). *Modern Control Engineering*.

Ørskov, E. R., & McDonald, I. (1979). The estimation of protein degradability in the rumen from incubation measurements weighted according to rate of passage. *The Journal of Agricultural Science*, *92*(02), 499-503.

Paiva, E., Lister, R. M., & Park, W. D. (1983). Induction and accumulation of major tuber proteins of potato in stems and petioles. *Plant Physiology*, *71*(1), 161–8.

Phenomenex (2016). BioSep: High Performance Size Exclusion for Biomolecules. http://www.phenomenex.com/Products/HPLCDetail/BioSep-SEC-S

Pots, A. M., Gruppen, H., Van Diepenbeek, R., Van Der Lee, J. J., Van Boekel, M. A. J. S., Wijngaards, G., & Voragen, A. G. J. (1999). The effect of storage of whole potatoes of three cultivars on the patatin and protease inhibitor content; a study using capillary electrophoresis and MALDI-TOF mass spectrometry. *Journal of the Science of Food and Agriculture*, *79*(12), 1557–1564.

Pots, A. (1999). Physico-chemical Properties and Thermal Aggregation of Patatin, the Major Potato Tuber Protein. Ph.D. Thesis, Wageningen University & Research.

Racusen, D. (1984). Lipid acyl hydrolase of patatin. *Canadian Journal of Botany*, *62*(8), 1640-1644.

Racusen, D., & Foote, M. (1980). A major soluble glycoprotein of potato tubers. *Journal of Food Biochemistry*, *4*(1), 43-52.

Ralet, M.-C., & Guéguen, J. (2000). Fractionation of Potato Proteins: Solubility, Thermal Coagulation and Emulsifying Properties. *LWT - Food Science and Technology*, *33*(5), 380–387.

Rutquist, P., Edvall, M. M. (2008). *PROPT - Matlab Optimal Control Software*. 1260 SE Bishop Blvd Ste E, Pullman, WA 99163, USA: Tomlab Optimization Inc.

Ryan, C. A. (1990). Protease inhibitors in plants: genes for improving defenses against insects and pathogens. *Annual review of phytopathology*, *28*(1), 425-449.

Schimke, R. T., & Doyle, D. (1970). Control of enzyme levels in animal tissues. *Annual review of biochemistry*, *39*(1), 929-976.

Shewry, P. R. (2003). Tuber storage proteins. *Annals of Botany*, *91*(7), 755–769.

Simon, M., & Azam, F. (1989). Protein content and protein synthesis rates of planktonic marine bacteria. *Marine ecology progress series. Oldendorf*, *51*(3), 201-213.

Sowokinos, J. R. (2001). Biochemical and molecular control of cold-induced sweetening in potatoes. *American Journal of Potato Research*, *78*(3), 221-236.

Strickland, J. a., Orr, G. L., & Walsh, T. a. (1995). Inhibition of *Diabrotica* Larval Growth by Patatin, the Lipid Acyl Hydrolase from Potato Tubers. *Plant Physiology*, *109*(2), 667–674.

Strnadova, M., Prasad, R., Kučerová, H., & Chaloupka, J. (1986). Effect of temperature on growth and protein turnover in *Bacillus megaterium*. *Journal of basic microbiology*, *26*(5), 289-298.

Suh, S. G., Peterson, J. E., Stiekema, W. J., & Hannapel, D. J. (1990). Purification and characterization of the 22-kilodalton potato tuber proteins. *Plant physiology*, *94*(1), 40-45.

Trewavas, A. (1972). Determination of the rates of protein synthesis and degradation in *Lemna minor*. *Plant physiology*, *49*(1), 40-46.

Viklund, G. Å. I., Olsson, K. M., Sjöholm, I. M., & Skog, K. I. (2008). Impact of harvest year on amino acids and sugars in potatoes and effect on acrylamide formation during frying. *Journal of Agricultural and Food Chemistry*, *56*(15), 6180–6184.

Weerstatistieken De Bilt – 2015/2016. Retrieved from https://www.weerstatistieken.nl on 21/10/2016.

Willigenburg van, L. G. (2016). Systems and Control Theory Part 2: Optimal Control.

Willigenburg van, L. G., & Ooteghem van, R. J. C. (2015). Modelling Dynamic Systems Course Guide.

World Congress on Root and Tuber Crops, Participants – GCP21 (2016, January). Retrieved from http://www.gcp21.org/wcrtc/

Appendix A: Optimal Control Parameters

Supplemental Table 1. Parameters used for the determination of the profit of protein sales.

Parameter	Description	Value	Units	Reference
C_{pa}	Specific heat capacity of air	1005	$[J * kg^{-1} * K^{-1}]$	Engineering Toolbox. *Air Properties.* Retrieved from http://www.engineeringtoolbox.com/ air-properties-d_156.html.
C_{pp}	Specific heat capacity of potato	3670	$[J * kg^{-1} * K^{-1}]$	Engineering Toolbox. *Food and Foodstuff – Specific Heats.* Retrieved from http://www.engineeringtoolbox.com/ specific-heat-capacity-food-d_295.html.
M_{bulk}	Weight of potato bulk	235 000	$[kg]$	Mol (2015)
V_{air}	Volume of air in storage facility	582.36	$[m^3]$	Mol (2015)
ρ_{air}	Density of air	1.208	$[kg * m^{-3}]$	Engineering Toolbox. *Air Properties.* Retrieved from http://www.engineeringtoolbox.com/ air-properties-d_156.html.
P_{prot}	Protein selling price	20-500	$[€ * kg \ protein^{-1}]$	https://www.alibaba.com/showroom /potato-protein.html
P_E	Energy price	0.20	$[€ * kWh^{-1}]$	https://www.milieucentraal.nl/energi e-besparen/snel-besparen/grip-op-je-energierekening/energieprijzen/

The specific heat capacity of the potatoes is determined by taking the average of the specific heat capacities of its components. Both the specific heat capacity and the density of air are dependent on the air temperature, therefore averages over the investigated temperature range were taken.

The protein price is a very rough estimate, and depends on the quality and purity of the protein.

Appendix B: Outdoor Temperature 2015-2016
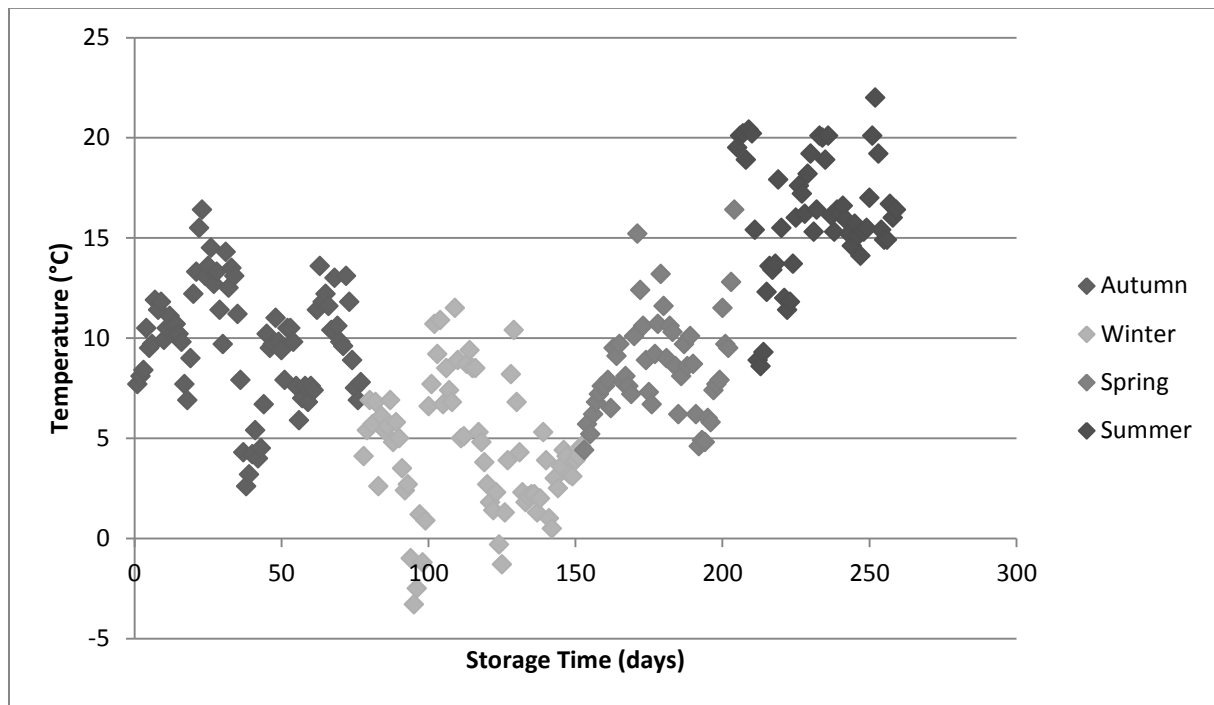
Supplemental Figure 1. Outdoor temperature of 2015-2016. Data starts on 16 October 2015 (Weerstatistieken De Bilt, 2015).

Appendix C: Experimental Data

Supplemental Table 2. Experimental data Miss Malina cultivar, including standard deviations. Values are averages of four measurements. SD=standard deviation. Values shown are in [g/kg fresh weight].

Week	Total Protein	Total Protein SD	Protease Inhibitor	Protease Inhibitor SD	Patatin	Patatin SD
1	2.33	0.024	0.82	0.022	0.42	0.015
2	2.54	0.287	0.89	0.093	0.45	0.083
3	2.15	0.233	0.80	0.090	0.36	0.034
4	2.20	0.091	0.77	0.015	0.37	0.021
5	2.00	0.039	0.71	0.012	0.33	0.010
6	1.96	0.08	0.69	0.019	0.34	0.033
7	2.00	0.325	0.71	0.115	0.36	0.084
8	2.09	0.092	0.72	0.025	0.37	0.001
9	2.21	0.207	0.76	0.077	0.39	0.066
10	2.89	0.061	0.96	0.005	0.52	0.004
11	1.96	0.066	0.68	0.011	0.31	0.017
12	2.23	0.111	0.77	0.036	0.40	0.028
13	2.88	0.174	0.95	0.070	0.54	0.064
14	2.10	0.106	0.71	0.034	0.35	0.019
15	2.21	0.067	0.75	0.017	0.39	0.008
16	1.90	0.003	0.66	0.003	0.31	0.017
17	2.27	0.215	0.75	0.068	0.39	0.061
18	2.02	0.201	0.68	0.054	0.36	0.064
19	2.11	0.089	0.72	0.022	0.37	0.008
20	2.26	0.176	0.77	0.052	0.39	0.045
21	2.07	0.157	0.73	0.064	0.34	0.015
22	1.64	0.192	0.57	0.078	0.25	0.037
23	2.01	0.082	0.70	0.036	0.33	0.033
24	2.03	0.057	0.72	0.022	0.36	0.005
25	2.10	0.11	0.72	0.041	0.36	0.020
26	1.91	0.083	0.68	0.035	0.31	0.003
27	2.08	0.071	0.71	0.001	0.33	0.033
28	1.81	0.241	0.63	0.076	0.29	0.070
29	1.99	0.24	0.68	0.084	0.33	0.047
30	2.07	0.079	0.70	0.028	0.33	0.038
31	1.90	0.106	0.67	0.043	0.32	0.010
32	1.76	0.082	0.62	0.039	0.26	0.028
33	2.07	0.197	0.70	0.062	0.34	0.046
34	2.04	0.234	0.70	0.077	0.31	0.063
35	1.90	0.147	0.66	0.042	0.31	0.035

Supplemental Table 3. Experimental data Agria cultivar, including standard deviations. Values are averages of four measurements. SD=standard deviation. Values shown are in [g/kg fresh weight].

Week	Total Protein	Total Protein SD	Protease Inhibitor	Protease Inhibitor SD	Patatin	Patatin SD
1	1.4815	0.2890	0.6747	0.0689	0.2314	0.0466
2	1.3255	0.3030	0.6097	0.1475	0.1892	0.0420
3	1.5185	0.1020	0.6698	0.0140	0.2411	0.0117
4	1.8525	0.1440	0.7622	0.0344	0.2722	0.0313
5	1.5135	0.0350	0.6331	0.0048	0.2151	0.0141
6	1.6585	0.2420	0.6909	0.1073	0.2832	0.0551
7	1.2975	0.1770	0.5477	0.0671	0.2066	0.0352
8	1.6995	0.0750	0.7211	0.0242	0.2716	0.0230
9	1.7085	0.0960	0.6649	0.0171	0.2825	0.0116
10	1.8265	0.1460	0.7212	0.0658	0.2649	0.0448
11	1.5905	0.0560	0.6001	0.0124	0.2482	0.0334
12	1.6705	0.0160	0.6246	0.0074	0.2782	0.0052
13	2.0495	0.1410	0.7568	0.0602	0.3441	0.0349
14	1.5415	0.0950	0.5786	0.0449	0.2286	0.0028
15	1.5615	0.1770	0.5716	0.0386	0.2440	0.0446
16	1.5285	0.0700	0.5504	0.0458	0.2345	0.0245
17	1.5165	0.0800	0.5526	0.0413	0.2278	0.0181
18	1.8845	0.1740	0.6386	0.0365	0.3183	0.0275
19	1.5995	0.0970	0.5544	0.0376	0.2614	0.0406
20	1.5045	0.1660	0.5428	0.0502	0.2332	0.0324
21	1.5165	0.1000	0.5452	0.0480	0.2398	0.0302
22	1.4195	0.0190	0.5266	0.0099	0.2143	0.0014
23	1.6045	0.0200	0.5561	0.0206	0.2560	0.0120
24	1.6415	0.0770	0.5886	0.0129	0.2648	0.0239
25	1.8385	0.1260	0.6461	0.0434	0.3290	0.0344
26	1.7415	0.0670	0.5418	0.0278	0.2834	0.0222
27	1.5875	0.0530	0.5585	0.0123	0.2448	0.0177

Supplemental Table 3. Experimental data Agria cultivar, including standard deviations. Values are averages of four measurements. SD=standard deviation. Values shown are in [g/kg fresh weight].

Appendix D: Moisture Loss Model

According to Lukasse *et al.* (2007), the moisture loss in potatoes is equivalent to the evaporation from the surface of the potatoes. The evaporation is driven by the moisture concentration difference between the surface area and the surrounding air. Supplemental Figure 2 shows an overview of the process. At the surface of the potato, the air is saturated.
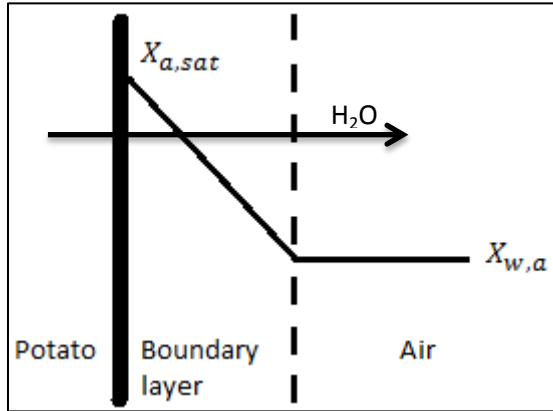
Supplemental Figure 2. Schematic overview of evaporation process on potato surface.

The rate of evaporation can be given by the following equation:

$$r_{evap} = K_{evap} * \frac{A_p}{\rho_{bulk}} * (X_{a,sat} - X_{w,a})$$

The amount of water that leaves the potatoes is equal to the amount that enters the surrounding air:

$$-\frac{dX_p}{dt} = \frac{dX_{w,a}}{dt} = r_{evap}$$

The saturated vapour concentration is assumed to be dependent on the air temperature as follows:

$$X_{a,sat} = \frac{e^{23.5 - \frac{3991}{T_{air} + 234}}}{\frac{\rho_{air} R}{M_w} (T_{air} + 273.15)}$$

Supplemental Table 4. Parameters used in calculations of moisture losses.

| Parameter/Variable	Description	Value [units] (Lukasse *et al.* 2007)
A_p	Specific surface area of potatoes	$60 \left[\frac{m^2}{m^3 \, bulk} \right]$
K_{evap}	Evaporation coefficient	$0.14 \, [kg \, air * m^{-2} \, potato * h^{-1}]$
M_w	Molar mass of water	$18.0115 \, [g \, mole^{-1}]$
T_{air}	Air temperature	0-30 [°C]
$X_{a,sat}$	Saturated vapour concentration	$-[g \, water * kg \, air^{-1}]$
$X_{w,a}$	Water vapour content of air	$-[g \, water * kg \, air^{-1}]$
r_{evap}	Rate of moisture evaporation	$-[g \, water * kg \, potato^{-1} * h^{-1}]$
ρ_{air}	Average air density at interval	$1.208 \, [kg * m^{-3}]$
ρ_{bulk}	Bulk density	$670 \, [kg * m^{-3}]$
R	Universal gas constant	$8.31446 \, [J * mol^{-1} * °C^{-1}]$

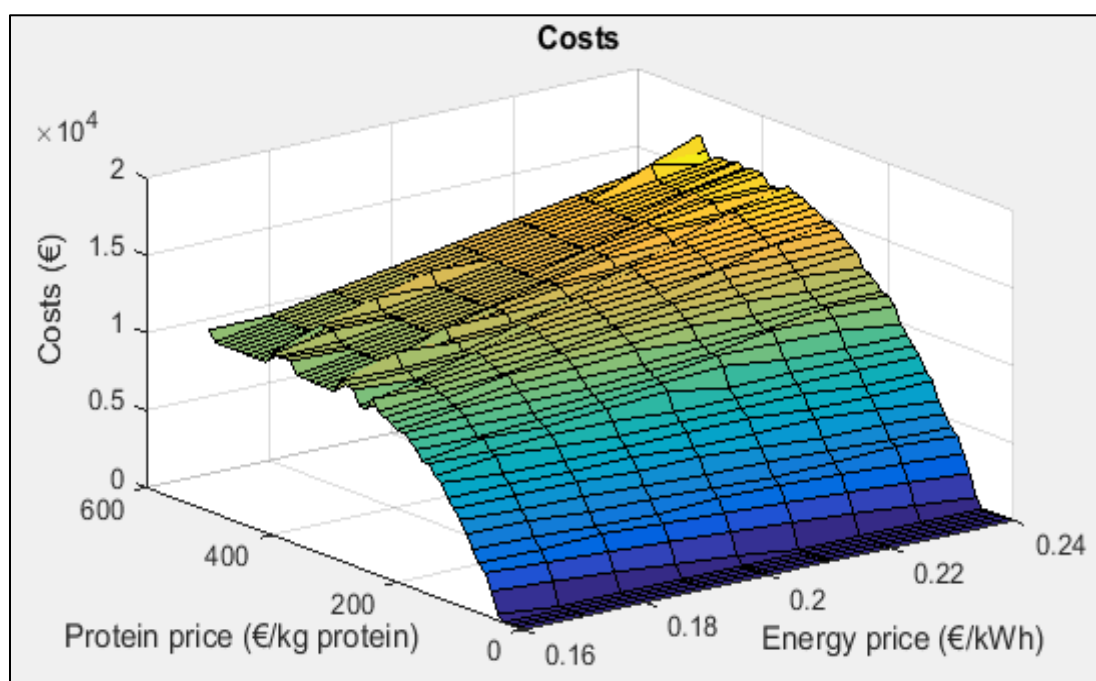## Appendix E: Profit, Income, and Costs
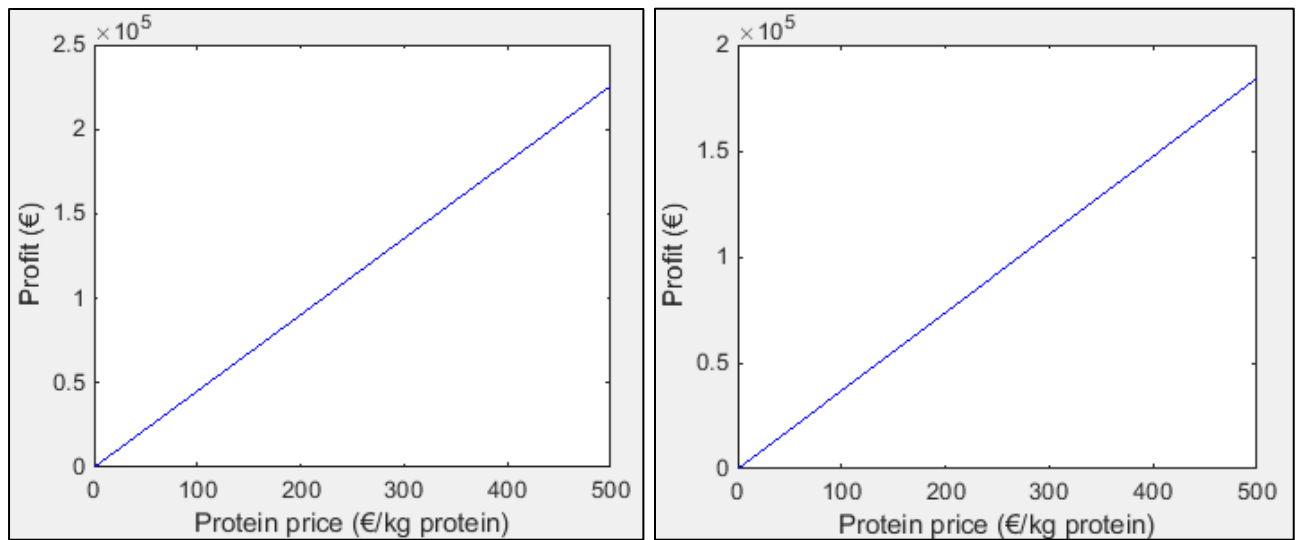
Supplemental Figure 3. Revenues without energy consumption. Left diagram shows Miss Malina profits, right diagram shows Agria profits.
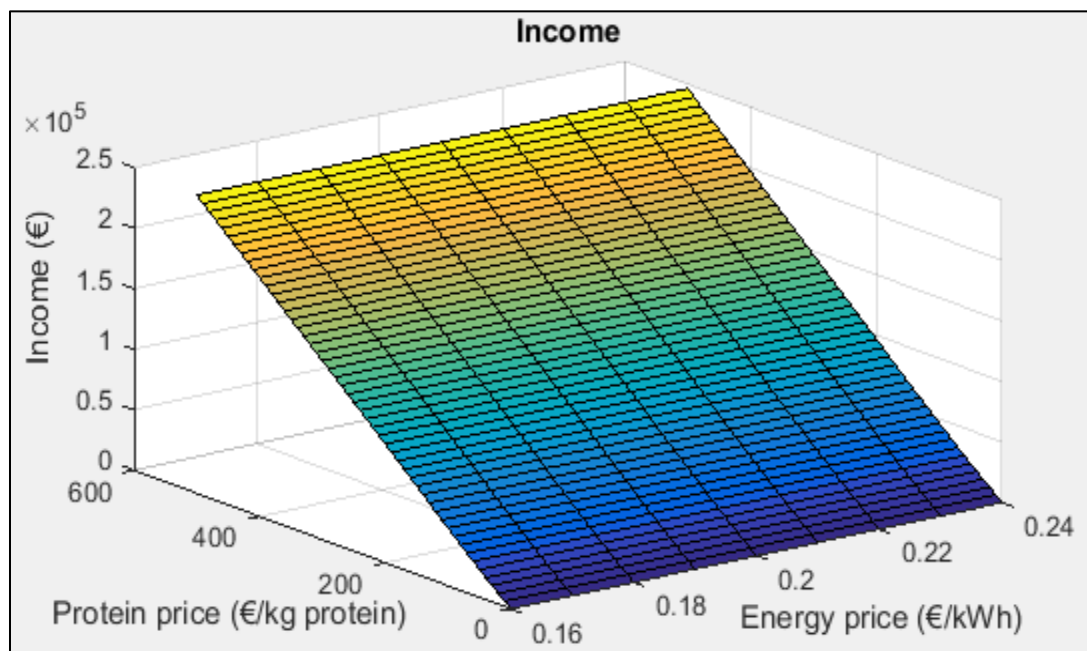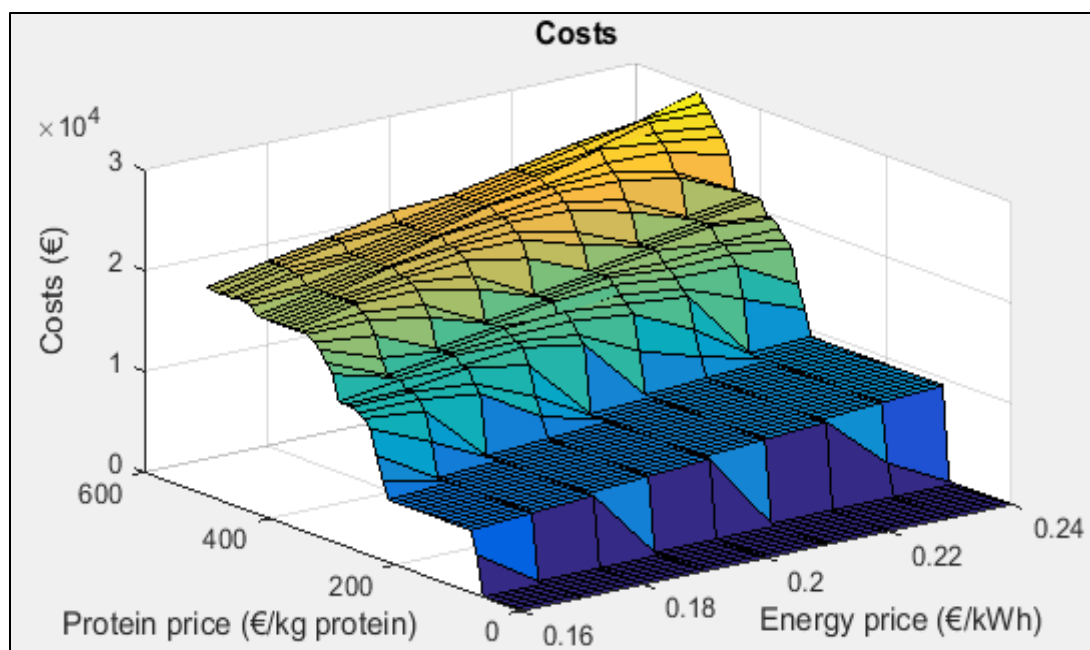

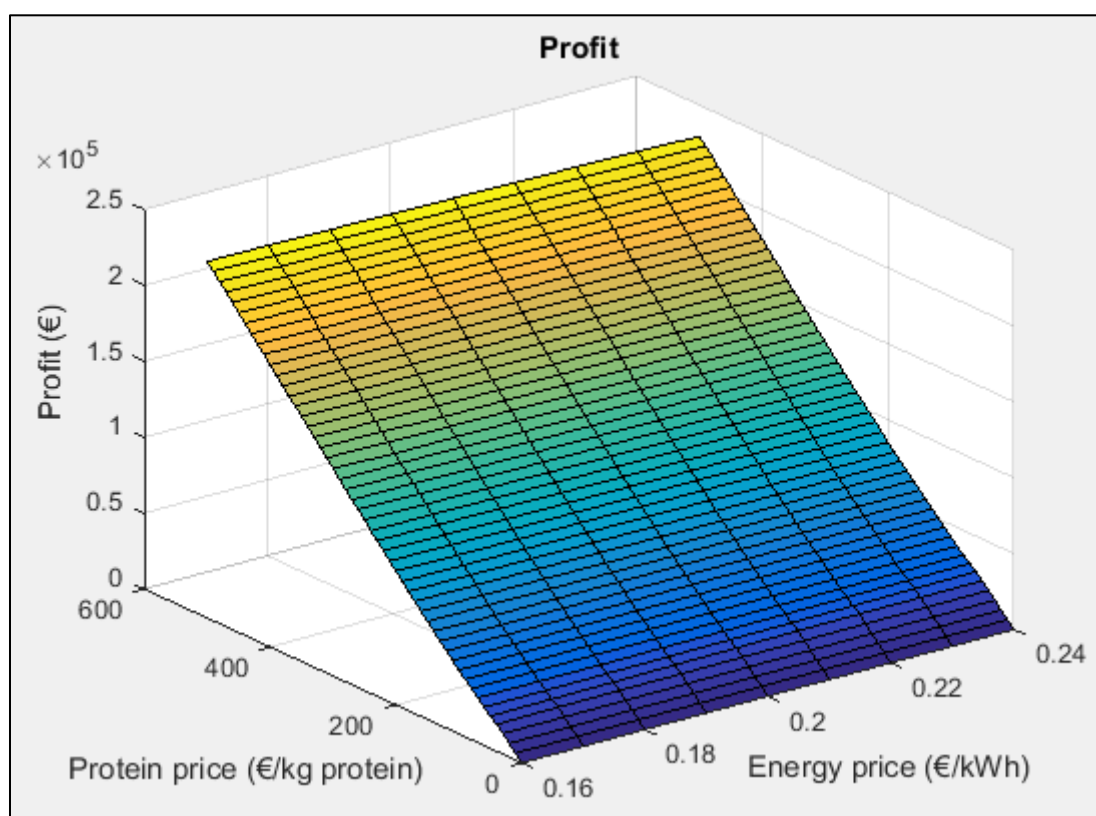**Supplemental Figure 4. Gross income Miss Malina potato sales with cooling.**
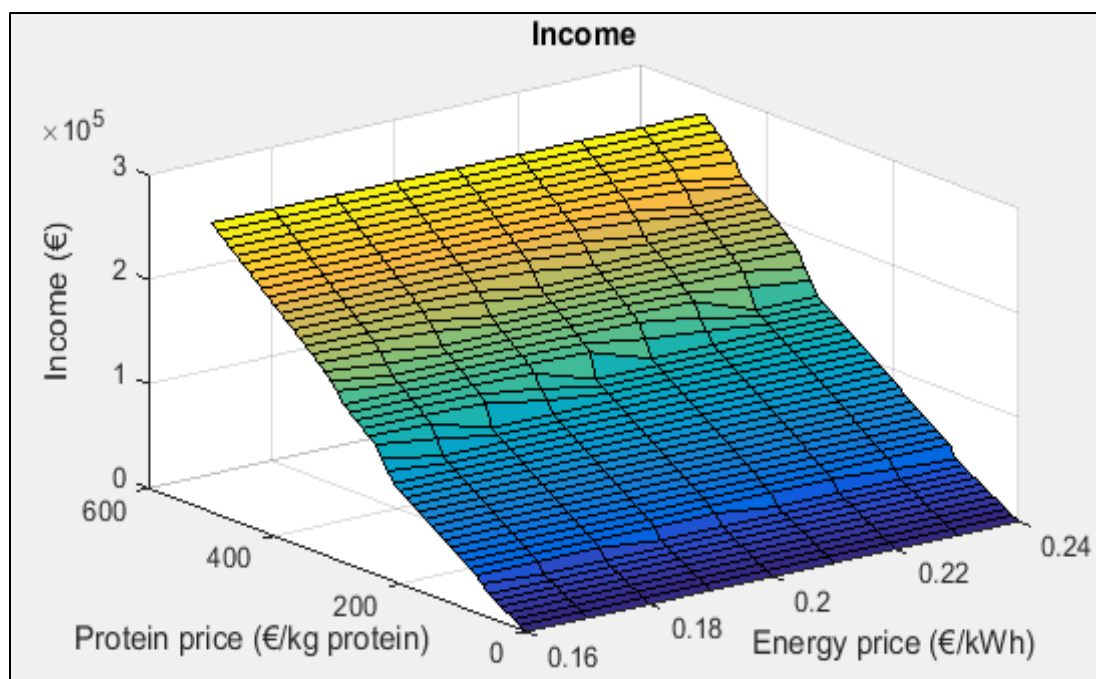
Supplemental Figure 5. Gross income Agria potato sales with heating.

Supplemental Figure 6. Costs of cooling Miss Malina potatoes.
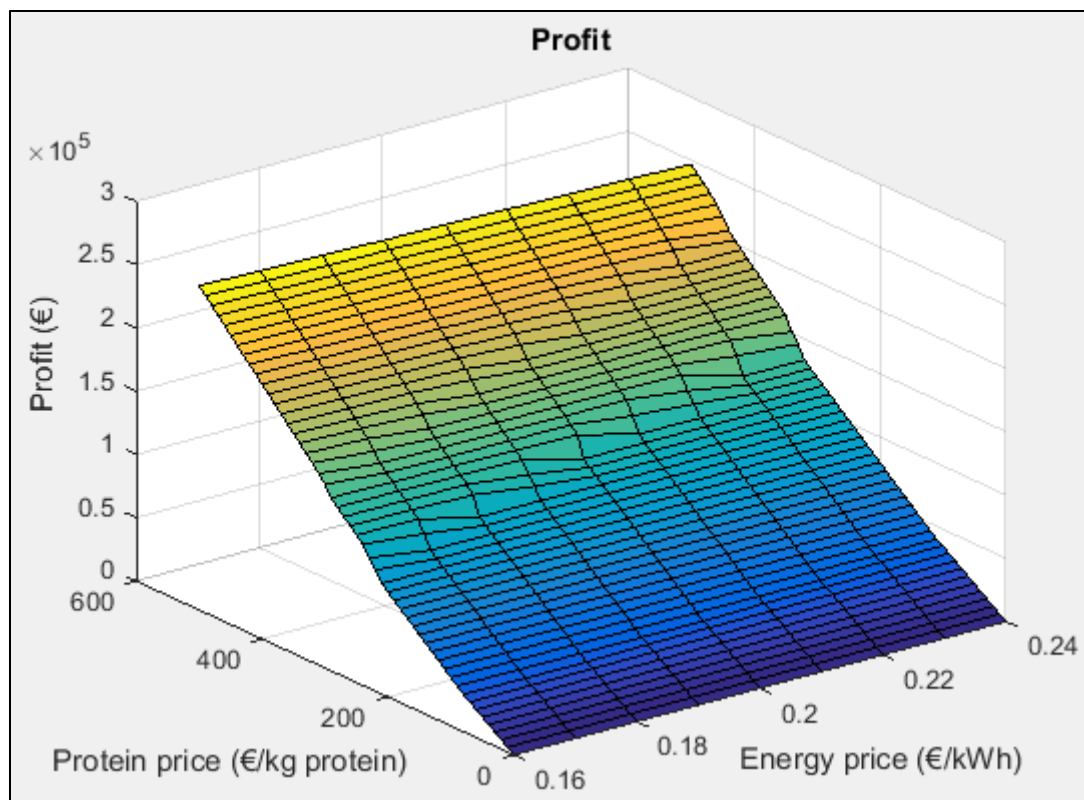
Supplemental Figure 7. Costs of heating Agria potatoes.

Supplemental Figure 8. Net profit after storage of Miss Malina potatoes.

Supplemental Figure 9. Net profit after storage of Agria potatoes.