for the degree of

Doctor of Philosophy

An intelligent strategy for endurance training based on a virtual lactate sensor

by

Urtats Etxegarai Susaeta

Under the guidance of

Dr. Eva Portillo Pérez

Dr. Jon Irazusta Astiazaran

Department of Systems Engineering and Automatics

University of the Basque Country

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

July 28, 2020

pausu bat, nire ametsa, nire hitza

Preface

Recherché: sought out with care.

This research, as any other, is exploration, discovery in essence.

And I find that this discovery process, despite being framed as a specific project, is mostly about ourselves. About expanding our intuition, our senses, overcoming our biases, discovering new perspectives and new tools. Improvements that eventually lead to better ways to approach and interpret whatever it is in front of us, in our never-ending endeavour of navigating through the unknown.

Using a language more fitted to the present thesis, it is about improving our abilities to filter, transform, represent and interpret the information about our context and make better decisions that would help fulfil or reach our goals. The main perspective of this thesis being the generalist point of view of system engineering.

A maxim of this area is that getting proper awareness of the foundations of an interesting and relevant problem is at least as important as having techniques and tools to solve a problem. E.g. with the objective of climbing the highest mountain, someone may try to climb the mountain that looks higher from his/her vantage point. And discover ex-post that the climbed mountain is not he highest but a close one. This analogy is not alien to anyone participating in some kind of research.

So expending the right energy in the problem-space identification is a must if we want to maximize the chances of reaching to the desired goal. The more complex and uncertain the problem, the truer this statement. And research is arguably one of the most uncertain areas, as we said before, ”sought out with care”.

Using this generalist point of view, this thesis attempts to provide novel solutions to: First, solve a specific problem that will help practitioners (coaches and athletes in this case) make better decisions for training endurance sports. Second, formalize the process behind this solution so that we provide a framework to help guide the steps when new similar problems appear in the future.

This dissertation is the exercise of accumulating, gathering, organizing, conceptualizing, synthesizing and making explicit the work done these last years.

Here it is my humble attempt to help science advance.

Acknowledgment

To family, friends, masters, colleagues, institutions and the huge amount of (usually invisible) people that shared the same path and/or made this possible:

To UPV-EHU (University of the Basque Country), specially to the Automatic Control and System Engineering, Physiology and Physical Education and Sport Departments. Thank you for bringing me the opportunity to invest my time in this exciting research.

To Campus Deportivo LTD. Without your work and trust this beautiful project would have been unfeasible. Thank you.

To the Basque Government. Without your support I would have not been able to enjoy working full time on this research. Thank you.

To AUT (Auckland University of Technology), KEDRI (Knowledge Engineering and Discovery Research Institute) and the European Commissions Erasmus Mundus PANTHER (Pacific Atlantic Network for Technical Higher Education and Research). Thank you for providing with so many important lessons and allow me to discover the beautiful country that New Zealand is.

To RWRI (Real World Risk Institute) and Nassim Nicholas Taleb. The lessons you gave me have been essential to improve the much desired rigour and care of this thesis. Thank you.

To Athletic Club Bilbao's Medical Services. Together we have been able to apply some of this thesis' lessons into an elite sport environment. Thank you for putting your trust in me.

But most importantly, to the people that compose these institutions. Thank you very much for your help. I very much enjoyed to invest time with you during this adventure.

To Eva, Jon, Ama, Aita, Aitzol, Neka, Iratxe, Nere, Asier, Adro, Iker, Chris, Iñaki, Jon, Eka, Iñi, Juan, Igortxu, Aritz, Alain, Ainho, Tito, Aitor, Alex, Ibai, Lucien, Elisa, Jacopo, familia CAG...

If I see something, its because I stand in the shoulders of giants, your shoulders.

My most sincere gratitude.

Abstract

In this thesis, a first fully operational *virtual lactate threshold sensor* was created for recreational runners based on a simple heuristic. From the additional knowledge gained, % of maximum heart rate at a given speed also showed the potential to be used in synergy with lactate threshold. This way, a so demanded operational solution to help the training of recreational runners was designed. Moreover, the *Lactatus* software was created to guide, ease the athletes' lactate threshold estimation process and implement the additional information in their training decision-making process. This way, the work of this thesis is made tangible, widely available and usable to recreational runners.

This solution grew from the creation and formalization of a strategy to apply machine learning to complex phenomena, an important contribution of this thesis. This strategy combined an iterative meta-process and a *satisficing* approach to deal with the problem boundary discovery and reduce the problem complexity. A methodology was created to define the collection and validation of the experiments. Then, the design of the *virtual lactate threshold sensor* was divided into three steps: *context characterization*, *content representation* and *next step decision*. The formalization of this methodology and a modification of *next step decision* are novel contributions. Additionally, several novel techniques are also used, including a standardization of the temporal axis, a modified stratified sampling and a computational algorithm to discover the inherent noise that input and output features may contain. This way, a robust strategy and methodology is created to design *virtual sensors* for problems with similar characteristics.

The application of this methodology led to an important conclusion. Concretely, Dmax individual lactate threshold's intrinsic error analysis showed that a higher accuracy of the *virtual lactate threshold sensor* was unnecessary and even non-characterizable. This fact manifested the importance of understanding the variability of the output features with respect to the input errors.

The computational algorithm could also be used to evaluate other lactate threshold protocols in order to quantitatively address their reliability. This may allow to make an objective cross-comparison of the accuracy of different lactate threshold protocols, something that, to the best of our knowledge, is not well addressed in the literature.

One of the possible limitations of this solution is that our population is drawn from local running clubs. This means that it is possible that the recreational runner population here characterized may not be representative of recreational runners of other culture, ethnicity or different contexts. However, one of the main advantages of providing a simple solution is that, unlike other black-box models, it is easily reproducible and adjustable, meaning that we have set a common ground for other researchers to evaluate the impact of our proposal. In the best case scenario, future experiments done in other contexts will validate it. In the worst case scenario, we have provided an easy to follow methodology and a strong prior that will allow to adjust the *virtual lactate threshold sensor* and knowledge to individual characteristics of different populations.

Table of Contents

List of Figures

List of Tables

Acronyms and Definitions

LT: lactate threshold.

Operational LT problem: the problem that this thesis tries to solve, i.e. creating a more operational (available, usable and with low interference with training) lactate threshold estimation method.

Features: variables related to certain phenomena.

Use modes: how the lactate threshold may be used for training decision-making.

MLSS: Maximum Lactate Steady State.

MLSSc: lactate concentration at Maximum Lactate Steady State.

MLSSw: workload at Maximum Lactate Steady State.

OBLA: Onset of Blood Lactate Accumulation.

IAAF: International Association of Athletics Federations.

HW: Hardware.

SW: Software.

ML: machine learning.

Dmax: method that calculates the LT in the maximum distance between the lactate curve and the line between the first and the last lactate values.

HR: Heart Rate.

HRR: Heart Rate Recovery.

HRDP: Heart Rate Deflection Point.

HRRT: Heart Rate Recovery Threshold.

%HRmax: % of maximum heart rate at a given speed.

Vpeak: maximum velocity reached in the treadmill test.

RPE: rate of perceived exertion.

Chapter 1

Introduction

A problem well put is half solved - John Dewey

Figure 1.1: The highest Mount Aoraki looks smaller than Mount Tasman

1.1 Motivation of the work: the problem space

The motivation of the present project, as usually happens, was born from a social demand. Nowadays, there is a huge recreational runner population that wants to train for performance. These athletes have particular interest of assessing the evolution of their performance to help optimize their training. In this regard, features related to the intensity of exercise at aerobic/anaerobic transition are good indicators of performance in endurance sports. Lactate threshold (LT) is probably the most used one with this purpose. In fact, current recreational runners, despite their limited resources, pay a reasonable amount of money to estimate their LT in specialised centres (see Figure 1.2). Thus there is a huge interest in obtaining more operational ways to do it. Despite there are other indirect more or less operational ways to estimate it, non seemed to have achieved the sufficient accuracy to displace the traditional non-operational ones.

Figure 1.2: Current lactate threshold determination method: invasive non-autonomous

Traditionally, training for performance has been based mainly on heuristics and good practices. Sport science, armed with the idea of applying the scientific method in the context of sport, has tried to provide additional tools, such as the LT, to further enrich the toolbox of coaches and athletes. Physiologists on their part, have tried to get a deep understanding about the LT phenomenon, not only for sport related implications, but also because of the relevance that physical performance measurements have in human health.

However, as in many other scientific areas, crossing the line between theoretical research and practice is always difficult. In this regard, this thesis is born with the objective of, using the system engineering mindset, closing this gap on the long-standing problem of operational LT estimation, focused on the recreational runner population. Moreover, another important motivation is about making an incursion on solving sport related problems using data based approaches and forming methodologies that will help future researchers.

Going deeper into the social demand, in recent years, endurance sports have dramatically increased in popularity, specially long distance running events and triathlon. These events usually count with thousands of participants, up to 42000 inscriptions in the Brooklyn marathon, 35000 in New York and 60000 in Paris in the 2019 editions (Figure 1.3). Even more popular are other endurance events such as half marathons, 10 to 20 km races and middle and long distance triathlons. Looking at the 17 million of people who finished a running event and the 4 million people who participated in triathlons in USA [2] gives an idea of the volume of athletes who are participating in this kind of events.

Figure 1.3: Runners participating in Paris marathon

By Marriot Bonvoy Traveller

This big recreational runner population does not appear to be diminishing in the short term. Physical activity is becoming a priority and an important part of the modern western society. Enjoyment, mental and physical health and the personal fulfilment that comes from establishing and achievement of objectives... are among the multiple benefits that physical activity has on individuals [3]. This has pushed people to seek in sport a tool to face the modern ailments, i.e. a modern society that poses high levels of chronic stress on individuals and our increasingly sedentary life distanced from the physical activity to cite a few.

Among these sport practitioners, nowadays many tend to go one step further and start training for performance [4]. Training for performance is driven by multiple intrinsic and extrinsic motivations such as personal growth and fulfilment, social recognition, material rewards etc. This motivation usually allows the practitioner to engage and adhere into an activity such as sport, that often asks for short term sacrifices for long term benefits.

To do so, this population trains methodologically and uses every kind of available tool that may help them improve their performance. This means that the current endurance recreational runner demands and consumes any type of training method, equipment and/or tool that allows them to reach their performance goals. In this regard, LT is a well known physiological indicator with a demonstrated power to aid training decision-making [5; 6]. However, since currently the estimation of LT requires attending to specialized centres and laboratories, there is a huge demand for estimating it in a more operational way. This is the problem that is to be solved in this work and we hereafter refer to is as *the operational LT problem*.

In order to provide an operational LT for training decision-making that is applicable in the real world, the general needs of sport performance must be aligned with the specific qualities of our solution. For instance, if we blindly try to solve one of the specific demands such us creating a non-invasive LT method, without considering why a non-invasive tool is demanded, we risk solving only part of the problem and creating a solution that is not going to be applicable. Thus, prior to getting into the details of the LT, in the following section we put the LT in the context of training for sport performance.

1.1.1 Contextualization of the lactate threshold problem in sport performance: Why rich and operational information about the athlete is key for training

Sport performance is the manner in which sport competition is measured. The main tool that a coach or an athlete can use to improve sport performance is purposely training to develop the qualities required for a desired discipline.

Super-compensation (see Figure 1.4) is a characteristic of the human body by which in the face of an external stimulus, it adapts or adjusts itself to a higher level of fitness and performance capacity with respect to its prior status. Taking advantage of this principle, coaches' work is to select and introduce appropriate stimulus and recovery time that facilitate the improvement in the desired qualities for the sport of interest.

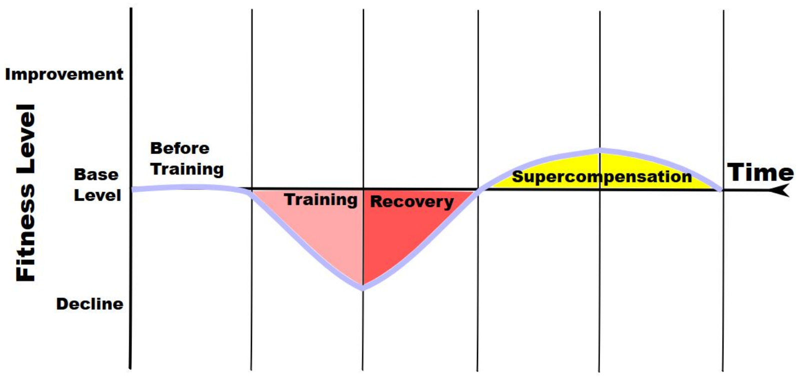


Figure 1.4: Super compensation

By Haus-Own work, CC BY-SA 3.0, https://commons.wikimedia.org/

These qualities are a complex mixture of tactical, technical, psychological, bio-mechanical, neuro-muscular, metabolic... abilities that the coach tries to maximize while minimizing injury and over-training.

The appropriateness of this stimulus, stands on several fundamental principles of sport performance:

- Overload: In order to elicit the human body's adaptive responses there must be a stimulus that goes over what the athlete is used to (in volume, intensity or density).
- Progression: The load applied must be progressively increasing to avoid both an insufficient stimulus that would not elicit adaptations or an excessive one that would in the worst case lead to an injury.
- Specificity: The adaptation of the human body is specific to the training stimulus applied. A simplistic example that illustrates this concept can be: if you train running fast you will

improve in running fast.

- Individualization: Every person responds differently to a stimulus, so this stimulus must be individualized to the characteristics of each athlete and its context.

- Reversibility: If certain stimulus is not applied during a extended period of time, the qualities start to revert to previous states.

Thus, training to improve sport performance is about selecting the appropriate combination of stimulus and recovery according to these principles. The coach then introduces this stimulus in form of a training program.

However, as it is known, human body is a complex system and there is an inherent uncertainty with respect to the response that the body-mind will have to the training stimulus. Moreover, the training is held in an highly uncertainty environment on which the athlete inevitably misses training days or is forced to change the objectives that were planned for the day. There are also specific dates for which the performance must be maximized, so the time factor is also considered. Training periodization and planning are related to how the coach selects the appropriate program with all these considerations in mind and for a specific athlete.

The current research on periodization and planing is based on a viewpoint that the aforementioned uncertainty of the training is an unavoidable part of the problem that must be considered and properly managed. From the scientific method perspective, the initial program is just an initial hypothesis that is adjusted on the go according to what it is observed using a priori established, heuristics, criteria and expertise to continually create a next best training guess. Due to the complex and ever changing nature of the context, the faster the feedback, the more appropriate the training stimulus will be for the current situation. Figure 1.5 represents this process from the systems perspective.

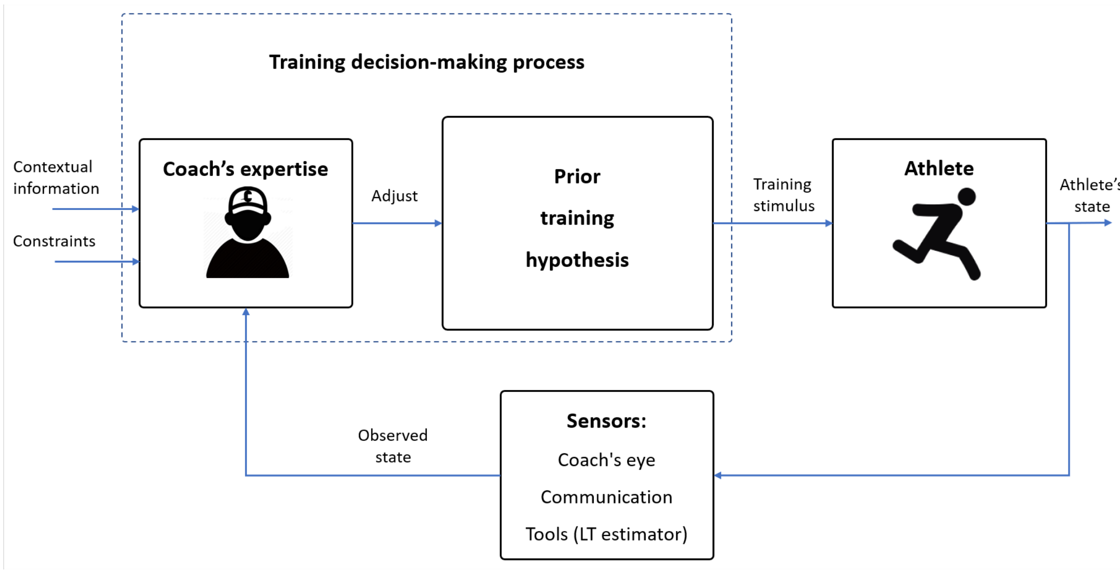


Figure 1.5: A systems view of appropriate training stimulus selection

The importance of having relevant information about the athlete and the surrounding context becomes manifest, meaning that every mean (coaches' eye, good communication, external tools etc) that allows to get valuable information about the athlete (LT in our case) has a major impact for improving their training.

However, sport performance is a highly optimized area which tries to make the most of the available resources with the objective of maximizing positive training adaptations. This means that the use of the resources is highly prioritized towards the most valuable.

In the case of the recreational runners, their economical, temporal and effort resources are even more limited compared to elite level athletes. For instance, recreational athletes expend their own money and usually their time is narrowed down to their free-time. Last but not least, the physical and mental effort needed are also limited. For instance, the effort done in a testing may have been better used in regular training.

Therefore, it is clear that the value of the method used to collect certain information (the LT in our case) is determined by both the relevance of the information collected and how operational is to collect and integrate it into the training decision-making process.

So, these two perspectives (the relevance and the operationality) will be used hereafter to analyse the importance of LT.

1.1.2 Lactate threshold, an indicator demanded by coaches and athletes

In the case of the LT, its relevance comes from its relation with the energy supply systems of the human body.

These energy supply systems are the mechanisms responsible for producing the energy that is used for physical activity. The effectiveness and efficiency of the energy production and how it is expressed in the executed task (a specific sport discipline in this case) is directly related to the performance that an athlete achieves.

Among other variables, the use of different energy supply mechanisms depends on the intensity and duration of the activity under execution (Figure 1.6). In long-duration exercises, as endurance sports, the oxidative or aerobic energy system is the main energy contributor as more powerful anaerobic systems are not sustainable in the long term without creating excessive fatigue.

Therefore, exercise intensity at the transition between the use of aerobic to anaerobic energy supply systems is determinant about the long term and sustainable energy production and plays a key role in the performance of the athlete [7; 5]. Thus, being able to obtain information about this transition of a particular athlete is of great interest for coaches as it can be used to enrich the training decision-making process.

The most relevant *use modes* of the aerobic to anaerobic transition zone are related to training monitoring (to evaluate the physiological or performance changes that training has caused) [5], training prescription (to aid in the prescription of training intensities) [6] and performance estimation (qualitative estimation of athlete level, pacing recommendation) [5].

However, despite the usefulness of the information about aforementioned transition zone is more than demonstrated, its characterization is not straightforward. The ample variety of terminology that has been used to name this transition zone exemplifies this difficulty. Moreover, sometimes the terminology is even contradictory and its meaning has even changed since it was

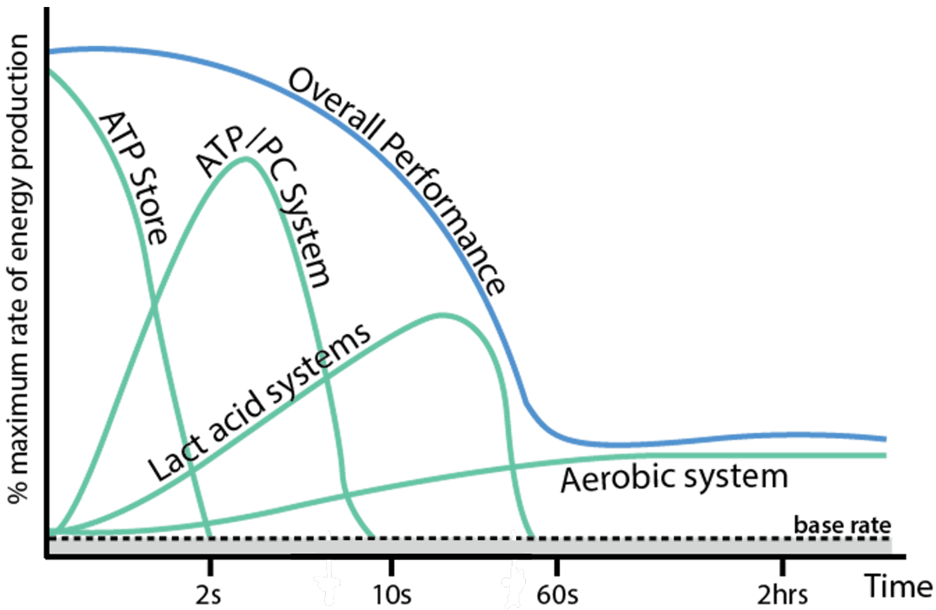


Figure 1.6: Energy supply systems

first coined [8; 9; 10; 11]. This clearly illustrates the fuzziness of the concept.

Consequently, multiple models, methodologies and techniques have been used so far to determine it. Moreover, there are multiple approaches that are still in use in practice, making it clear that there is no closed nor universal characterization of it. Among these approaches, the respiratory gases based methods and the blood lactate based approaches stand out.

Regarding the respiratory gases approach, multiple protocols are available. However, all these protocols have common limitations. First, the equipment needed to make the tests is very expensive, much more than the equipment needed for the lactate based approaches, which is already a big operational limitation. Moreover, the respiratory equipment may also interfere into the normal exercise of the athlete altering the results.

Lactate based methods on their part rely on much more accessible portable lactate measurement devices which have also a lower interference compared to respiratory approaches.This made the lactate based determination methods to the most extended ones in sport performance for anaerobic to aerobic transition zone characterization.

However, analyzing the lactate based approaches from the operationality perspective, they still require external equipment and the extraction of blood samples, which are inconvenient for frequent monitoring. Furthermore, most recreational runners do not have access to routine assessment of their physical fitness by the aforementioned equipment so they are not able to calculate LT without resorting to an expensive and specialized centre. Consequently, there is high demand for a more operational LT estimation.

Moreover, this interest exceeds the recreational runner population. Nowadays, the Spanish Athletic Federation uses lactate tests for the selection process of marathon runners to compete in international championships such us the Olympic Games, World Championships and European Championships.

However, as already mentioned, the lactate threshold problem is a limited resources problem where material, facilities, money, time, effort etc are finite. In the case of endurance runners, their limited resources problem materialises in a demand of tools and methods which are or provide:

- A non-invasive solution which avoids taking blood samples
- Easy-to-use tool
- Cost efficient tool, affordable solution for recreational runners/coaches
- Autonomy to the recreational runners/coaches
- A solution without the need of a new/additional wearable

Actually, the relevance and potential impact of an operational LT estimator is further strengthened by the interest that this research has risen among several important actors in the sport performance industry. These interest groups include: the huge recreational runner volunteers (more than 800) who inscribed for the experimental test performed under this research with little publicity efforts, the FIPSE (Consejo Superior de Deportes), the Campus Deportivo company that supported this research, the Department of Economic Development and Competitiveness of the Basque Government (Gaitek 2015), managers of Spanish Kayaking and Rowing Federation, Basque Rowing Federation, managers of the Basque Public School of Sports (Kirolene), athletic coaches etc

1.2 General objectives

The objectives of the present work are then defined with this operationality demand in mind. Creating a model that would estimate LT from easily obtainable input features, fits very well with the characteristics of this demand. As illustrated in Figure 1.7, virtual or soft sensing techniques are essentially this, approaches used to provide feasible and economical alternatives to costly or impractical physical measurement instrumentation. This approach uses information available from other measurements and process features to calculate and estimate the outcome of interest (LT in our case).

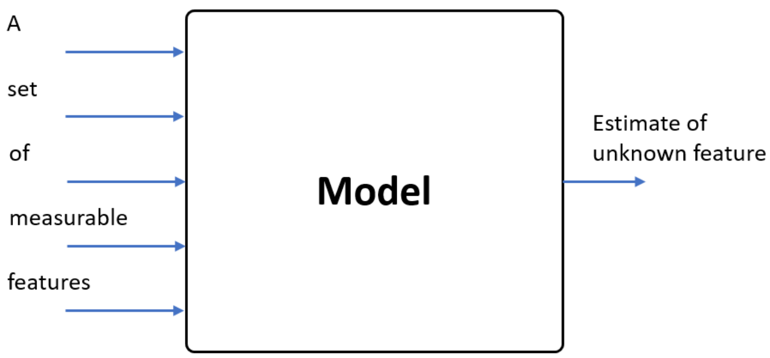


Figure 1.7: Virtual sensor concept

To do so, a variety of virtual sensing techniques have been proposed, while the vast majority of them fall into two major categories: analytical or empirical.

Since the relationship between LT and the other easily obtainable physiological features is complex and with multiple inter-dependencies, creating an analytical model that could characterize this relationship is not a viable approach.

In complex problems such this one, where the number of features involved in the process are bast and, even more importantly, the relationship between them are opaque, empirical virtual sensing approaches are, a priori, a more appropriate way to go.

Thus, creating a virtual sensor based on empirical data arises as an interesting approach with the potential to fulfil both requirements. Since the objective is to learn the relationship between certain inputs and output (LT), and the output is continuous, this problem can further be classified into the supervised-learning category done for regression purposes.

Therefore, we can formulate the main objective of the present work as:

Create a empirical virtual sensor to estimate the lactate threshold that can be easily integrated into the endurance recreational coach/athlete toolbox. The objective can be further divided in:

1. Providing an easy to integrate LT virtual sensor to help recreational running coaches and athletes in training decision making.
2. Gain knowledge about other physiological features that may be key performance indicators of endurance athletes, specially the easily measurable ones, and give guidance to integrate it into the training decision making process.
3. Create a methodology for applying virtual sensing techniques to solve problems related to sports so it could be extended to other future demands of this area.
4. Acquire know-how about the specific demands of the problem to be able in the future to extend the solution proposed to other type of users and/or disciplines.
5. Design a final prototype that demonstrates the validity of this proposal.

1.3 Structure

The present document is divided in five additional chapters:

Chapter 2 makes the state-of-the-art analysis of the LT determination methods used so far from an operational perspective. To do so, first we set and describe the operational qualities that are to be looked. Then, both the traditional LT determination methods and the operational attempts are analysed with these qualities in mind.

Chapter 3 determines the strategy and formalizes into a methodology with specific steps.

Chapter 4 is about designing the *virtual LT sensor* according to the strategy and methodologies established in Chapter 3.

Chapter 5 deals with the implementation of the designed *virtual LT sensor* and acquired additional knowledge for aiding the training decision-making of coaches and athletes. To do so, the *Lactatus* software (SW) is created to work as guide to the athlete interested in using and applying the knowledge of this thesis.

Chapter 6 gather the main conclusions and contributions of this work. Additionally it provides suggestions and possible directions for future research.

Chapter 2

Lactate threshold in training decision-making: value of current arts & direction towards and operational solution

To add value to others, one must first value others - John C. Maxwell

Figure 2.1: Giving a closer look to the landscape

This chapter deals with the analysis of the state of the art with regards to the consolidated LT methods and the operational solutions that have been proposed so far as alternative. In order to evaluate the state-of-the-art from the operational perspective and in a systematic way, we create a tool based in the concept of added value. This tool is then used to place all the proposed approaches (consolidated and alternatives) according to the operational qualities that are relevant for the objectives of this thesis.

As stated in chapter 1, this state of the art analysis highlights that the relevance of the LT comes from its relation with the aerobic to anaerobic energy supply transition and that, due to the fuzzy nature of this threshold, there are multiple LT determination methods available in the literature. Additionally, we also present the idea that the relevance of the information and its operationality (in the recreational runners context) are the two mayor characteristics that make a LT determination method more or less valuable for the user (recreational runners in this case). This observation evidences that, there are as many values as LT determinations are.

The first purpose of this chapter is to elucidate and select, among the multiple LT determination methods, the reference that will be used as the output labels of the (supervised learning based) virtual sensor. The second purpose of this chapter is to analyse the most relevant solutions that have been proposed so far trying to to provide some more operational LT determination method compared to the consolidated ones.

To do so, as illustrated in Figure 2.2, and prior to digging deep into the literature, we set the qualities that will be used as standard by which the *value* of every LT determination method will be judged.

Then, we present the multiple different ways that historically have been used to characterize the LT concept. Using the previously defined criteria, we map the value of both the consolidated approaches and the operational attempts that have been made so far to surpass some of the former's limitations. This allows to have an overview of the current arts space from the recreational runners perspective.

Finally, from this analysis, we conclude which is the reference LT determination method that will be used hereafter and we adjust and detail the objectives stated in the previous chapter according to the findings of this one.

2.1 Value determinants of lactate threshold methods for recreational runners training: Desired qualities and value mapping

As we saw in chapter 1, the value of integrating certain information in training decision-making is related to first, the relevance of the information collected and second, the operational burden that collecting and integrating it carries.

However, the value is related to how the inherent qualities of the method (quality of the information and operational qualities in this case) unfold in the context of application, i.e. the recreational runners context. That is to say, certain LT determination method may pose a high value for an elite runner because of their higher economical resources and be of zero value for a

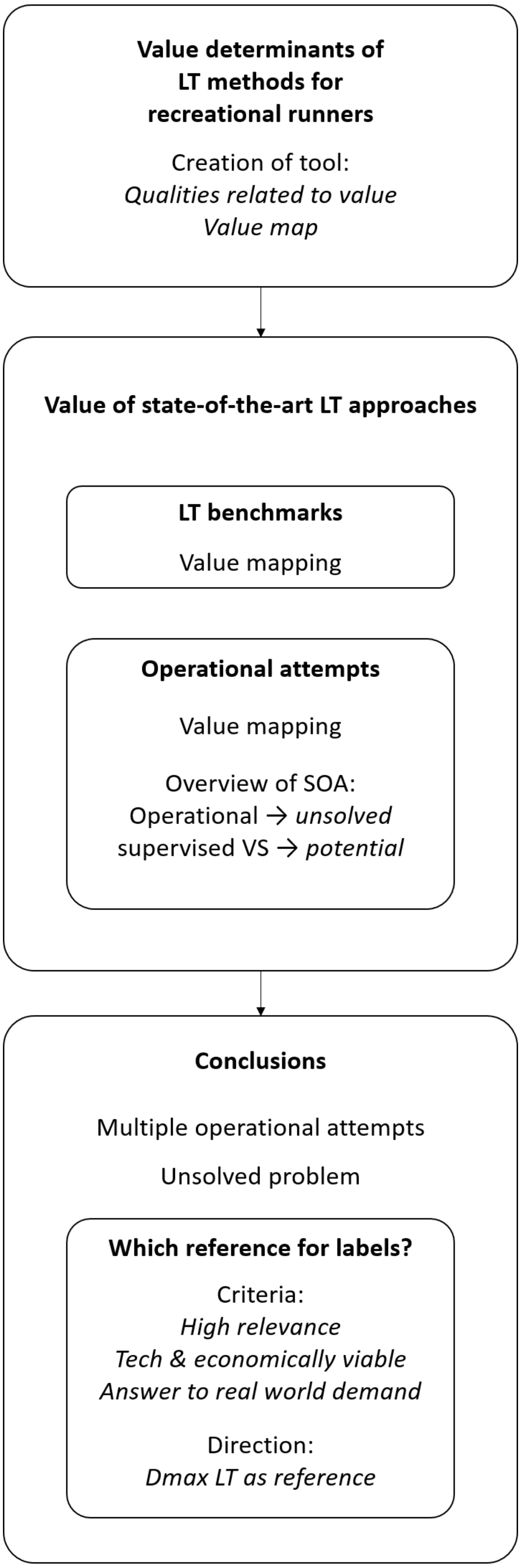


Figure 2.2: Overview of Chapter 2: state-of-the-art

Abbreviations: SOA, State of the art; VS, virtual sensor; LT, Lactate threshold recreational runner. Figure 2.3 represents this idea and shows how both these general attributes combine in our context to determine its value.

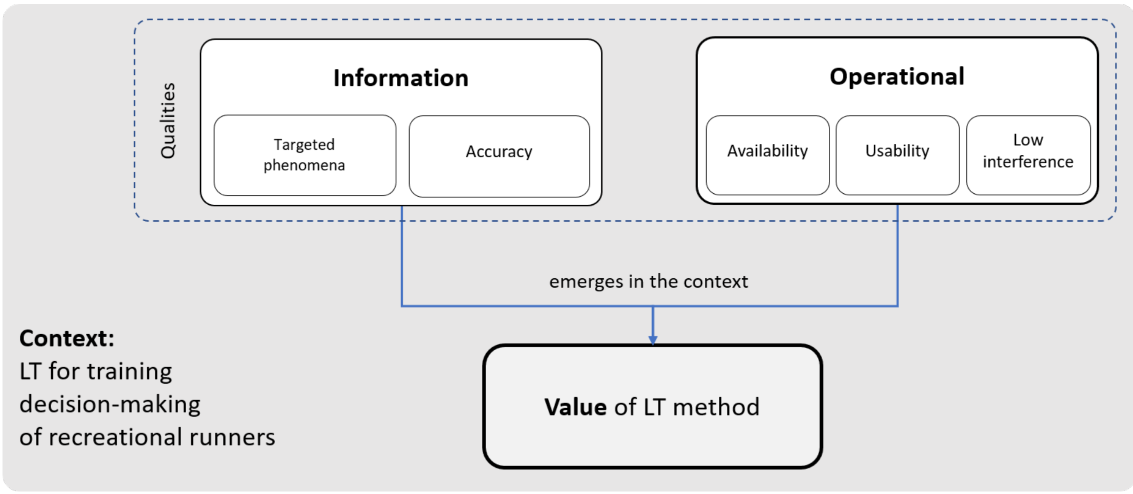


Figure 2.3: Value determination from information relevance and operational perspectives

So, clearly defining the qualities by which the value of LT determination methods are judged contextualizes them to the characteristics of the problem in hand. Based on the generalities and specificity's of the LT problem already analysed in chapter 1, in this section we formalize these qualities that will allow to sort the state-of-the-art approaches (both consolidated and operational tries) according to the perspective of creating a valuable LT estimator for recreational runners and systematically place them in a *value map*.

This gives an overview of the current arts with the glasses of operational estimation of LT for recreational runners. Consequently, the *value map* will facilitate to identify the *added value* space of our problem, delimit the *operational space* and pave the way for selecting the consolidated LT estimation method that will work as our *reference value*.

2.1.1 Qualities that determine the value of a lactate threshold method: Breaking down 'information relevance' & 'operationality'

In chapter 1, we stated that the relevance of the information that a LT determination method characterizes and its operationality are the two major qualities related to the value provided to recreational runners. Here, we go deeper into this reasoning so that we can pose the detailed qualities that are important for our context of application and by which the state of the art is going to be analyzed in the next section.

The value of a method comes from its ability to be used by the recreational runners in the three *use modes* previously seen. As already mentioned, there are many degrees of fulfilment of these *use modes*, as many as ways of determining the LT. If we look in more detail to the *use modes*, we can observe that each determination method has a different degree of fulfilment of each *use mode*:

LT has been successfully used for training monitoring purposes by assessing the physical condition from an endurance performance perspective [12; 5]. However, the monitoring may be done by different means. Sometimes a LT method is able to characterize information about an underlying relevant physiological phenomena such the maximum aerobic capacity. Other methods characterize a phenomena that is indirectly related to performance. Hence, the former is richer and more informative when using it for monitoring purposes compared to the latter.

Something equivalent happens when using it for training prescription. It is well known that lactate threshold is useful for training prescription [13; 12; 6]. Similarly to the monitoring use case, some methods are able to target an underlying relevant physiological phenomena (such as one that helps to identify training zones) that may be used to prescribe exercise intensity from this perspective [13; 14; 6]. However, other methods are not able to do so, limiting their training prescription capacities to those conclusions derived from the monitoring capacities [15].

The third use mode is for performance estimation. It has been observed that LT is a good predictor of performance, both in elite and recreational runners [16; 5]. Furthermore, nowadays it is fully demonstrated that the LT is more decisive for endurance sports performance than other variables such as the maximal oxygen uptake (VOmax) or the running economy [7; 17; 18]. However, the relationship with performance is also dependent on the methodology used. Some method have a strong direct relation with performance and others are just indirectly related. The former are strong enough so that can be used to prepare pacing strategies, rough athlete level evaluation... on the contrary, the latter methods only can be used to determine qualitative performance improvements [15].

Thus, in this section, it is reinforced that the value of a LT method is strongly related to the relevance of the information that characterizes from the target population perspective and that the operationality of the method.

The relevance of the information characterized can be further divided into lower level qualities such us: the targeted information (more or less relevant for the different use modes) and the accuracy with which it is characterized.

Additionally, in Chapter 1 we saw that multiple operational characteristics (non-invasive, easy-to-use, affordable, autonomous, without the need of a new/additional wearable...) are demanded. All these characteristics serve to higher operational purposes that are important to explicitly state here. To minimize the risk of solving only part of the problem by forgetting to comply with the general needs of sport performance. Thus, the operationality of the solution can be described by its qualities of: availability (location, requirement of specialized equipment, excessive associated costs etc.), interference with training (training time loss, testing associated fatigue) and usability (facilitating adherence etc).

Finally, as illustrated in Figure 2.3, the value of our solution is subjected to these qualities in the context of training decision-making of recreational runners.

In the following sections we will make a detailed analysis of the aforementioned qualities.

Qualities related with information relevance

The relevance of the information characterized by the LT determination method comes from the degree of fulfilment of two lower level qualities:

- Targeted phenomena: We have observed that the phenomena being characterized by a specific LT method impacts on the maximum potential value that can be obtained in the different use modes. Therefore, the relevance of the phenomena the LT method is characterizing is an important matter to evaluate the value that can provide.

- Accuracy: Even if the relevance of targeted phenomena is high, the accuracy with which it is being determined has a major importance. Not only that the margin of error of the estimation is important for the coach to be more or less certain about the conclusions, but also that not achieving a minimum accuracy may invalidate the estimation for training decision-making. The accuracy needed for the estimation is very dependant on the application and thus, it is an important quality on which to keep a close eye to evaluate the value of the LT determination method.

Therefore, these two qualities are to be used to determine the value of the state-of-the-art LT methods from the information relevance perspective.

Qualities related to operationality

The operationality of the LT determination method comes from the degree of fulfilment of three lower level qualities:

- Availability: The collection of certain beneficial information is usually subjected to some sort of specialized equipment, facilities, associated costs, location, need of expert personnel... The value provided by increasing the availability of a solution comes not only from the obvious reduction in the resources used, but also because increasing the availability indirectly facilitates the adherence to it and reduces the interference with the training as less time is invested in it. Increasing the availability of the tools and methods is therefore highly and directly related with the value of the solution and may be achieved by creating methods that do not depend on additional equipment (or equipment already available) nor require help from an expert.
- Usability: The usability of a solution is one of the main qualities that facilitates the adherence and consistency. Usability of a tool is described by the following five characteristics: effectiveness or how well the tool meets the specific task (including a clearly understandable use) ; efficiency or ability to meet the task with minimum effort (including the easiness of interpretation and communication of results); engaging or pleasant and satisfying to use; error tolerance or being robust to the errors that inevitably are introduced by the misuse of the tool (beyond the estimation error) and easiness to learn [19]. This quality is fundamental for adherence and consistent use of a LT method, specially, for repetitive tasks as continuous monitoring, where the easier collection and integration of information, the more consistently is going to be used and established as a habit. Therefore, increasing the qualities related to usability of a product that increase the adherence is also directly related with the value that the solution provides.
- Low interference with training: Whenever a test or measurement is done, there is a cost opportunity loss for training or recovering. The ability to provide reduced interference with

training is therefore a operational quality that a valuable LT method must have. This interference may come for the mere fact of performing the test, measurement etc or because the testing process creates residual fatigue that interferes with subsequent training sessions. Therefore, the ideal scenario is this in which a test or measurement is fully integrated with the training (a.k.a. invisible monitoring) becoming part of the training activity, without modifying nor hindering it. Thus, any improvement in this direction increases the operationality of the solution.

Therefore, these three qualities are to be used to determine the value the state-of-the-art LT methods from the operationality perspective.

2.1.2 Value map: emphasizing the value differences between lactate threshold methods

The value that a certain LT determination method has for recreational runners is thus directly proportional to the aforementioned qualities. The *value map* shown in Figure 2.4 illustrates this relationship and determines three regions according to the degree of value that is provided (low, medium high). This map will be henceforth used to place the different LT determination methods according to their value.

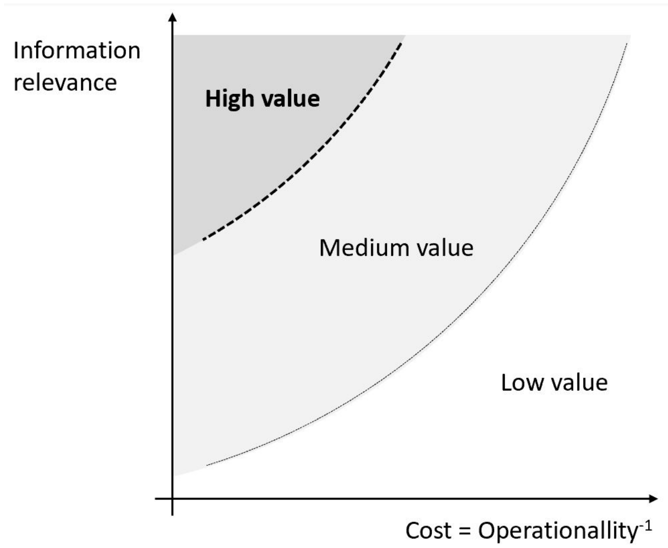


Figure 2.4: Value map: information relevance and operational qualities in the context of application

Since neither of the general qualities are quantifiable, this map allows to place the different state-of-the-art LT estimation methods according to the relative value they have.

In this regard, as we already mentioned in Chapter 1, there is currently no operational LT determination method available in the literature that suits the needs of recreational runners. Therefore, as illustrated in Figure 2.5, we can delimit the value space of the currently available arts for our objective population to a medium-low value region. This indirectly defines the region on which an added value solution should reside, i.e. the high value region.

The added value space is further narrowed by introducing the constraint of the cost limit that the recreational runners have, delimiting a both operational and added value space as illustrated in 2.5.

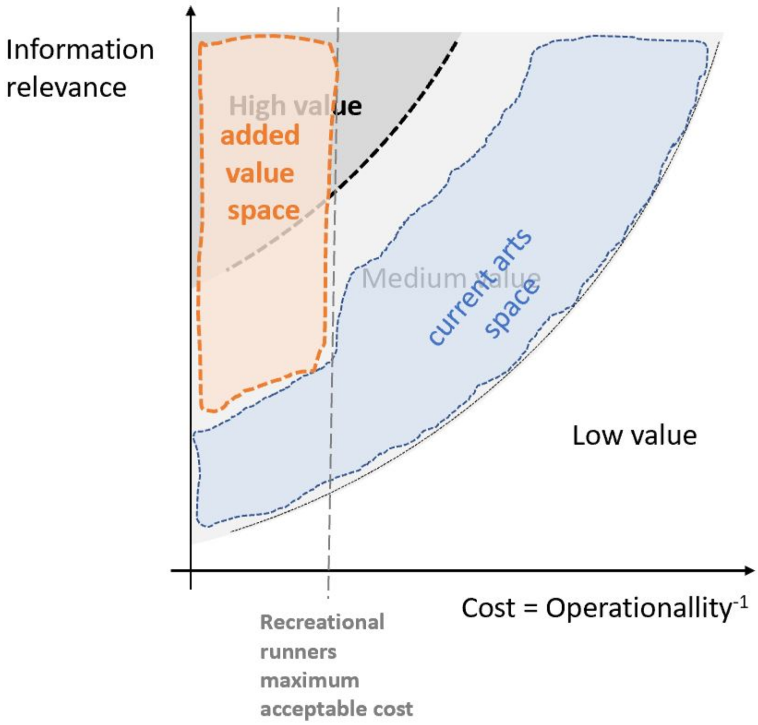


Figure 2.5: Value map: Added value space

Consequently, the *value map* will facilitate to identify the *added value* space of the research field, delimit the *operational space* and place the LT methods and operational solutions available in the literature and practice in the same space to better decide the consolidated method that will be used as reference for the design of our *virtual LT sensor*.

2.2 Value analysis of current art lactate threshold determination approaches: Consolidated approaches & operational attempts

In the previous section, we introduced the qualities that will be used to determine the value of a LT determination method for recreational runners and created a *value map* on which the different methods will be drawn.

In this section, we delve into a detailed analysis of the different ways that historically have been used to characterize the LT. We first determine the value of the consolidated approaches that are used in practice according to the qualities established in the previous section.

This analysis allows to have an overview of the state-of-the-art space, present the most important LT estimation methods and solutions that have been proposed so far and clarify the operational limitation that the consolidated LT methods have. Then, the most novel approaches that have tried to improve all or some of the operational qualities are analyzed and knowledge about the different solution paths that have been already walked so far acquired.

As already mentioned, our approach uses empirical data to infer a relationship between input and output features by supervised learning. In this regard, this analysis also enlightens the selection among the consolidated LT determination approach that will later be used as reference labels for our input features. At the same time, this indirectly sets the *reference value* to be beaten by the solution proposed in the present work.

To finish this analysis, we gather them an make a summary about the options and the possible directions that exist towards a operational LT estimation.

2.2.1 A review of the reference state-of-the-art lactate threshold determination methods

Among the multiple lactate determination methods proposed so far, there are certain methods that can be considered consolidated due to the support they have both in theory and practice.

From a detailed analysis of their strength and weaknesses we will determine the value and its place in the *value map*. Later on, we will select the referent method for the empirical inference needed for the design of the *virtual sensor* among these methods.

- Maximal Lactate Steady State (MLSS): MLSS is defined as the highest blood lactate concentration (MLSSc) and work load (MLSSw) that can be maintained over time without a continuous blood lactate accumulation [14].

  From the physiological point of view, MLSS represents the maximum workload which the oxidative metabolism can sustain [20; 21; 22; 23]. In lactate steady state, the process of lactate appearance is balanced by the process of lactate disappearance, i.e. there is an equilibrium. The MLSS represents the maximum point in this equilibrium [21; 22; 24]. This concept is illustrated in Figure 2.6 where a typical MLSS estimation protocol is represented.

  To look for the potential benefits that MLSS can provide in the previously defined use modes (monitoring, training prescription and performance estimation), we now analyse the qualities of the MLSS. It has been demonstrated that the workload (speed, power...) at MLSS can be used to characterize information about athletes endurance capacity [14] and the accuracy of the original determination method illustrated in Figure 2.6 is high. Furthermore, it represents a useful quantitative measure of the exercise-related behaviour of the blood lactate concentration [25; 23].

  This means that MLSS can be used to monitor the evolution of this physiological phenomena in absolute terms and thus, it is more powerful than other monitoring types that only provide relative and general information about training induced adaptation.

  From the training prescription perspective, MLSS has shown to be beneficial [14]. This is so not only by the relative conclusions about performance changes that can be derived from its monitoring qualities, but because it is considered that MLSS can discriminate qualitatively between sustainable exercise intensities on which continuous work is limited by stored energy and exercise intensities that have to be terminated because of a disturbance of cellular homeostasis.

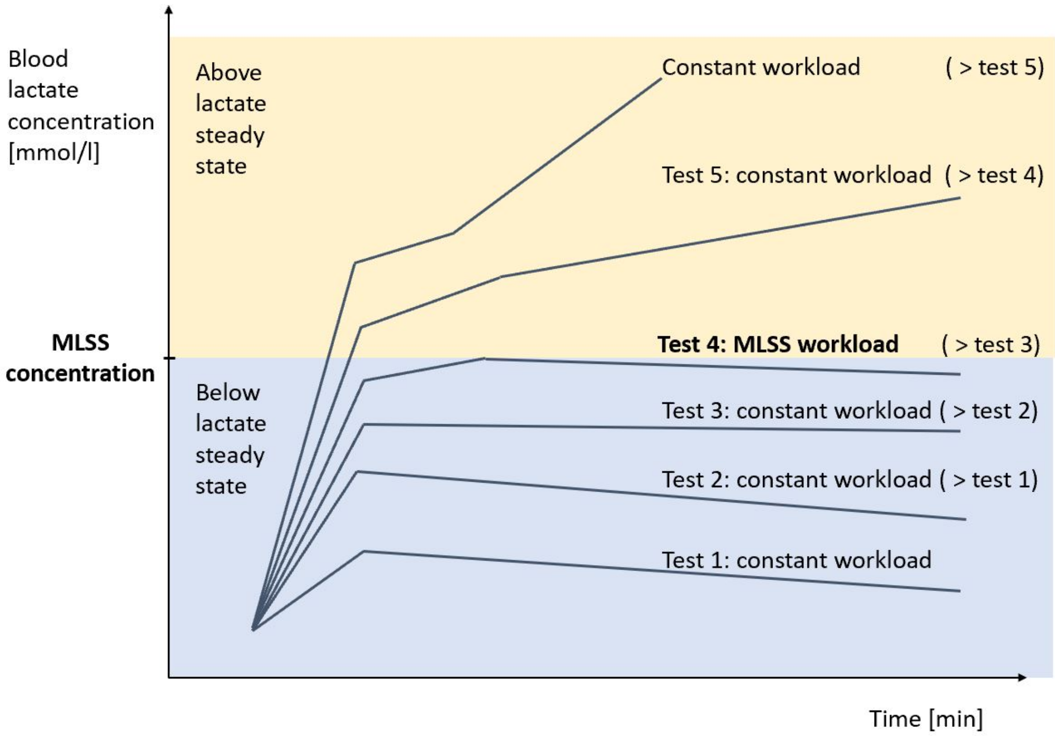


Figure 2.6: Representation of the Maximum Lactate Steady State testing and determination process

Additionally, a close relationship between endurance sport performance and the workload at MLSS has been reported, as the average velocity over a marathon is just below this workload. Thus, the prediction power of MLSS is above a purely relative evaluation provided by other methods and can serve as a rough estimation of athlete level [14; 23].

Actually, this method is highly beneficial in the three use modes, since it helps to monitor relevant physiological phenomena which at the same time is related to performance. Additionally, it may influence in the training prescription not only by indirect means of the monitored performance but also by determining training zones. Thus, the MLSS is considered the gold standard in terms of the information that characterizes about the LT concept [11].

However, the determination of MLSS is difficult, invasive, cumbersome and requires from 3 to 5 tests in a specialized centre to obtain an accurate result. More precisely, the testing protocol works as follows: a long duration test (usually up to 30 minutes) is performed at a fixed work load where blood lactate measurements are taken. These tests are repeated (with long proper rest between them and may be done in different days) until blood and lactate concentration increases continuously during the constant load. As represented in Figure 2.6, the MLSS is identified as the maximum sustainable blood lactate concentration and workloads [14; 23].

Analyzing the availability qualities of this method we observe that, access to specialized center, personnel and equipment is necessary. Moreover, sometimes several days are needed. Therefore, the availability of this method is low and limited to few people with high resources. Similarly, the usability of the method is very low, the athlete is dependant on multiple external tools and expert agents that understand the way of performing the tests and the testing process in unpleasant and unfriendly. Finally, it creates a big interference with training because several intense tests must be performed usually separated in different days followed. This creates both a high opportunity cost due to the repeated days required plus some residual fatigue that comes from it.

These results show that the cost of calculation the MLSS through the traditional method is high even for elite athletes, inconceivable for recreational runners.

Therefore, some authors propose to use a single test to indirectly determine the MLSS value [21; 24]. However, the former did not achieve a minimum necessary accuracy to stick as a referent. The latter and most important one is presented hereafter, showing strengths and weaknesses compared to the MLSS.

- Onset of Blood Lactate Accumulation (OBLA): It was observed that workload at MLSS elicits a blood lactate concentration average of 4.0 mmol/L [17; 14]. For that reason, this workload has long been estimated by the OBLA which is the work load corresponding to blood lactate levels of 4.0 mmol/L determined in an incremental test.

  However, as represented in Figure 2.7, the blood lactate concentration at MLSS has been reported to have great variability between athletes (from 2 to 8 mmol/L in capillary blood) and the conclusions obtained from this static point do not take into account individual characteristics [25]. In this regard, some researchers disagree with using the speed at OBLA as indirect marker of MLSS [14; 22], as multiple factors such as the aerobic training may affect in the lactate concentration that corresponds to the MLSS [26]. Ultimately, these facts suggest that the value of 4 mmolL-1 does not consider the inter-individual variability of the MLSS [27] and consequently it is considered that the speed at OBLA has too many limitations and does not characterize the aerobic capacity.

  This means that, as the phenomena characterized by OBLA targets less relevant information, the value for monitoring purposes is reduced in comparison with the MLSS. However, the workload at 4.0 mmol/L, is still a good indicator of the training adaptation produced in relative terms [21]. Actually, to its capacity to monitor the relative adaptations it has been long used for indirectly aiding training prescription.

  Finally, the speed at OBLA has been proposed as an effective variable to qualitatively determine the performance in several different sports such us track & field [15], swimming [28], soccer [29], hockey [30], cross-country skiing [31] or road cycling [32]. Nevertheless, and as already explained, it has significant limitations. Its main use is to estimate performance improvements and also has been suggested as a discriminator between well trained and elite athletes [33].

  Indeed, the value of this method in the three use modes is lower than the MLSS since it characterized only relative information.

  Similar to MLSS protocol, it is an invasive test that requires from blood lactate measurements and specialized equipment, but despite the MLSS, OBLA measuring protocol requires

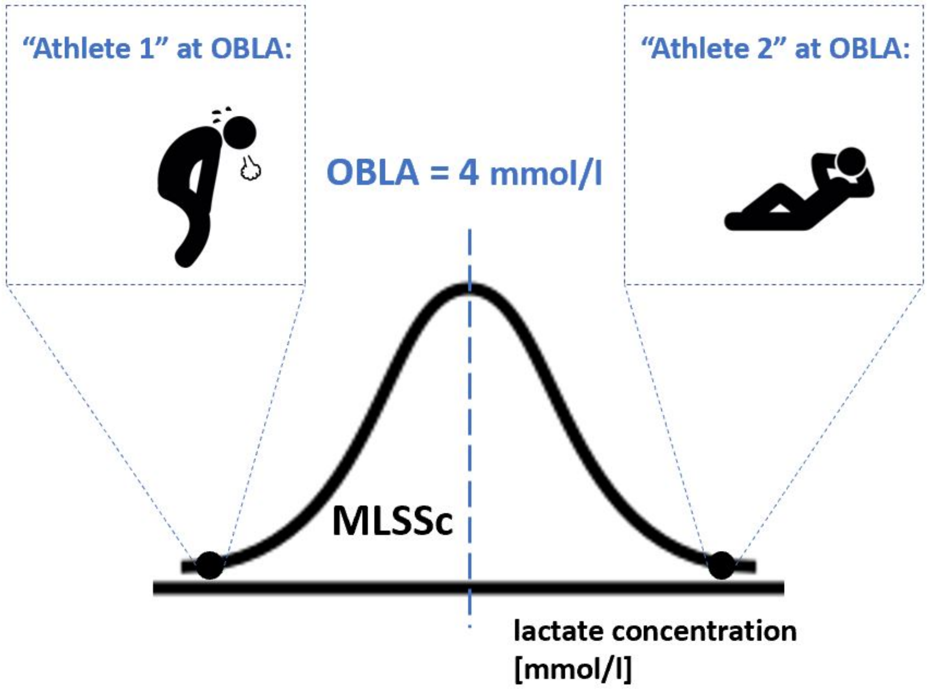


Figure 2.7: concentration at Maximum Lactate Steady State vs Onset of Blood Lactate Accumulation: inter-individual variability

only one test.

Thus, since the determination still needs from specialized equipment, it is only available to those with access to this resources. Similarly, the usability of the method is still low as the athlete is dependant on external tools and the test is still cumbersome and invasive which makes it unpleasant and unfriendly. However, in comparison with the MLSS, both the availability and usability are much higher. Finally, the interference that this method creates in the training is far-lower than the MLSS counterpart. This is mainly because a single test is needed to determinate it.

This results in that the operationality of the OBLA method is much higher in comparison with the MLSS. This made that a lot of elite coach and athletes have use it so far, but it is still costly for recreational runners.

- Individual lactate threshold: The Individual LT is defined as the maximum workload at which a sharp increase of the lactate occurs [34] calculated from the lactate curve obtained from a graded exercise (illustrated in Figure 2.8).

  The calculation of individual lactate threshold involves the measurement of blood lactate during an incremental step-wise exercise followed by a recovery. In this test, multiple lactate measurements are taken and use to form a convex curve from which the LT is determined. Multiple methods have been proposed to determine the LT from this curve such us the popular method that determines the LT as the first rise of blood lactate greater than 1 mmol [35]. Among these methods, maximum distance or Dmax method stands out [34]. In this method, the maximum distance between the lactate curve and the line between the first and the last lactate values is calculated and considered as the LT. This methodology is

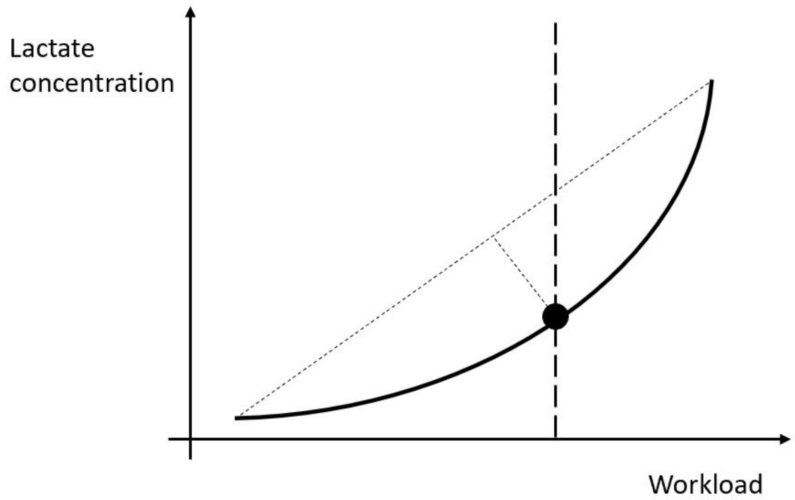


Figure 2.8: Individual lactate threshold calculated by Dmax method

fully automatizable and allows to calculate the LT in all the convex curves using a single experiment (something that is not possible with other methods) and was initially proposed by Cheng et. al. [36].

Some authors suggested that the individual lactate threshold workload calculated by Dmax method is closely related with the workload at MLSS [37]. However, the duration and size of the intensity increments have been found to influence the value of the lactate threshold [38], something that should be taken into account. In any case, this method is the most recommended methodology nowadays [39; 34].

Moreover, it has been observed that this method predicts the performance accurately, especially in recreational athletes [12; 40; 41; 39]. Thus the benefit that provides for this use mode are high. It can be concluded that, despite being lower than MLSS, the value of this method in the three use modes is high.

From the availability perspective, this method shows the limitations of the previous methods, in the sense that the test must be done in an specialized centre and needs from specialized equipment. The usability of the method is still low as the athlete is dependent on external tools and the test is still cumbersome and invasive which makes it unpleasant and unfriendly. Regarding the interference with training, it is much lower to the MLSS and similar to the OBLA. Therefore, the operationality of this method is much higher than the MLSS.

As we have seen, the consolidated methods have operational qualities that are not sufficient for most of the recreational runners. It is interesting to note that, all the consolidated approaches rely on invasive blood lactate measurements, and/or on specialized equipment and/or personnel to obtain it (Figure 2.9), meaning that there are accessible to few people.

Even that the process of taking blood samples has improved due to the simplification of the measurement devices, is still cumbersome and uncomfortable, which makes it inconvenient for its consistent use. Therefore, they do not fulfil the operational qualities necessary for our objective population, i.e. recreational runners, specially due to the poor availability and usability qualities they provide.

Figure 2.9: Equipment needed for the consolidated lactate threshold determination methods

So, the detailed analysis done in this section allows to place the consolidated LT methods in the *value map* according to their relative cost-benefit and consequently value characteristics. Figure 2.10 shows the landscape including these consolidated methods already placed.

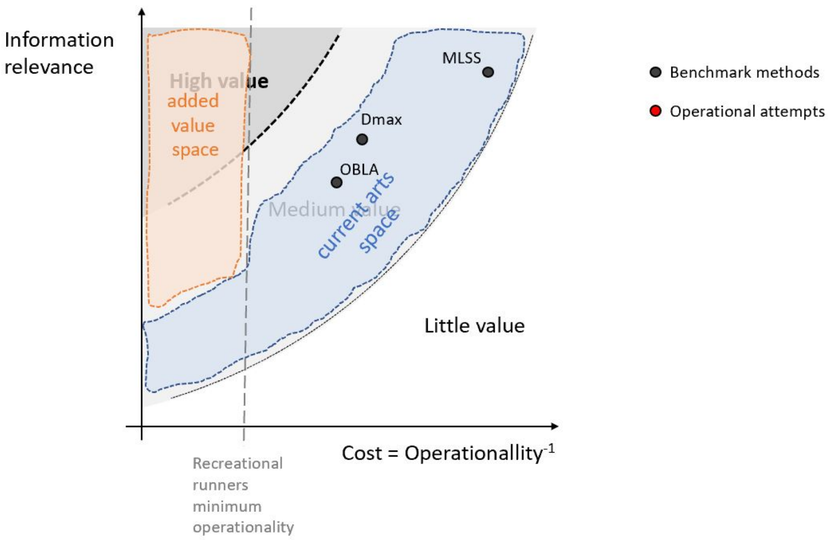


Figure 2.10: Value of consolidated state-of-the-art approaches

2.2.2 Attempts to improve the operationality of the consolidated lactate threshold determination methods

Given that the consolidated LT determination methods are not able to answer to the operational needs, multiple attempts have been made so far to overcome some or all of these limitations. As shown in Figure 2.11, very diverse approaches have been proposed in the literature.

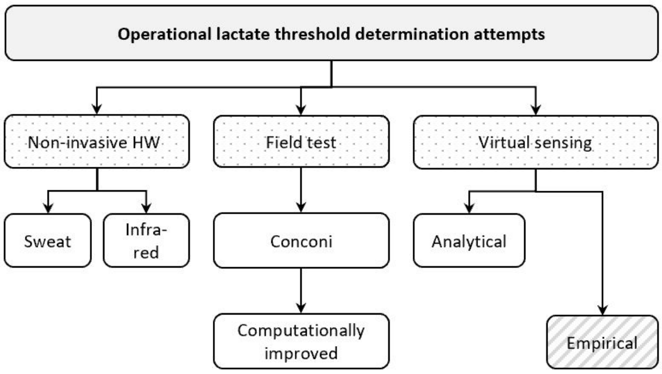


Figure 2.11: Classification of operational lactate threshold estimation determination attempts

Non-invasive hardware

Among these approaches, some focused on providing alternative non-invasive ways to determine LT by means of additional hardware (HW).

Historically, sweat lactate measurement has been seen as a potentially interesting manner to estimate the LT. It promised the opportunity of a non-invasive continuous monitoring with non extremely expensive devices [42] (Figure 2.12). However, there is a lot of controversy about the relation between the blood and sweat lactate levels. The review made by Derbyshire et. al. [43] pointed out that most of the studies showed no evidence of a direct relation between sweat and blood lactate levels. This means that, before even entering into a deeper analysis of the operational qualities of these methods, it is already discarded, since the value that this measurement could provide is low due to its incapability of getting relevant and accurate information about blood lactate.

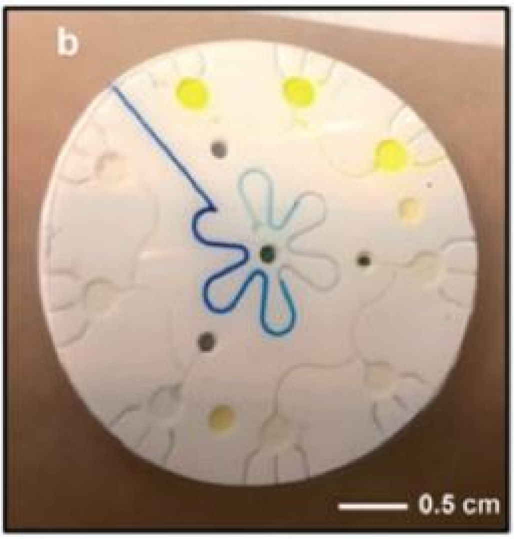


Figure 2.12: Sweat lactate meter

By Seshadri et. al.

Optical non-invasive methods to estimate the lactate levels by examining the connection between the physiological tremor occurring during muscle contracts and the lactate blood levels have also been proposed [44]. However, as represented in Figure 2.13 the solution requires from a very expensive HW, what puts it out of the scope of this work.

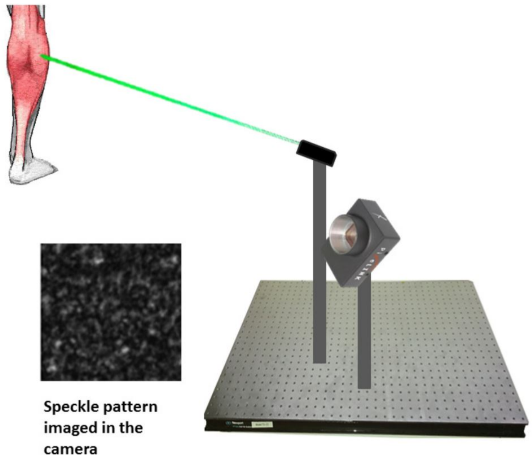


Figure 2.13: Photonic non-contact lactate measurement

By Abraham et. al.

Other approaches proposed non-invasive individual Dmax LT determination through muscle electrical impedance [45]. This LT estimation was addressed to professional rowers since, by eliminating the invasive nature of serial blood sampling, it would increase its usability to increase the number of LT assessment done in a fixed period of time. However, the accuracy of the bio-impedance measurements is a well known limitation [46]. In addition, caution is needed about its accuracy beyond the professional rowers population.

Among the non-invasive HW approaches, the most relevant one is probably a wearable stock presented in the work done by Borges et.al. [47]. This device is directed to runners and has already been validated against 7 male and 7 female athletes from recreational to highly trained levels.

Unlike the rest of the non-invasive HW approaches, which provide low quality information about the LT, this approach seems that may be in comparable terms with the Dmax individual lactate threshold. However, the validation sample is small (14 athletes) and thus the conclusions must be taken with a grain of salt.

From the operational perspective, they do not address the needs identified in the present work. It requires an additional equipment that need to be placed aligned with the thickest section of the gastrocnemius. In this regard, the need of and additional and expensive equipment goes against the availability that we seek in the present work. Moreover, the usability lowers due to the extra discomfort that this solution may entail, specially in sports such as triathlon were the transitions between disciplines are determinant and even creating a big interference. Therefore, despite it may have some interesting operational qualities such as continuous monitoring, this solution does not address the needs identified in section 2.1, probably because targeted at recreational runners with high economical resources.

In any case, all of the non-invasive HW based approaches, by definition, need additional equipment to estimate the LT, and in most cases this equipment is even more expensive than the portable lactate measurement devices used nowadays. This goes in opposite direction of the availability

Figure 2.14: Near-infrared lactate threshold detector

By Borges et. al.

qualities sought in the present work and therefore are not valid for the recreational runner population.

Field tests

Field tests are another extended approach to indirectly assess LT using means available to everyone. Among the field tests, probably the most known and used one is known as the Conconi Test [48]. This test is based on the relation between LT and the heart rate deflection point (HRDP). HRDP is a deflection from linearity of the heart rate (HR) with respect to the workload and it is related with the MLSS [11].

This test consists of running laps of 400 meters until exhaustion. After setting the initial speed and the running speed is increased slightly (0.5 km/h) every 200 meter while a heart rate monitor records the HR. As illustrated in Figure 2.15, the running speed - heart rate relationship was in part linear and in part concave. The point where both superimpose is defined as the HRDP.

This method only requires from a HR monitor and a stopwatch, a equipment already available to the majority of recreational runners and their coaches. Moreover, it can be done in any widely available athletic track. Despite a coach must manually measure the times, the availability of this method is high since the external resources are already widely available to recreational runners. Compared to the consolidated methods, the usability of this method is also much higher because of the non-invasive an simplicity of the test. In terms of interference, a separated test is still necessary which creates interference with training comparable to the consolidated methods.

However, one of the mayor drawbacks of this method is its accuracy. In this regard, several studies pointed out its deficiencies [46; 49] and evidenced that, in spite of being a commonly used method, the accuracy of this method is questionable. Factors such as lack of control of the

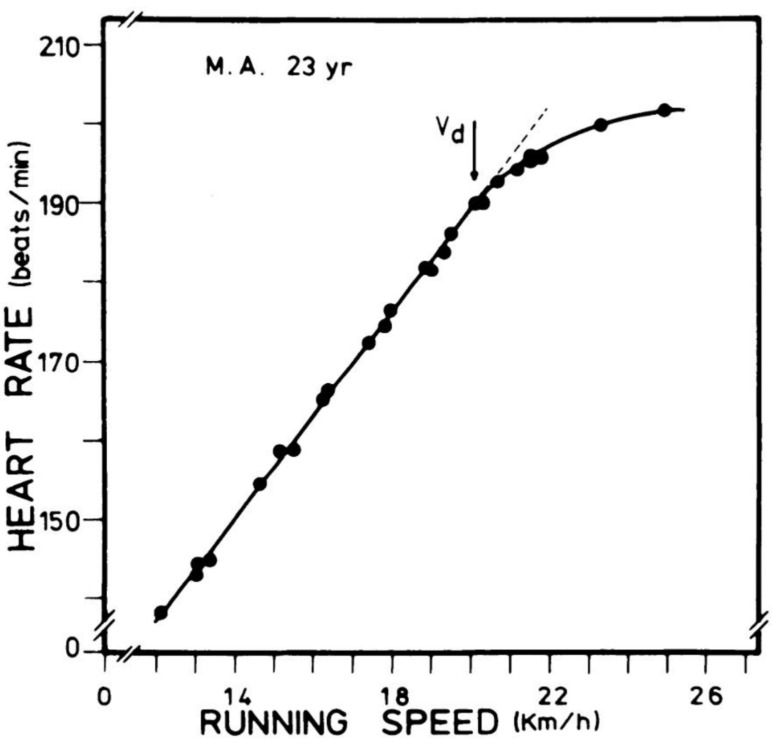


Figure 2.15: Indirect estimation of lactate threshold by "Conconi test"

By Conconi et. al.

conditions during and previous to the test and the impossibility of finding the HRDP in all the cases are probably among the reasons which make this method inaccurate.

Therefore, despite providing greater operational qualities compared to the consolidated methods, the value of a method with such a questionable accuracy are greatly reduced. In this regard, a computational solution has also been proposed to improve the accuracy of the Conconi test [50]. More precisely, as illustrated in Figure 2.16, they proposed to use two models (a linear and non-linear auto-regressive exogenous, ARX and NARX) in combination with fuzzy interpolation to model the heart rate (HR) dynamics and to better discern the HRDP.

However, although the work of Ringwood et.al. presented an interesting proof-of-concept, the obtained accuracy was low. Moreover, the population studied was very small (9 athletes) and thus, considering that the parameter identification was made from this sample, it is not reasonable to draw conclusions about its applicability, specially about the generalization capabilities of the model. More importantly, it does not compare the results to the ones that would have been obtained through the original Conconi test, which would have provided a much more meaningful conclusion about the improvement that this methodology provides against the original test.

Virtual sensors

As with the HRDP, it is known that the blood lactate concentration is related with multiple other features such as the HR at a given speed (or the speed at a given HR), the rate of reduction of the HR after an exercise or heart rate recovery (HRR), the rate of perceived exertion (RPE) as

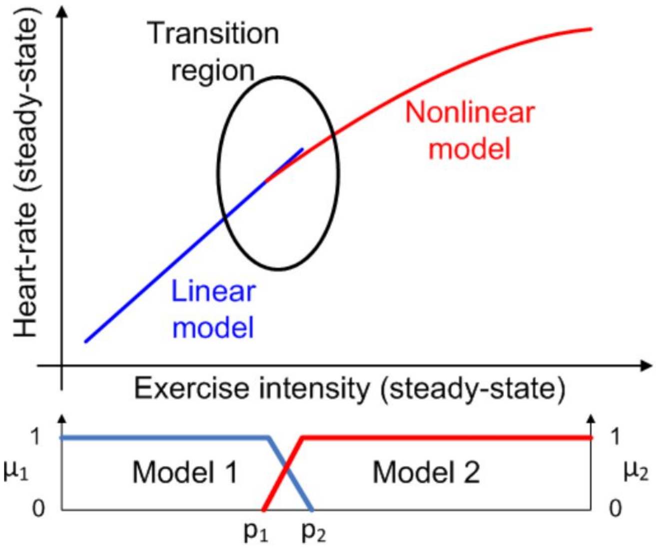


Figure 2.16: Improved Conconi test by using machine learning techniques

By Ringwood et. al.

Abreviations: p1, p2, fuzzy interpolation optimal values

described by Borg [51], gender, age, diet or athlete level [52]. Therefore, and similarly to what we hypothesize in this thesis, the indirect measurement of the LT by means of other accessible and easily measurable features has been proposed as an alternative, i.e. using a virtual sensing approach.

From the operational perspective, given that the *virtual sensors* are based in easily available input features (see Figure 1.7), the creation of a model that can estimate LT can provide multiple desired operational qualities.

In this approach, all the three availability, usability and interference would be subjected to the work and resources needed to obtain the input features. Therefore, *virtual sensing* has intrinsically embedded the potential to provide a high quality operational solution.

One of the approaches possible for creating a *virtual sensor* is creating an analytical model from indirect features related with lactate. This is precisely the approach followed in the work of Proshin et. al. [53]. They proposed a mathematical model of human lactate metabolism gathering several physiological models and merging them into a single one by extending a previously created cardiac system model and including it in a system of equations that describe the dynamics of lactate metabolism processes in the organism (Figure 2.17).

However, the purpose of the work presented by Proshin et. al. was to create a model that would be parametrized to be used in individual athletes. This means that multiple measurements including hemoglobin, blood pressure and saturation measurements, blood lactate measurements and a respiratory metabolism analysis etc were needed to fit the parameters. Thus, the cost of parametrizing this individual model is already huge, leaving it out of the scope of this thesis.

What the work of Proshin et.al. clearly shows is that creating an analytical model that could explain the complexity of the lactate metabolism, not only that of individual athletes, but for the

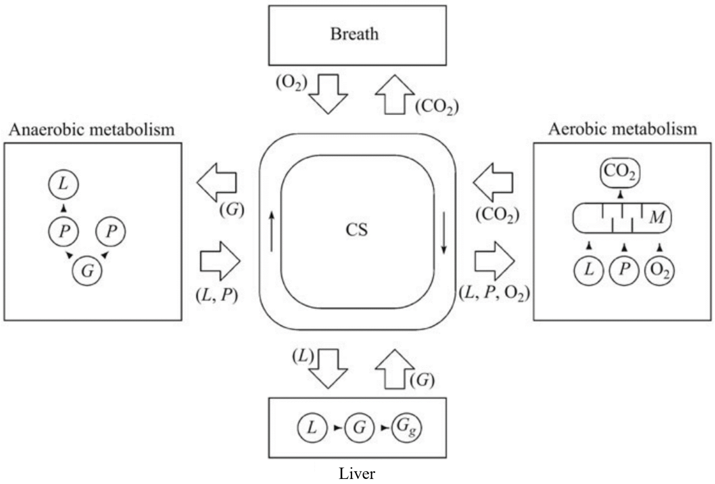


Figure 2.17: Analytical modeling of lactate metabolism

Modified from Proshin et. al.

Abbreviations: G, glucose; Gg, Glycogen; P, pyruvate; L, lactate; M, mitochondrion; O2, oxygen; CO2, carbon dioxide; CS, cardiovascular system;

entire recreational runner populations, is far from being feasible. It is even more difficult to do it without additional measurements.

As already mentioned in Chapter 1 in complex problems such this one, empirical virtual sensing approaches are, a priory, a more appropriate way to go. In this regard, Machine Learning (ML) techniques such as Artificial Neural Networks (ANN) are widely used to create models of complex non-linear dynamic problems. These computing systems that are inspired by biological neural networks, helped by their complex network style interconnected architecture shown in Figure 2.18, are able to model complex relationships between inputs and outputs.

Actually, ANNs have also been scarcely used to model lactate production in athletes. In this kind of empirical modelling approaches of complex phenomena, the main inherent difficulty is to obtain a generalizable solution. In other words, a solution that is accurate both in the sample from which the function is inferred and in unseen data.

Erdogan et al. [54] proposed a model based on a multi-layer perceptron (MLP) to estimate the HR at onset of blood lactate accumulation (OBLA) point, with its strengths and limitations. This means that, the maximum possible value of the model of the work by Erdogan et. al., by definition is set by OBLA's relevance which, as we have already seen, it is limited compared to other methods such as the Dmax individual LT.

From the generalization point of view, the studied population was coming from an homogeneous football player sample and, as the authors themselves acknowledged, more training and testing cases from heterogeneous groups are needed in the future for better generalization.

A more recent paper [55] also proposed a machine learning model to estimate LT. Thirty-one healthy male and female participants made a cycle ergo-meter tests to gather data and create the model. A limitation of the work by Huang et. al. is that the population from which the participants

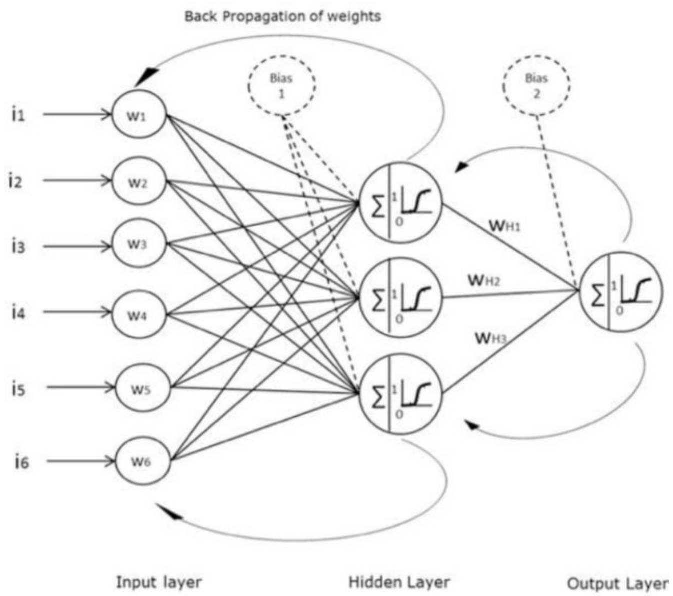


Figure 2.18: Artificial neural network architecture

By Liana et. al.

Abbreviations: I, input; w, weight

were gathered is not clear. From the generalization perspective, the data-set is considered insufficient to achieve a generalizable accuracy, because it has a reduced sample size. Furthermore, it comes from a different population to the recreational runners.

More importantly, from the operational perspective, multiple non-invasive but still costly cardio-respiratory and anthropometric factors were used for the model. So, as already mentioned, the availability and usability of this model becomes dependent on the cost of obtaining these measures, high in this case.

Therefore, the validity of both models presented above [54; 55] is limited by their methodology which, due to the homogeneity of the population in the former and the inadequate sample in the latter, are prone to create over-fitted and thus non-generalizable models. Moreover, neither of the models have the desired operational qualities.

From the analysis of the operational attempts done we identify and represent in the *value map* the most relevant approaches among the already explored paths (Figure 2.19).

This map, apart from gathering the value of the most relevant methods, also highlights the importance qualities identified in section 2.1 to determine what value is in the context of our problem. Actually, non of the approaches provide a valuable solution to the needs of the recreational runners to aid their training decision making (Figure 2.19).

This analysis further strengthens that *virtual sensing* has the potential to give an answer to the still unsolved operational LT estimation problem. More precisely, after discarding the analytical path, supervised learning based *virtual LT sensor* arises as the way to go, as it has been already hypothesized in the present work.

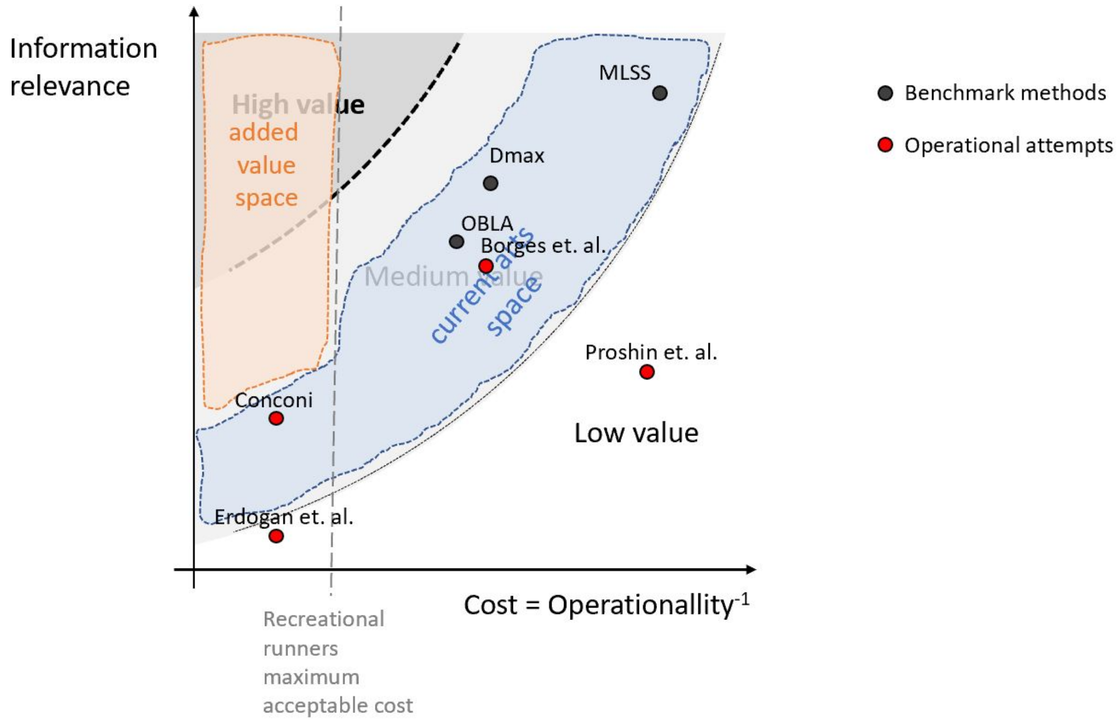


Figure 2.19: Value map with operational state-of-the-art lactate threshold determination methods

2.3 State of the art conclusions

This chapter served to make an analysis of the most valuable LT determination methods available. The multiple efforts that have been made to determine the LT and obtain a more operational LT determination method confirms that there is a huge interest in this matter.

To be able to determine the value of different methodologies, we first established the qualities used to determine the value in the context of training decision making of recreational runners. Additionally, we created a *value map* that served not only to organize all the important LT determination methods in the same place, but also to, in future steps, be able to place our solution and compare with the rest of the proposed approaches. This framework arises as a tool that allows to map the state-of-the-art of different proposals according to the value that they provide for the context of application.

Using these criteria, we made a deep analysis of first, the consolidated LT determination methods available nowadays and second, the attempts that have been made to improve the operationality of these approaches.

None of the proposed alternatives are able to solve the operational problem of the current LT determination methods. However, supervised learning based *virtual LT sensor* has shown to have potential to answer this problem. This is precisely the first hypothesis of this thesis.

Moreover, to create a supervised learning based *virtual LT sensor*, a reference LT method must be selected for labelling purposes. In this chapter we also set the criteria for the reference selection and the Dmax individual lactate threshold aroused as the chosen.

The criterion used to select the reference method starts from the obvious need of characterizing highly relevant information for the three use modes. This criterion already discards the OBLA.

The second criterion, relates with establishing a reasonable sample size, including economical and time constraints to perform the experiments. The more observations the more robust the solution. This criterion discards the in comparison much more expensive MLSS method since 3 to 5 tests are needed for a single observation, something that would reduce from 3 to 5 the amount of observations that we would be able to collect.

The third and last criterion relates with the once and again stated need of addressing the demand coming from practice. Nowadays, the Dmax individual LT method is the most recommended method [15] specially for recreational runners. Moreover, it straight addresses a real demand of the community of recreational runners and coaches that use this LT method

Therefore, the Dmax individual lactate threshold is the reference selected in the present work and is one of the major conclusions of this chapter. Consequently, as illustrated in Figure 2.20, by the selection of the reference method we set the maximum relevance and value that our solution will have and we indirectly define the problem space in which our solution must fall.

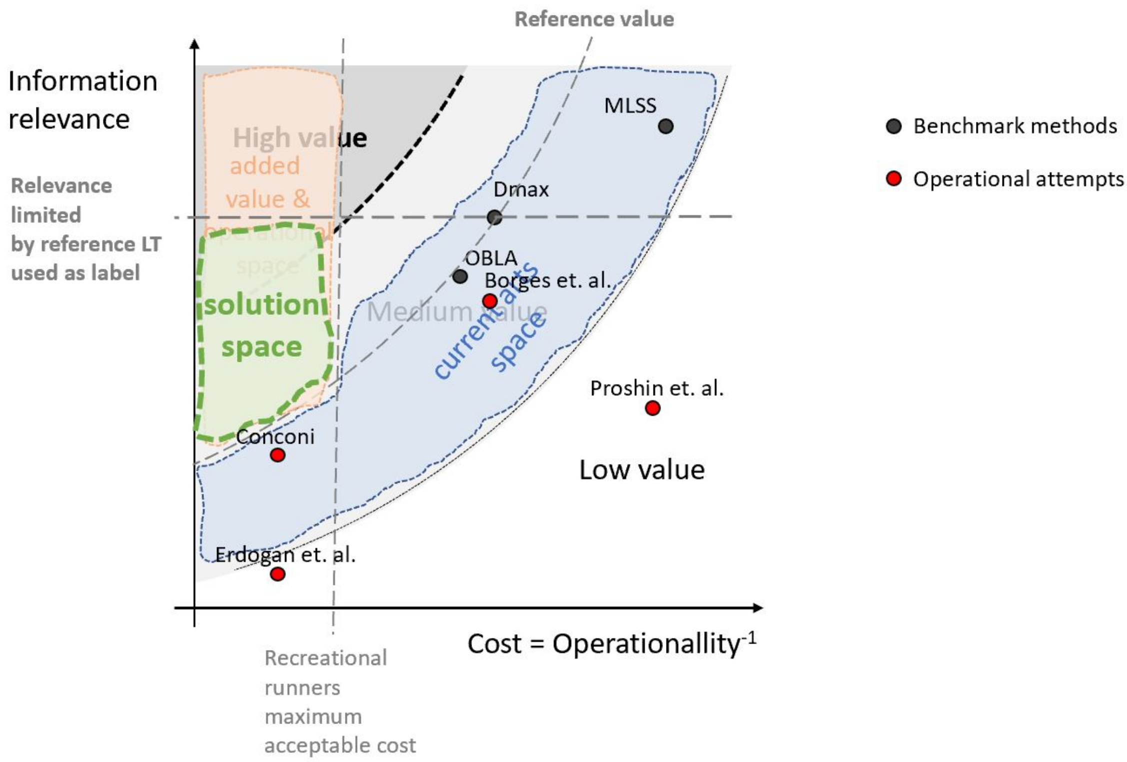


Figure 2.20: Value map with the solution space

Chapter 3

Strategy and methodology to design a data based virtual lactate threshold sensor

Strategy without tactics is the slowest route to victory. Tactics without strategy is the noise before defeat - Sun Tzu

Figure 3.1: Visualizing the route to the top

This chapter defines the methodology that is followed in the design of the *virtual LT sensor* according to the needs identified in the previous chapter.

A proper methodology entails creating a robust and systematic way to face and create solutions for certain types of problems. Thus, the main purpose of this chapter is to create a framework that, by taking into account the characteristics of *the operational LT problem*, provides a robust and systematic way to design a *virtual LT sensor*.

To do so, we first make a high level analysis to detect the main characteristics and inherent difficulties of *the operational LT problem* and provide precise strategical means to deal with them. More precisely, we define the meta-process and the performance perspectives that are going to be used in the present work.

With the aforementioned strategical principles in mind, and combining with the identification of the inherent difficulties that creating a supervised learning based *virtual LT sensor* has, we create a framework that will afterward guide the design phase.

As illustrated in Figure 3.2, the framework grows from two strategical principles for solving complex ML problems: (1) Using an iterative approach for boundary discovery and (2) setting a *satisficing* accuracy to minimize the problem complexity (see section 3.1). Then, as the *virtual LT sensor* is a data based ML model, the experimental methodology defines how the experiments are to be performed and validated for their use in the *virtual LT sensor* design. Section 3.2 goes into detail of the preparations, requisites for realization and the definition of the validity of an experiment. Finally, with the principles and the experimental methodology in mind, the design of the *virtual LT sensor* is divided into, *context characterization*, *content representation* and *deciding next step*, three steps that are common in supervised learning approaches.

3.1 Strategy for the design and development of a virtual lactate threshold sensor

In Chapter 2, we concluded that a supervised learning based virtual sensor has potential to be an operational LT estimator for recreational runners and that the use of the Dmax individual LT as reference for labelling the outputs is appropriate for the *virtual LT sensor*.

Supervised learning can be described as the ML task of learning a function that maps an input to an output based on sample input-output pairs. ML approaches enable to model more complex systems in comparison to analytical approaches (i.e. approaches that try to find a closed form solution expressed as a mathematical analytic function) that tend to fail when the explicit input-output feature relations are unknown. For instance, ML techniques have been used for estimating features such as grinding energy [56], which would be much more difficult or even impossible to do analytically. In our case, as already stated, the output of interest, i.e. the output that the virtual sensor must estimate from a set of easily measurable input features, is the Dmax individual LT.

This kind of learning that infers a general function from specific observations, as represented in Figure 3.3, follows an inductive reasoning.

The main characteristic of this kind of reasoning is that, unlike deductive reasoning where the

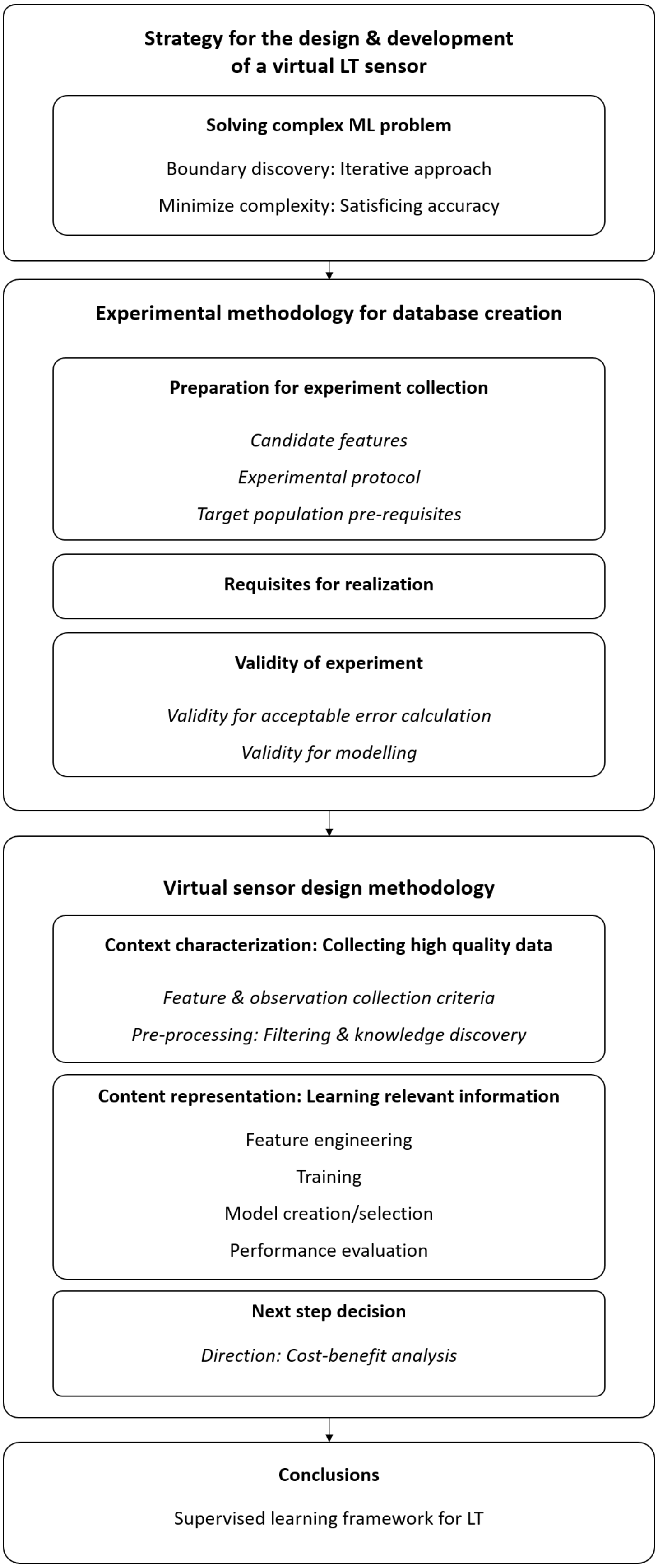


Figure 3.2: Overview of Chapter 3: strategy, experimental methodology & design methodology

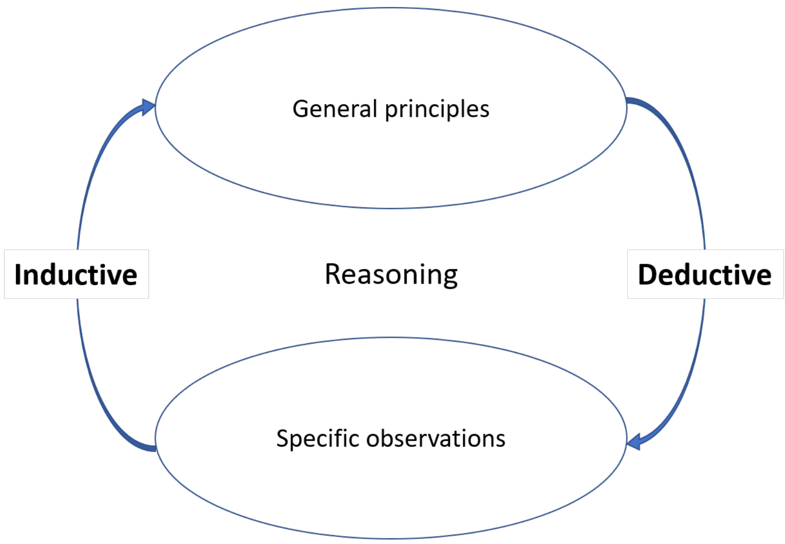


Figure 3.3: Inductive vs deductive reasoning

conclusion of an argument is certain, the truth of the conclusion of an inductive argument may be probable (i.e. weak or strong), based upon certain influencing aspects. Those aspects are ingrained into the next question: which input features and how many observations sampled from the target population do we need to make a strong (hereafter robust) learning about the underlying general function?

Given that the observations are collected from the right population (i.e. recreational runner population), the selection of the input features and the number of observations needed for a robust estimation is directly related to the complexity of the problem to be solved. In other words, given that the features are relevant, the more complex the problem the richer the set of features and observations needed. In the present work, the term complexity is understood as the inherent characteristic of a system or problem due to the number of relevant features, and the inter-dependencies between these features and their context. The higher these numbers the more complex. More precisely, the complexity relates to characteristics such as: the system being bounded in the feature space (i.e. when all or most of the relevant features can be clearly identified and measured), the knowledge about the underlying rules of the system, and if these rules change over time. So, not only that the higher the complexity the higher the features and observation needed, but it also must be considered that, for certain systems with fast changing rules, the needed resources may change over time.

Regarding the system targeted in this work (i.e. the recreational runners LT derived from their lactate metabolism), as stated in Chapter 2, only fuzzy rules are known about the lactate production and there is a fair amount of uncertainty due to the multiple features related to LT in an inter-dependent way. However, the rules are known to be stable in a healthy person. Therefore, we are targeting a moderately complex phenomena which, despite having unknown boundaries, the uncertainty related to it is finite and stable.

Being the *targeted system's complexity* stable, the overall complexity of the problem (the design of an operational *virtual LT sensor* in our case), can be formulated as being proportional to the combination between the desired performance (how the *virtual LT sensor* should behave) and the *targeted system's complexity* (1). This can be illustrated as:

Problem complexity $\propto$ Desired performance of the *virtual LT sensor* x Targeted system's complexity (1)

Hence, the higher the *problem complexity*, the more resources are needed to appropriately characterize it. This means that more relevant features and their corresponding observations are needed to allow a robust inference about the input-output relationship [57].

However, due to the already mentioned uncertain boundaries, the exact number of appropriate features and observations that are necessary for a robust inference is unknown ex-ante. Usually, expert knowledge about the phenomena being characterized (LT in this case) can give some hint about the most important features. Then, the number of observations required can be estimated accordingly. However, due to the inherent characteristic of this kind of problems, the boundaries may only be discovered ex post, once a model is created. Therefore, it is fundamental that the strategy and methodology used embraces the discovery of the problem boundaries as part of the problem solving process [58].

Approaches inspired by the conceptual framework of evolution are well suited in this task [58]. Iterative strategies are among their more simple types and fit very well to this problem, since it allows to learn from the experience of previous iterations. This allows to adjust the features, observations and even the learning approach used to make the most from the available data.

In our particular case, as formulated in formula (2), the goal of the iterative strategy is to discover a successful solution by adjusting our steps towards matching the available resources (features, observations, computational power...) with the *problem complexity*.

Available resources $\equiv$ Problem complexity $\rightarrow$ successful solution (2)

As shown in Figure 3.4, this adjustment can come from any of the following two means: a) from increasing the resources used and/or b) from reducing the complexity of the problem (while its still useful for application purposes).

What is known as *brute forcing* is the most extreme case of solving a problem making use of resources. This approach may only serve when, apart from targeting to a system with static uncertainty, high amounts of relevant data (features and observations), powerful algorithms and HW that minimizes the computational cost are easily available. Problems such as games [59] and image processing [60] tend to fall into this category. One of the most recognized materialization of the *brute forcing* concept is the well-known *deep learning* [61].

Regarding the applicability of *brute forcing* to design the *virtual LT sensor*, the cost of determination of the Dmax LT is around 100 euros per experiment if done in a specialized center, which impossibilities the chance of collecting high amounts of data. This fact makes the *brute forcing* path not viable for us. In any case, it highlights the importance of maximizing the proper use of the available resources (good and enough amount of data in our case) and further strengthens our selection of the Dmax individual LT as reference instead of other more costly alternatives such as MLSS.

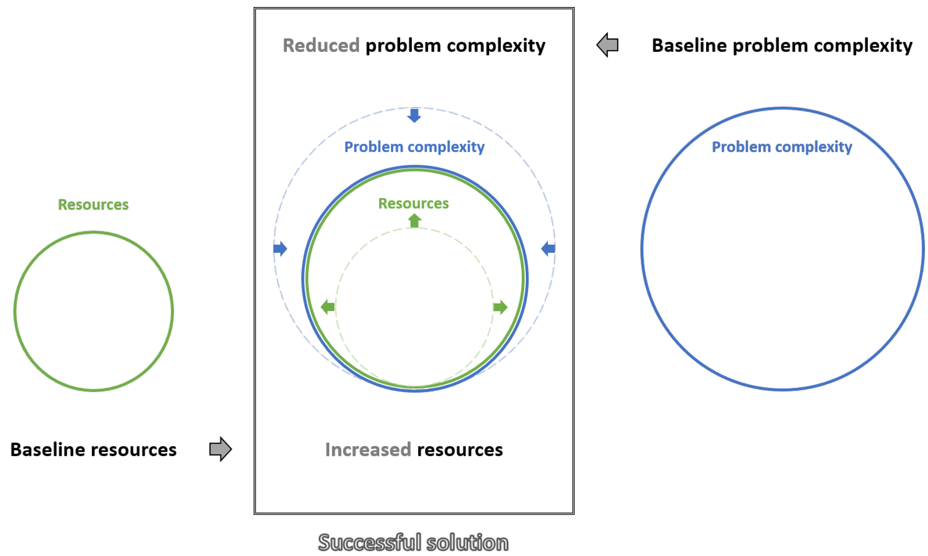


Figure 3.4: Successful solution: Trade-off between resources and problem complexity

Apart from using resources, as represented in Figure 3.5, another strategy for solving the LT kind of problem is by reducing the *problem complexity*. As illustrated in equation (1), this can be made from two means. Minimizing the *targeted system's complexity* (not viable in our case as already mentioned) or minimizing the *desired performance*, i.e. reducing the accuracy or performance objective of the virtual sensor.

Therefore, this is a fundamental principle that will be used in this work by aiming first for a minimum necessary *satisficing* (i.e. the combination of satisfy and suffice [62]) performance and grow to more ambitious objectives afterwards. Moreover, this approach suits very well with working iteratively.

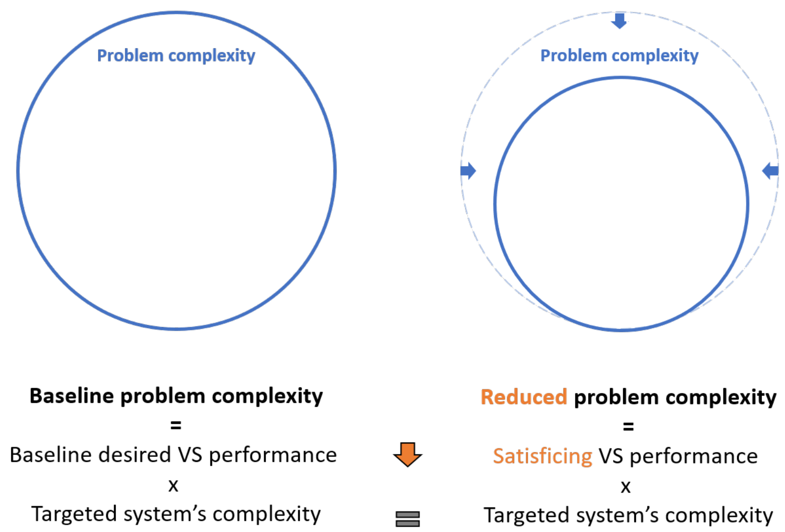


Figure 3.5: Reduce complexity: Aim for *satisficing* performance

Abbreviations: VS: virtual sensor

In the following sections, we will delve deeper into these strategies: the development of an iterative strategy and the general considerations for performance evaluation of the ML system towards achieving a *satisficing* solution.

3.1.1 Meta-process and process for a machine learning based virtual sensor design: Combination of iterative and traditional approaches

As already mentioned (see formula 2 in section 3.1), a successful resolution of complex problems comes from having enough and appropriate resources to characterize the complexity of the system for the desired performance. However, apart from having tools (ML techniques in this case) that enable to create models that represent the complexity of the problem in hand, the main practical difficulty that this kind of problem poses is that, which and how many resources (features and observations) are needed is a priory unknown.

First, as previously mentioned, a meta-process that takes into account and provides ways to manage the uncertain boundaries of the problem is necessary. In this regard, traditional engineering approaches, i.e. those that make several planning estimation for different project phases (functional specification, design, validation, manufacturing and implementation), tend to fail under problems with uncertain boundaries as the design of the *virtual LT sensor*. For this purpose, iterative approaches tend to work better since they enable to explore the problem in small batches and to continually adjust the direction according to what is learnt from the previous iteration. In the case of designing a *virtual LT sensor*, the aforementioned agility to adjust direction facilitates to work with incremental objectives towards matching the available resources (features-observations pairs) to the problem complexity (see formula (2) in section 3.1).

The application of this kind of iterative design is not uncommon in ML [57]. However, instead of being a conscious methodological decision, its use tends to naturally arise from the ease of applying it when practically unlimited and cheap resources are available. In other words, when it is a path of low resistance. Image processing problems are an example of this [60]. Therefore, and to the best of our knowledge, iterative strategies have not been explicitly formalized for ML application. This formalization is an important contribution of the present work.

Second, at process or iteration level, a traditional problem-solving perspective is used. By applying reductionism, each iteration is further divided into sub-parts and planned according to estimations of needed resources. More precisely, each iteration follows a methodology that is separated into context characterization (gathering and appropriately preparing data), content representation (making the best learning from it), and deciding next steps (adjusting the knowledge of the problem boundaries and the future direction).

Finally, the iterative meta-process and the traditional process are combined to create a framework to design the *virtual LT sensor*. As represented in Figure 3.6, the framework starts with a description of the experimental methodology that is to be used in each design iteration. Then, the design starts according to the described meta-process and traditional process perspectives.

Following this framework, the objectives will be progressing from less to more ambitious goals through the iterations using the experience of the previous ones to adjust the direction. This strategy permits to minimize the chances of making big backward steps that too ambitious

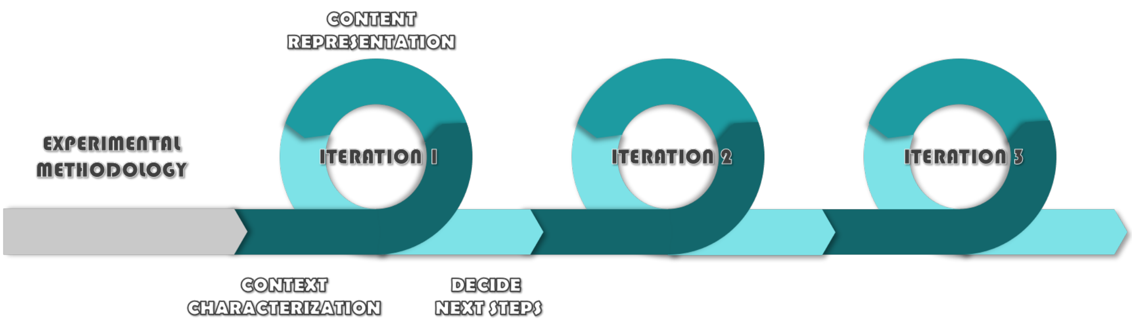


Figure 3.6: An iterative strategy and classical methodology for an operational lactate threshold estimation

objectives use to lead to.

3.1.2 Desired performance for the virtual lactate threshold sensor: *satisficing* accuracy

Based on what has been previously stated, the performance objectives selected for the ML based *virtual LT sensor* (desired performance of the *virtual LT sensor* in formula (1) section 3.1) indirectly defines the *problem complexity* and consequently its viability and approach with respect to the available resources for a successful solution (formula (2) section 3.1). Moreover, we mentioned that setting a *satisficing* performance reduces the problem complexity and consequently increases the chances of solving the problem with the available resources.

In this regard, setting proper performance metrics is key and usually one of the most determinant steps for a successful application of a solution in the real world. But prior to setting any metric, the following question must be answered: "what makes our system good?". It seems obvious to say that, what we are looking for is an accurate estimation of LT of recreational runners, so we can reformulate the question to: "what makes our system accurate?" According to the *satisficing* principle, our system is accurate if it achieves a *satisficing error*. In other words, if the ML system error (hereafter *system's error*) is below a *satisficing error*.

$$\textit{system's error} \leq \textit{satisficing error} \quad (3)$$

The *system's error* refers to the error that our ML system has with respect to the underlying joint probability distribution (i.e. the real probability distribution of the recreational runner population). However, as already mentioned, under induction the determination of the *system's error* is not straightforward. Under inductive reasoning, the conclusions, instead of right or wrong, are more or less robust. In other words, the ML system infers a function that has weaker or stronger generalization capabilities to unseen athletes. This means that the *observed error* may differ from the *system's error*.

In this regard, the ML *system's error* can be broken down into a combination of *observed error* (i.e. the error made in the available data) and *generalization error* which is related to the robustness of the ML under unseen data [63] (see Figure 3.7 question 1a):

system's error = *observed error* + *generalization error* (4)

This means that, for a solution valid for the usual recreational runner, the combination of *observed error* and *generalization error* must be below a *satisficing error* (5).

observed error + *generalization error* ≤ *satisficing error* (5)

On the one hand, the *observed error* is quantified with the available data. On the other hand, the *generalization error* is not strictly quantifiable nor fully observable, as there is no way to test it in the entire underlying probability distribution (which would mean that we are able to gather all the data). Traditionally different ways of estimating the *generalization error* have been used (e.g. comparing the difference between the error in the sampled data and an out-of-sample data). However, the correctness of this estimation is subjected to the robustness of the learning and evaluation methodologies (see Figure 3.7 question 1b).

Therefore, it is fundamental to minimize the *generalization error* to be able to confidently conclude that the calculated *observed error* can be considered the *system's error*, the cornerstone of ML [64] (see Figure 3.7 question 1c). To solve this, apart from raw estimated performance meta-attributes as the robustness of the methodology and the final model must also be considered. So that, beyond achieving a low observed error, we maximize the chances of the inference to be generalizable to the remaining unseen recreational runner population.

Wrapping up, as represented in Figure 3.7, the desired performance of our system is to be achieved by following two approaches: first, from the ML methodology perspective, the minimization of the *generalization error* is sought so that the observed error approximates as much as possible to the *system's error*. As represented in Figure 3.7, the maximization of the methodology's robustness is a necessary condition for proper evaluation of the created *virtual LT sensor system's error*. Second, the evaluation of this performance with a *satisficing* perspective according to certain thresholds is sought in order to make our system applicable in the real world.

Then, using these evaluation principles, we can confidently assess where in the trade-off between *problem complexity* and *available resources* are we (see formula (2) in section 3.1), so that we can make next step decisions towards a successful solution if necessary.

Machine learning methodology performance: Robust learning and evaluation

The main purpose of the ML methodology is to both accurately capture the regularities in our database and generalize well to unseen data. In other words, there is a trade-off to properly match the degrees of freedom of the learning approach with the degrees of freedom of the relevant information collected in our database. This trade-off is known as the bias-variance dilemma [57] and is directly related to the *observed error* of our system. An incorrect trade-off may lead to a model that either under-fits or over-fits the data. In this regard, Figure 3.8 represents three different functions that could be fitted to the same set of data points, with completely different consequences in terms of the bias and variance of the system. There, it is shown that the error is high both if the learning approach has less degrees of freedom compared to the phenomena being targeted (Figure

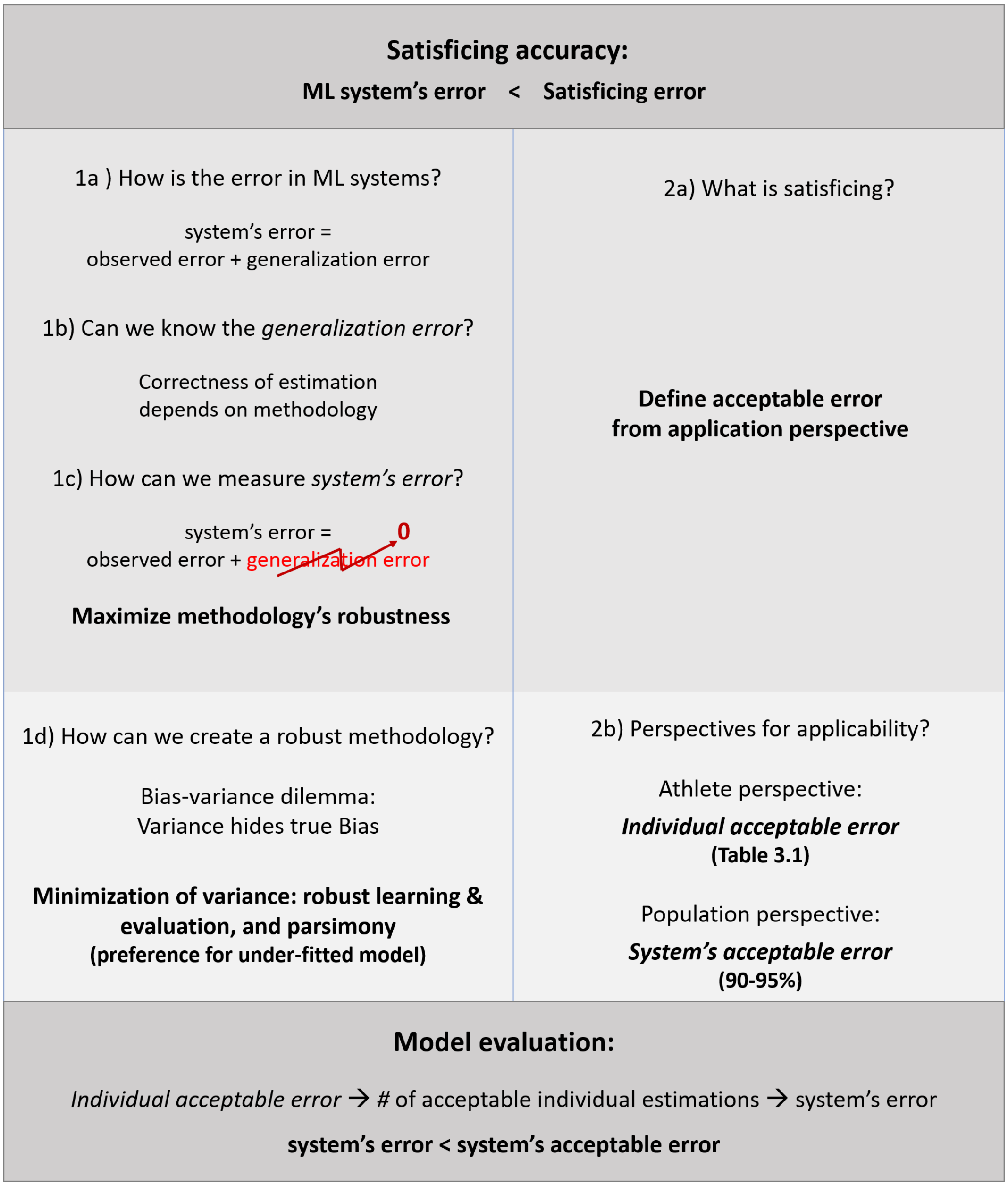


Figure 3.7: Achieving satisficing accuracy: Proper Machine Learning system's error calculation and satisficing error definition

3.8 image 1) and if the degrees of freedom of the learning approach are higher than the phenomena being targeted (Figure 3.8 image 3).

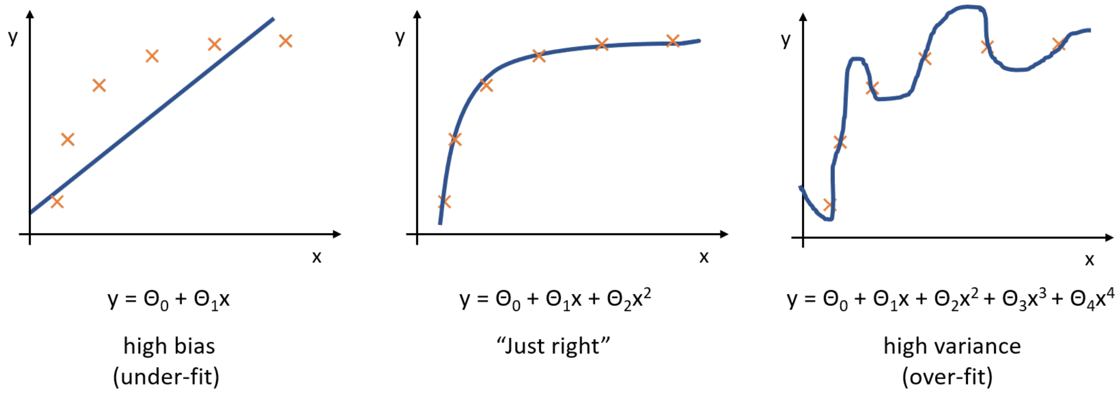


Figure 3.8: Representation of bias-Variance trade-off

Hence, with respect to the purpose of maximizing learning the relevant information, neither the under-fitted nor the over-fitted models are optimal. However, regarding the performance observability attributes, while an under-fitted model behaves equally biased under seen or unseen data, an over-fitted model hides its error in the variance term (see Figure 3.8 question 1d). In terms of the ML *system's error*, the over-fitted model has a high *generalization error*, which goes against our aim of creating a robust system with observable error (see Figure 3.7).

Therefore, in order to properly observe the *system's error*, the aim is to minimize the chances of creating an over-fitted model by maximizing the robustness of the methodology. This implies that, in the absence of a perfect model, an under-fitted model is prioritized over creating and over-fitted one. In this work, this is what we define as creating a robust ML system. More precisely, robust learning and performance evaluation techniques are to be used to minimize the degrees of freedom of the learning approach. To do so, a combination of techniques and principles such as re-sampling, regularization, parsimony and ensembling are to be used so that, applied to multiple layers, help create a robust ML system. This approach is described in detail in section 3.3.

This way, the *system's error* can be evaluated against the *satisficing error* and decide what to do next (section 3.3.3).

Application level performance: *satisficing* accuracy at individual and system levels

The previously mentioned *satisficing error* that is to be achieved comes from two considerations, the error that is acceptable at individual level and the error that is acceptable at population level (see Figure 3.7). In this regard, the *system's acceptable error* serves as a *satisficing error* to be achieved at system level. At the same time, the *system's error* (see Figure 3.7) is calculated from the computation of every individual estimation and comparing it to a *individual acceptable error*. In both cases, expert knowledge serves as the criterion to set robust *acceptable errors*.

From the individual perspective, a valid solution for training decision-making resides in providing estimations for individual athletes under a maximum acceptable error that ensures that the estimation is useful. As represented in Figure 3.9, there is an error in the LT estimation above

Table 3.1: Lactate threshold *individual acceptable error*

Pace at the LT (min/km)	Maximum error in the LT ±(s/km)	±(%)
≤3*	3	1.7
[3, 3.5)	5	2.4
[3.5, 4)	10	4.2
[4, 4.5)	15	5.5
[4.5, 5)	20	6.6

*Out of scope: Fitness level above target population

Abbreviations: LT, lactate threshold

which the information collected is not useful and is considered noise. In this regard, we define the *individual acceptable error* as the *satisficing* threshold that sets the maximum error in the LT estimation for a particular athlete.

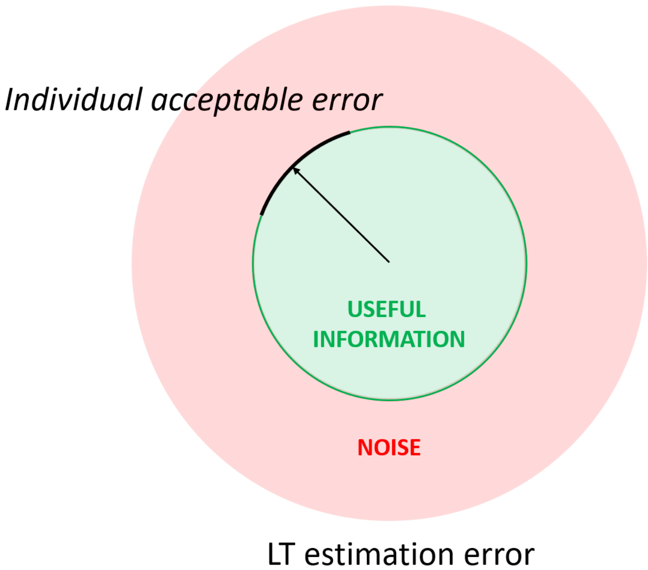


Figure 3.9: *Individual acceptable error: the threshold between useful information and noise*

Abbreviations: LT, lactate threshold

To calculate it, a maximum acceptable error for individual athletes is defined according to the experience of experts in physiology and exercise performance. In particular, we propose an *individual acceptable error* for recreational athletes (see table 3.1). The physiology perspective is applied in the following way: since higher level athletes require higher individualization in their daily training, a higher accuracy than the baseline is deemed necessary. On the other hand, this is the opposite for the less trained athletes since the individualization is less critical [65; 66]. Therefore, as shown in table 3.1, errors of 3, 5, 10, 15 and 20 seconds / kilometer are found to be acceptable for athletes with running paces above 3.5, 4 and 4.5 minutes / kilometer, respectively.

Looking from the population perspective (see Figure 3.7), the recreational runner population, as any other heterogeneous population, has individuals with features that are far from what it can be considered usual from the physiological perspective. Moreover, the efforts and resources required to improve the estimation power for athletes with less common features grow non-linearly. In this regard, as already stated, this work is directed to the majority of the recreational runner population. This means that there is also an acceptable error in terms of the percentage of the recreational runners for which the *virtual LT sensor* is valid. We define this *satisficing* threshold as the *system's acceptable error*.

Regarding defining this threshold, as it is usual in many other engineering problems, creating a system valid for $90 - 95\%$ (two standard deviations) of the target population, recreational runners in this case, is considered *satisficing* for two reasons. First, it covers almost every athlete and second, because going beyond these numbers usually is extremely costly in terms of the resources needed. Therefore, this criterion is used in the present work as a base *System's acceptable error*.

Then, the *individual acceptable error* is used to calculate the validity of every estimation and the computation of valid / invalid estimations are compared to the *system's acceptable error* to evaluate if the overall performance of the ML system is *satisficing*. Additionally, section 3.2 explains how these *satisficing* errors are validated.

3.2 Experimental methodology for database creation

This section sets the steps for a proper experimental methodology according to target population and the required information. As represented in Figure 3.10, we first make the necessary preparations for the experiment, so that it facilitates and ensures the proper collection of the relevant features of the target recreational runner population. Second, we define the athletic, health, legal and ethical requisites necessary to make the test. Finally, we state the criteria that are to be used to decide whether the experiments are valid for *satisficing errors* calculation and/or for modelling purposes.

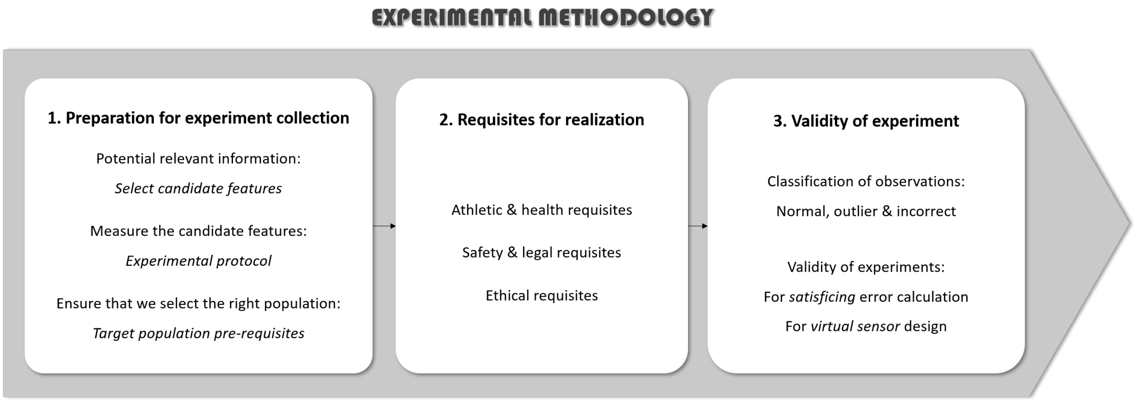


Figure 3.10: Steps for obtaining valid experiments

3.2.1 Preparation of experiments

The aim of making experiments is to measure the easily available features that hold relevant information about the LT and ensure that we do it for the target population. To attain so, the preparation of experiments is further divided into three steps.

First, as represented in Figure 3.11, the candidate easily measurable features are selected from the features with potential relevant information known in the literature.

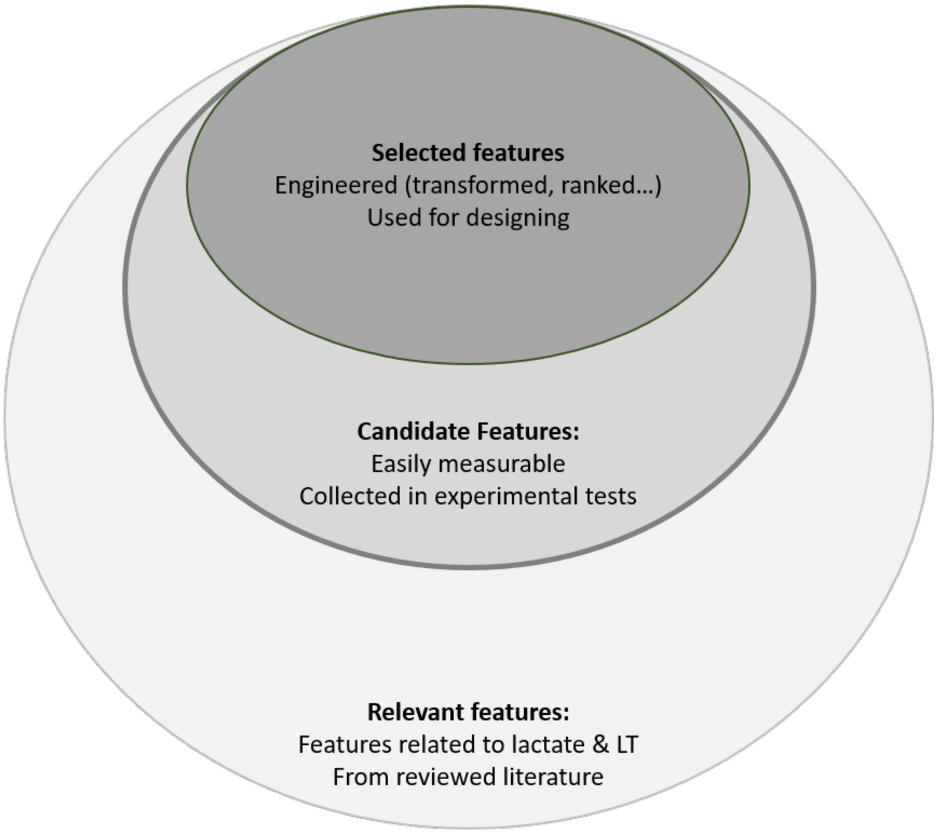


Figure 3.11: Relevant, candidate and selected features

Despite all the candidate features are probably not to be used for modelling, the final selected features will arise from this list in subsequent steps. Therefore, the second step deals with creating a protocol that allows to collect all these candidate features. Especially, it deals with the definition of the incremental treadmill speed test protocol to be performed for collecting the time-series data. Third, we define a set of pre-requisites that will ensure that the population that we are sampling corresponds to the target recreational runner population.

Candidate features

The selection of the candidate features is done according to expert knowledge so that those features with potential to have relevant information about LT are collected.

It is important to note that, not all these features are to be used for modelling purposes. The number of observations needed to correctly balance this amount of features is huge and not attainable for an efficient trade-off between the resources used and the potential value of the results of

Table 3.2: Candidate features: raw collected features during the incremental treadmill speed test

Feature	Time format
Discipline	static
Physical condition	static
High intensity interval training	static
Years train	static
Sex	static
Birth-date	static
Height	static
Weight	static
Body fat index	static
Abdominal diameter	static
Hip diameter	static
Fat percentage	static
Water percentage	static
Fat percentage	static
Personal best (IAAF points)	static
Vpeak	static
Resting HR	static
Maximum HR	static
HR values at different running stages	static
Maximum muscular Borg	static
Maximum respiratory Borg	static
Muscular Borg values at different running stages	static
Respiratory Borg values at different running stages	static
Resting Lactate	static
Lactate values at different running stages	static

Abbreviations: LT, lactate threshold; HR, heart rate; IAAF, International Association of Athletics Federations; Vpeak, Maximum velocity obtained in the experimental test.

this work provide.

However, there are three main reasons to collect most of these features. First and foremost, it allows us to make a descriptive analysis of the features that help us get a deeper understanding about the lactate and their related features from the physiological perspective. Second, during the feature engineering phase, dimensionality reduction and feature aggregation techniques may be considered. Finally, it also leaves an open door for the present work to be extended in future works by adding further observations. The complete list of features an their format is organized in table 3.2.

Additionally, based on expert knowledge, some additional well known features are created from the raw collected ones and gathered in table 3.3.

Experimental protocol

Among the multiple features that are defined as candidate to be used in the *virtual LT sensor* design we can make a separation between two kind of features according to their time characteristics:

Table 3.3: Candidate features: transformations from collected raw features

Feature	Time format
HRDP	static
HRRT	static
%HRmax at different running stages	static
HR evolution	time-series
HRR evolution	time-series
Muscular Borg evolution	time-series
Respiratory Borg evolution	time-series
Lactate evolution	time-series
Lactate threshold	static

Abbreviations: LT, lactate threshold

static and time-series.

On the one hand, the collection of the static features is quite simple since it can be done by means of questionnaires and direct measurement. On the other hand, as already explained, the determination of LT and the collection of the rest of time-series features is done by a incremental treadmill speed test protocol that must be defined prior to starting the data acquisition.

As any other protocol, the precise incremental treadmill speed test protocol used for individual Dmax LT determination arises from certain criteria that afterwards materialize in precise rules. In the present work we explicitly define these criteria from which the incremental treadmill speed test protocol rules are defined according to the expert knowledge of the experimenters. The criteria are described below:

- Criteria for heart plateau: 4 minutes of effort are considered necessary to be able to reach at the end of each stage to HR plateau, i.e. the stationary state in each of the running stages.
- Criteria for heart rate recovery measurement: HRR is among the features to be calculated and it has been observed that 1 minute of recovery is appropriate for measuring HRR [67].
- Criteria for LT finding: Avoiding sharp increments in lactate concentration is important for proper LT finding. To do so, the increments in speed are regulated.
- Criteria for warm-up without extra fatigue: The starting speed is important to let the athlete face the test in progression so that it enables a proper warm-up while avoiding excessive fatigue. To do so, the starting speed was determined in 9 kilometer/hour.

From the aforementioned criteria and considering the population under study, as illustrated in Figure 3.12, the protocol to be used in the experimental tests is defined as: A maximal incremental running test at 1% slope on a treadmill, started at 9 kilometer/hour without previous warm up. The speed is increased by 1.5 kilometer/hour until 13.5 kilometer/hour and then by 1 kilometer/hour until the participant reaches exhaustion. The duration of each running stage has been set in 4 minutes with 1 minute of recovery between them.

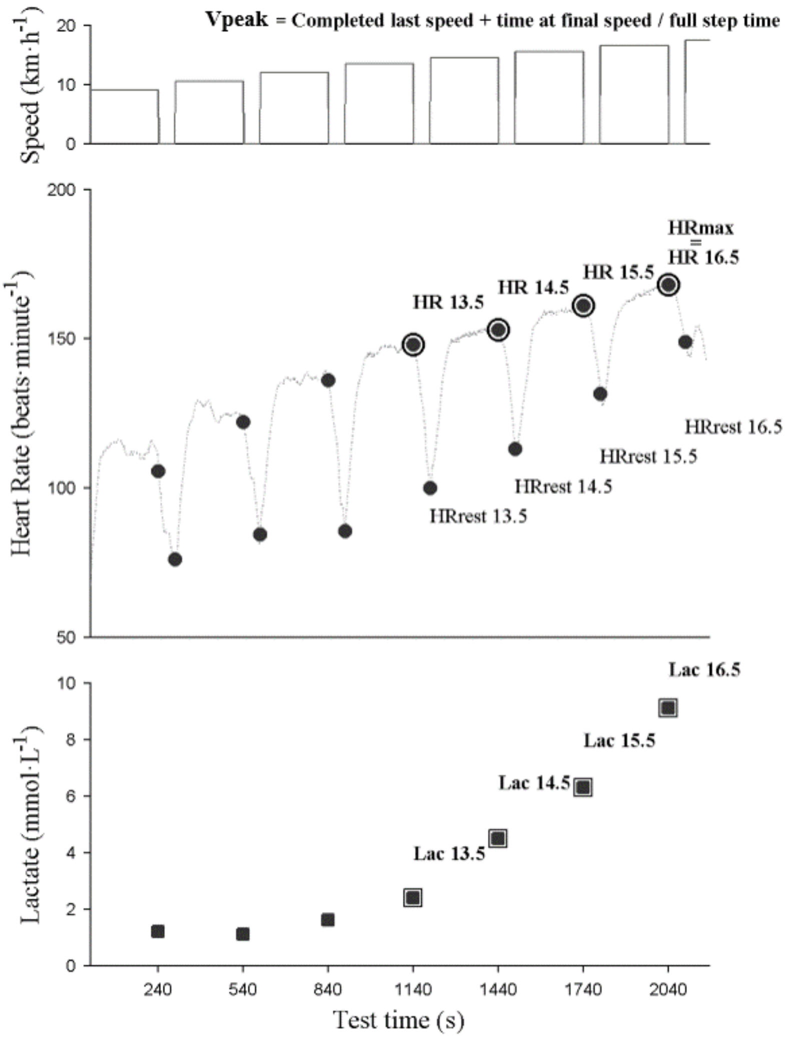


Figure 3.12: Representation of the incremental speed test protocol and heart and lactate related feature collection

Target population definition and pre-requisite formalization

As already defined, the *virtual LT sensor* is directed to recreational runner population. However, due to its heterogeneity, simplifying and clearly defining the boundaries of the recreational runner population is not straightforward. Therefore, prior to defining the pre-requisites for the target population, we shall first define the target population or 'what a recreational runner is' in terms of this work.

In this regard, the recreational population term has been commonly used to describe a wide variety of athletes which include different sets of: beginner, well-trained and sub-elite athletes. The recreational runner community interested in an operational lactate threshold estimation has certain characteristics.

The first characteristic of the recreational runners (from the perspective of the present work) can be derived directly from the population in demand of an operational LT estimator. These are athletes currently participating in long distance endurance races from 5 kilometers upwards. It includes several disciplines such as track, road running, cross country running, trail running and triathlon.

Second, there is both a lower and an upper athletic level that limits what can be considered recreational runner.

On the one hand, the usefulness of the virtual LT sensor for an athlete below certain athletic level is very limited and the athlete would better benefit from following a introductory training plan. In this work we set this lower limit in being able to finish the 14.5 kilometer/hour running stage in the treadmill speed test. On the other hand, there is a level above which the athlete can not be considered recreational.To determine the upper level, two sub-elite athletes were recruited and experiments made to be used as reference. Elite and sub-elite athletes are scarce and thus two well-known athletes in good physical condition where selected according to their recent athletic performances. The tests performed by the sub-elite athletes reached 20.5 kilometers/hour and thus this speed is used as upper limit.

Summing up all the previous considerations and as represented in Figure 3.13, the following athletic characteristics are requisites for an athlete to be considered part of the target population:

- Endurance athletes training for and participating in running races from 5 km upwards.
- Currently running at least 3 days a week.
- A running experience of at least 1 year.
- Athletic level according to the maximum running stage reached in the test herein assessed between 14.5-20.5 kilometers/hour.

Target population pre-requisites

Purpose:

Objective criteria to discern between recreational & non-recreational runners

Criteria:

- Participating > 5K running races
- Running > 3 days/week
- >1 year running exp.
- Level 14.5-20.5 km/h running stages

Figure 3.13: Pre-requisites to consider recreational runner population

3.2.2 Requisites for realization

The previously defined experiment preparations tries to ensure that the experiments are aimed to the proper population and that the appropriate features are selected.

In this section, we describe the requisites that every athlete participating in the experiments must fulfil prior to being allowed to perform it. These requisites include athletic, health, legal and ethical considerations that either protects the athlete and/or ensure that the experiments are performed under the appropriate conditions.

The athletic-health requisites that each athlete must fulfil includes:

- Be well rested and to abstain from hard training sessions and competition for 24 hours before testing.
- Abstain from eating for 3 hours before testing.
- Abstain from taking stimulant substances before testing, including coffee or tea.
- Be familiarized with running on a treadmill.
- Being healthy and lacking on infections.

Additionally, there are certain safety and legal requisites that must also be fulfilled:

- To be federated in their respective disciplines
- Provide a medical certificate that ensures that they are able to perform the test.

Finally, according to the Ethics Committee for Research on Human Subjects of the University of Basque Country UPV/EHU (CEISH/GIEB) that approved this study with M10/2015/203 reference number. The application of this protocol is subjected to fulfilling the following requisites:

- The participant has read the information sheet (see annex A).
- The participant has provided a written informed consent acknowledging that has been informed about the possible risks of the tests and giving their consent (see annex A).

3.2.3 Experiment validity definition

Despite great efforts are put into maximizing the quality of the experiments, there is always a chance to collect experiments that are invalid. The validity of an observation depends on characteristics like correctness and/or application type. Therefore, we first make a classification of different type of observations that we can encounter and that could potentially invalidate an experiment.

In any population, there are always some rare observations that fall out of what it can be considered as normality of a targeted feature. While these rare observations may still be correct, distinguishing them from incorrect observations is sometimes not straightforward.

Nonetheless making this distinction is fundamental, the validity of the observation may depend on its characteristics. Figure 3.14 is representative of the classification according to type of observations (gathered in columns) into *normal* and *rare*. Then, the *rare* is further divided into *outlier* and *incorrect*. Moreover, this distinction is done both at sample level (considering if the athlete is inside the target recreational runner population) and data level (if the data has been erroneously measured or estimated).

A difference between incorrect and outlier data may be illustrated using the HR measurement as example. On the one hand, an observation giving a negative value of HR or a HR of 300 beats per minute is clearly considered incorrect data, since they are not within the range of what is possible for the phenomena under scrutiny, and therefore would be categorized as invalid data. On the other hand, an unusual high HR value of 210 beats per minute may be due to the inherent characteristics of the athlete and therefore considered as outlier. In this work, we hold a conservative approach when labeling data as incorrect so that we only do so for the flagrantly incorrect as the former example.

Additionally, as represented in Figure 3.14, the incorrect and outliers may appear after any of the data handling steps and we divide the sources of error in three types: sampling errors (major contingencies and non-recreational runners), measurement errors (pulsometer, lactate measurement device error...) and transformation errors (%HR max, LT due to Dmax LT method error).

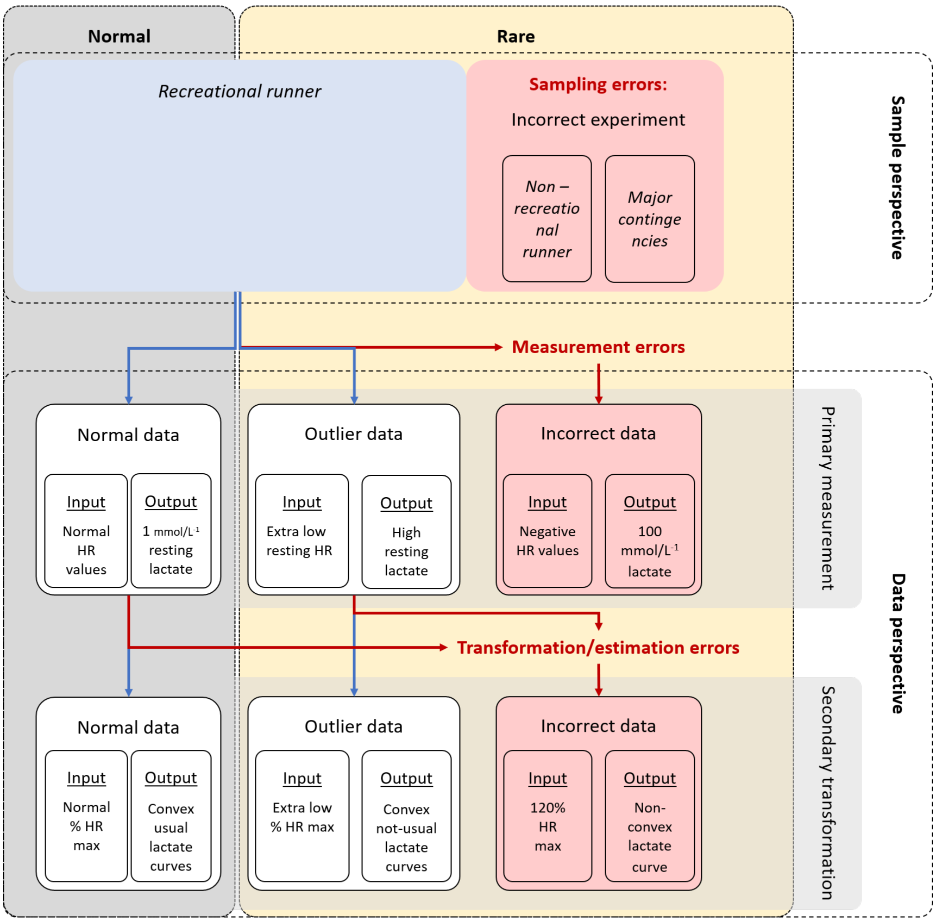


Figure 3.14: Categorization of correct and incorrect experiments and data

Therefore, Figure 3.14 sets a framework to evaluate which kind of data is valid for one of the two main purposes: calculating the *satisficing* error and designing the *virtual LT sensor*.

Validity for *satisficing* error calculation: individual and system's acceptable error

As already mentioned in section 3.1, there are two acceptable thresholds, *system's acceptable error* and *individual acceptable error*. These two acceptable errors were defined using expert knowledge. However, to add further robustness, these criteria is to be validated with the analysis of the LT threshold data. As represented in Figure 3.15, normal, outlier and incorrectly lactate data (measured lactate values and transformed LT values) are valid for this purpose.

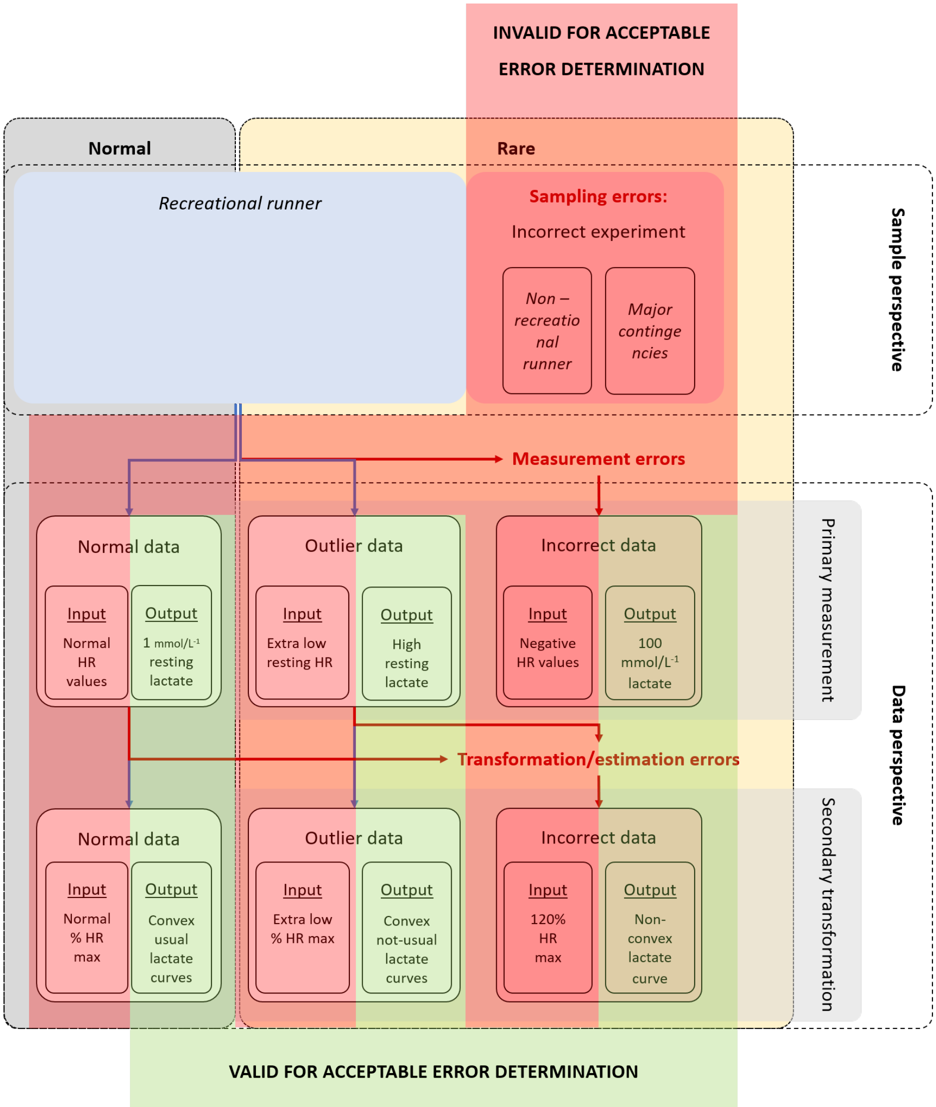


Figure 3.15: Valid data for individual and system's *acceptable error* determination

A technique such as the individual Dmax LT method that is used in practice serves as the proper reference to asses how much is "good enough" for nowadays standards. The Dmax LT estimation method, as any other methodology that tries to estimate or measure a variable, suffers from approximation errors. In the present work we refer to this error as the *unavoidable error*. This *unavoidable error* is part of the output of each observation that is used to create the ML system, so, it also sets the upper bound or the ideal performance that the *virtual LT sensor* could achieve. As represented in Figure 3.16, the *unavoidable error* has two components which can be directly mapped with the previously defined individual and system's *acceptable errors*:

- *Dmax LT method precision error*: The individual Dmax LT determination method produces precision errors in each observation that comes from the propagation of the *blood lactate measurement error* (a combination of lactate measurement device error, inherent errors of the physical measurement such as sweat in blood sample...) to the determined LT. The quantification of this error can be used as reference to validate the *individual acceptable error* based on expert knowledge. To do so, the normal and outlier Dmax LT estimations are used.

- *Incorrect Dmax LT rate*: The incorrect Dmax LT observations (see Figure 3.20) are valid since they give information about when the individual Dmax LT fails (those with non-convex curves...) for the sampled population and thus defines the limit that the individual Dmax LT has for its application to the entire recreational runner population. This limit is also to be used in combination with expert knowledge to set a robust *system's acceptable error* as it sets the upper limit that any ML system created on these labels has.

In the present work and as represented in Figure 3.16, the Dmax error analysis is used to validate the *acceptable errors* previously defined using expert criteria.

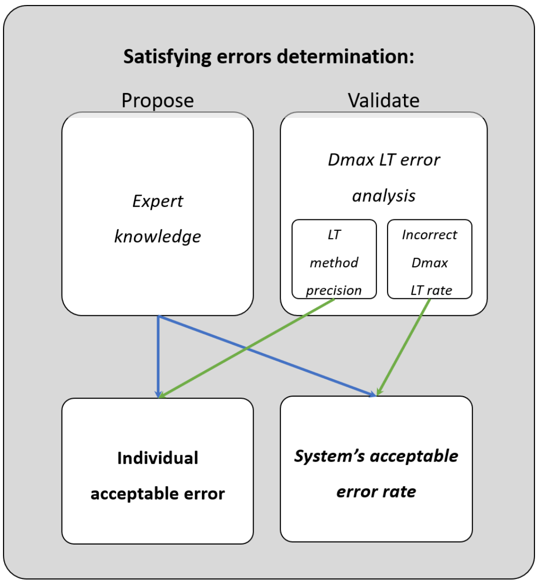


Figure 3.16: *Satisficing* errors: Validation of expert knowledge from Dmax lactate threshold error analysis

Individual acceptable error determination

The *Dmax LT method precision error* is well known in the literature from a measurement perspective [68; 34] and is here used as the validation reference for the *individual acceptable error*. The *Dmax LT method precision error* is related to the initial and final point selection, regression type, number of blood measurements... but it has not been quantitatively addressed so far. However, this error arises from a propagation of the *blood lactate measurement error*, which is dependent on unavoidable small errors such as the measurement device error, blood sampling errors related to sweat, timing... and has been well characterized and quantified in the work done by Tanner et. al. [69]. As represented in Figure 3.17, the *precision error* of the Dmax LT estimation is caused by how *blood lactate measurement error* propagates through the individual Dmax LT method to the determined LT.

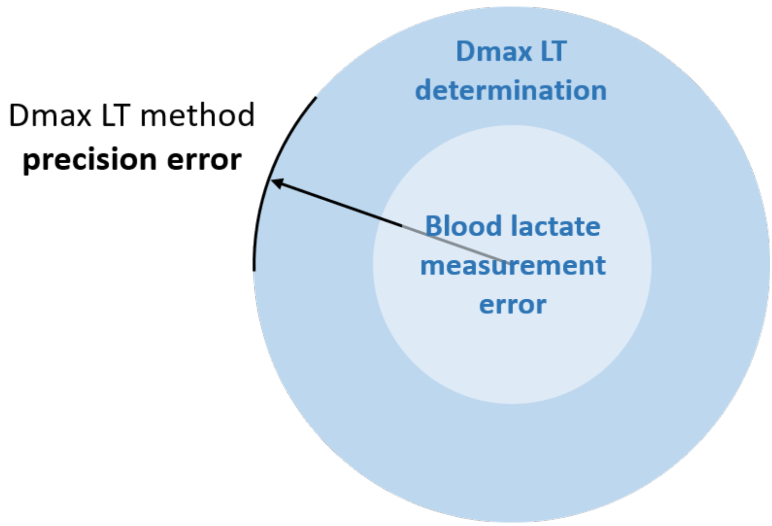


Figure 3.17: Dmax lactate threshold method precision error caused by blood lactate measurement error

Abbreviations: LT, lactate threshold

Since the process between the blood lactate measurements and the final individual LT estimation follows a defined protocol, we can formalize an algorithm that creates as many simulated LTs as required using hypothetical (hereafter plausible) measured blood lactates. Hence, in the present work we propose a computational algorithm to estimate the *Dmax LT method precision error* by unfolding how the *blood lactate measurement error* propagates and eventually materializes in the determined LT.

As represented in Figure 3.18, the accuracy of the blood lactate measurement can be divided in its trueness and precision components. As already mentioned, the blood lactate measurement device used in this work has been validated as an effective analyzer for lactate measurements [69]. Thus, it can be confidently stated that it has negligible trueness error. The rest of the possible sources of error (blood sampling errors related to residual sweat, inevitable small timing differences etc) are of random nature and thus they contribute to the precision component [70]. As illustrated in Figure 3.18, the precision error and its random nature can be represented as a probability distribution. The precision's standard deviation error has already been estimated by Tanner et. al. by making a test-retest reliability analysis [69].

Figure 3.19 exemplifies the computing process for the *Dmax LT method precision error* estimation of a single athlete. With the previously explained precision error as reference (Figure 3.19 step 1), the computational method starts from taking multiple ('Z' in algorithm 1) random samples from the blood lactate precision distribution to estimate new blood lactate measurements (hereafter plausible blood lactate measurements) (Figure 3.19 step 2). These plausible measure-

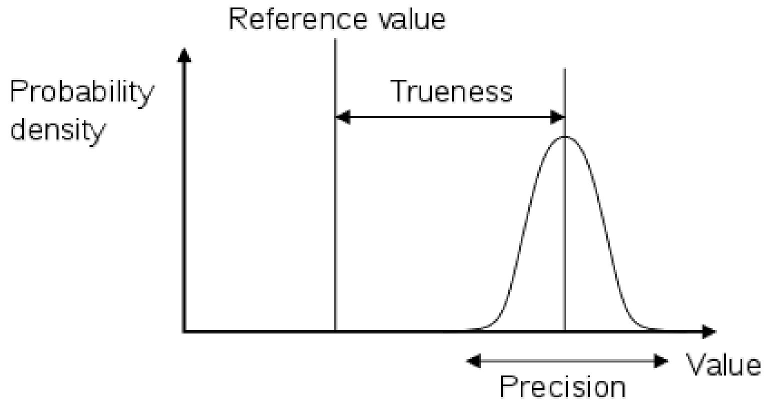


Figure 3.18: Precision error

ments for a particular athlete are represented under 'blood lactate measurements' name. In other words, these plausible blood lactate measurements correspond to the blood lactate measurements that could have been measured in the treadmill speed test due to the precision error of the blood lactate measurement device. Using these plausible blood lactate concentrations, their corresponding lactate curves are calculated (Figure 3.19 step 3), resulting in multiple different combinations of lactate curves that could have been derived for a particular athlete. These lactate curves are represented in Figure 3.19 under the 'lactate curves' name. Finally, the Dmax LT of each curve is calculated and represented in Figure 3.19 under the 'LT' name, which brings to light the inherent variability of the Dmax LT estimation protocol and the calculation of the Dmax *LT method precision error* for a particular athlete (Figure 3.19 step 4). To calculate the *Dmax LT method precision error* for the whole recreational runner population, this computational algorithm takes random samples with replacement from the sampled population to better represent the underlying population (Figure 3.19 step 5). Finally it calculates the mean of all the calculated precisions to obtain the final *Dmax LT method precision error* (Figure 3.19 step 6). Algorithm 1 formalizes this process.

This calculation process is repeated in each design iteration (as explained in section 3.3) so that the calculation grows in robustness together with the increase in sample size. Thus, the determination of this errors is done in Chapter 4 in each iteration. Then, this calculation is used to validate the *individual acceptable errors* defined according to expert knowledge stated in table 3.1. Therefore, the validation is done in Chapter 4. The W and Z re-samples are experimentally selected in Chapter 4 by increasing the size until reaching the "diminishing returns" phase. X is increased together with the sample size of different iterations.

***System's acceptable error* determination**

As in any other method there are certain recreational runners from which the application of the individual Dmax LT method derives into incorrect LTs from data perspective (see Figure 3.14). Analysing these incorrect LTs is a good way to estimate which proportion of the recreational runners can use the individual Dmax LT determination, serving as a good reference to validate the base *system's acceptable error* created with expert knowledge.

As we have seen before (Chapter 2), the lactate determination method selected in the present

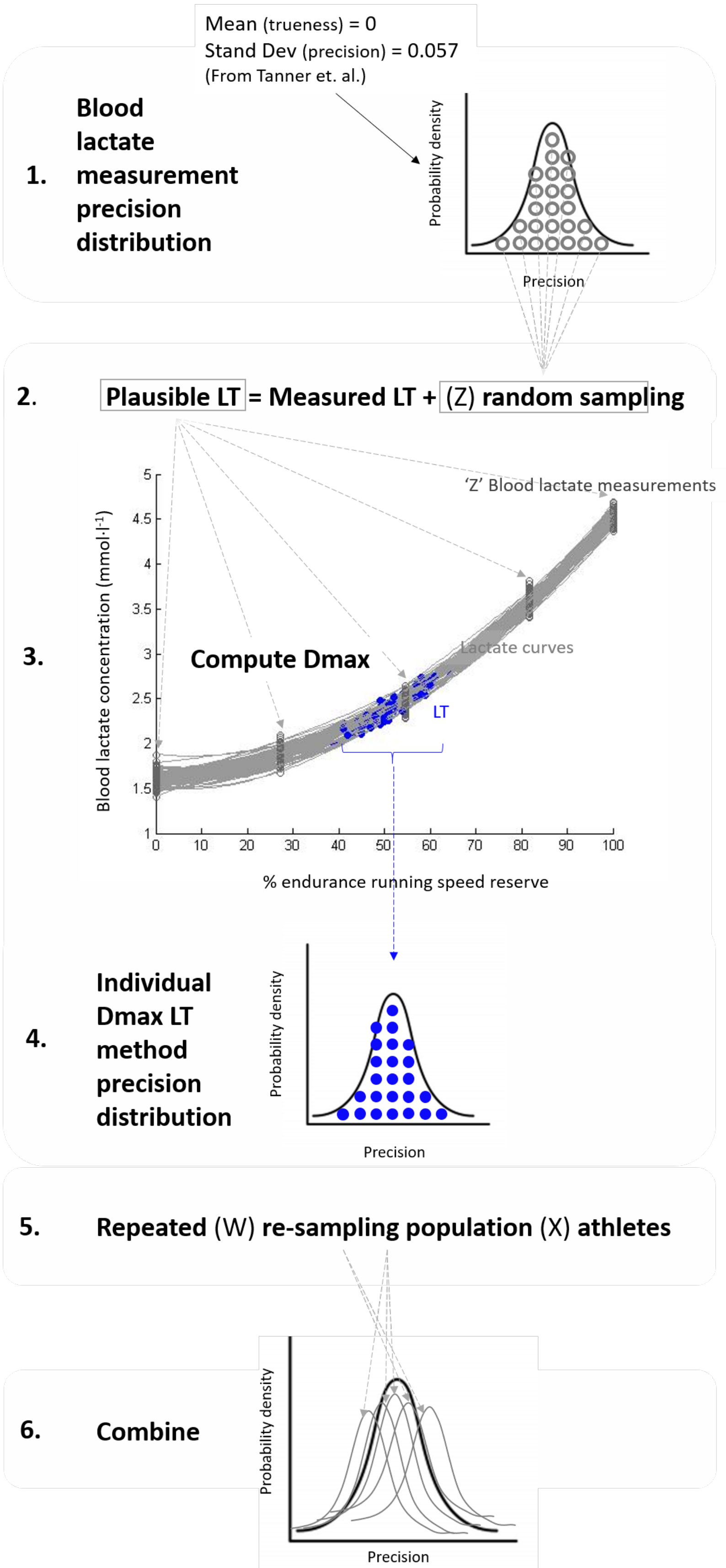


Figure 3.19: Computational *Dmax lactate threshold method precision* error calculation

Algorithm 1 Algorithm to compute the error caused by measurement: Dmax LT method precision error

Require: # lactate points per athlete
  SDMeasurement = Blood lactate measurement precision standard deviation ([69])
  Precision distribution = Normal distribution(SDMeasurement)
  for W bootstrap re-samples **do**
    for X athletes **do**
      for Y lactate points **do**
        for Z random samples **do**
          Plausible measurement error = Random sample from Precision distribution
          Plausible blood lactate concentration = Measured lactate + Plausible measurement error
        end for
      end for
      Plausible LTs = fDmax(Plausible blood lactate concentrations)
      Dmax error per athlete per re-sample = Plausible LTs - mean of Plausible LTs
    end for
  end for
  Dmax error distribution aggregating Dmax error per athlete per re-sample
  Dmax LT method precision = Standard Error Measurement of Dmax error distribution

work is among the most reliable methods and there is no ground truth with which it can be compared. Therefore, discerning between an outlier from an incorrect LT (see Figure 3.14) is not straightforward. In this work, a conservative ”least favorable” approach has been followed and only what experts considered flagrantly incorrect data is considered so. As represented in Figure 3.20, in our case this includes the lactate thresholds estimated on non-convex curves or highly undulated curves among others.

Thus, it can be used as a reference Dmax LT determination error and use it as a reference to set the *system's acceptable error* of our estimator, as well as to evaluate the maximum room for improvement of our system. Similar to the *individual acceptable error*, this calculation process is repeated in each design iteration (as explained in section 3.3) so that the calculation grows in robustness together with the increase in sample size. Then, this calculation is used to validate the *system's acceptable error* defined according to expert knowledge (90-95% error). This validation is done in Chapter 4.

Validity for *virtual LT sensor* design

As represented in Figure 3.21, the validity of the observations for designing the *virtual LT sensor* consists of detecting the incorrect observations at every level. More precisely, incorrectly sampled athletes (non recreational runners), incorrect measurements or incorrect transformations that compromises a relevant feature may invalidate the entire experiment.

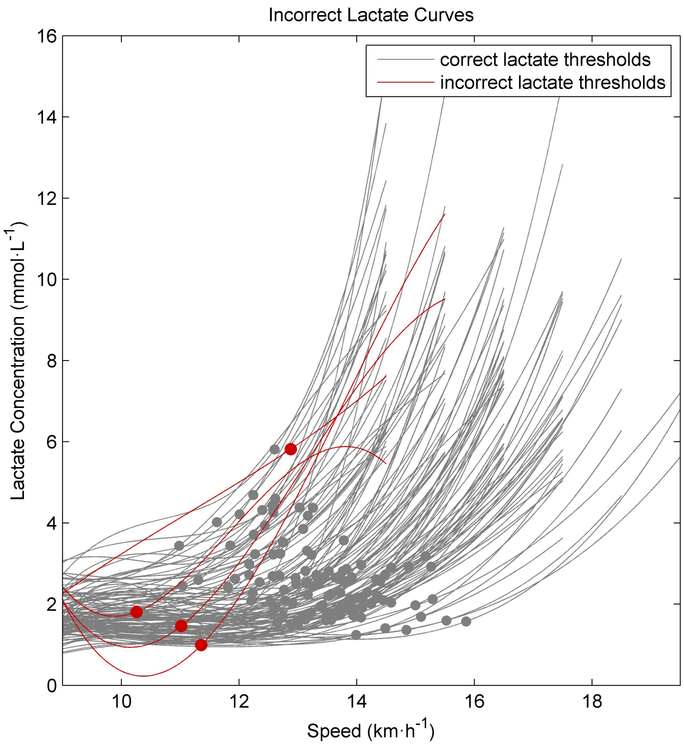


Figure 3.20: Incorrect data from Dmax lactate threshold method perspective

3.3 Methodology for *virtual lactate threshold sensor* design and development: characterize, represent and decide

The methodology here proposed describes the intra-iteration process. The objective of each iteration is to create a virtual lactate threshold sensor and to do so, as already mentioned and illustrated in Figure 3.6 section 3.1.1, this process is divided into context characterization, content representation and next step decision making.

As illustrated in Figure 3.22, the steps are divided according to two main inherent difficulties that creating a ML system has, plus a step that systematizes the decision making process for the next step to take. The purpose of these steps can be summarized as:

1. Context characterization: How and which data we collect (features and observations) to best represent the LT phenomena by minimizing the selection biases that may be introduced.

2. Content representation: This step deals with making the best use of the data so that the function that ML system infers best approximates the underlying function. This entails maximizing the robustness of the ML system as explained in section 3.1.

3. Deciding next steps: The final step deals with formalizing a method for deciding which is the most appropriate direction that the next iteration should follow (if any). More precisely, the decision-making process that leads to accepting, improving or definitively stopping the design of the *virtual LT sensor*.

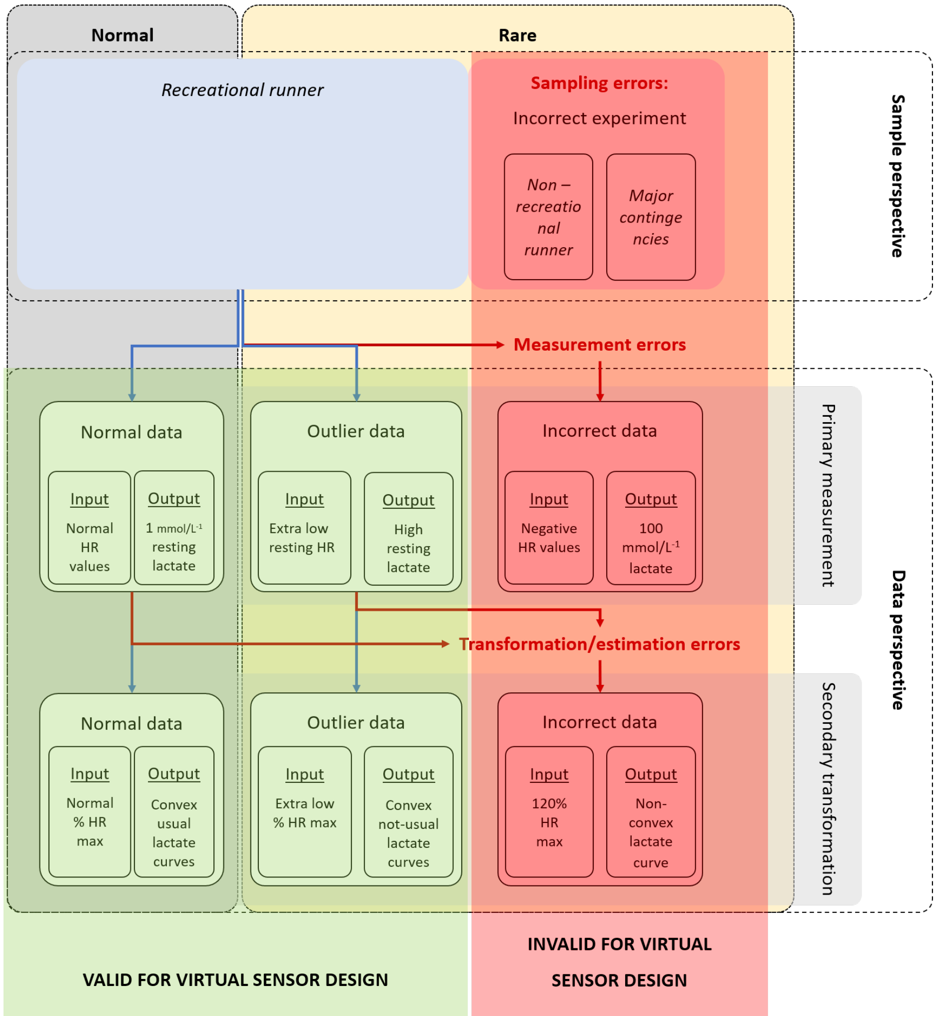


Figure 3.21: Valid data for *virtual lactate threshold sensor* design

3.3.1 Context characterization: Collecting high quality data

The quality of the gathered data is one of the key characteristics for a successful ML solution and where most of the effort must be focused on. Thus, the purpose of this step is to maximize the quality of the data collected in terms of its relevance with respect to the LT. In other words, to characterize appropriate context so that we have enough relevant information to make the inference about the outcome of interest.

Collecting quality data implies to select and collect the features and corresponding observations that have a close relation with the outcome of interest as well as to do it minimizing the selection bias. Selection bias is the error introduced in the selection process of individuals in such a way that the data sample obtained does not properly represent the population intended to be analysed. In our case, this can come from improper selection of the athletes and/or improper collection of feature-observation pairs.

A proper experimental design is therefore fundamental to minimize this source of error and to create a robust *virtual LT sensor*. To do so, given that the population of interest, i.e. recreational runners, is correctly defined and the protocols created (see section 3.2), it is fundamental to ensure

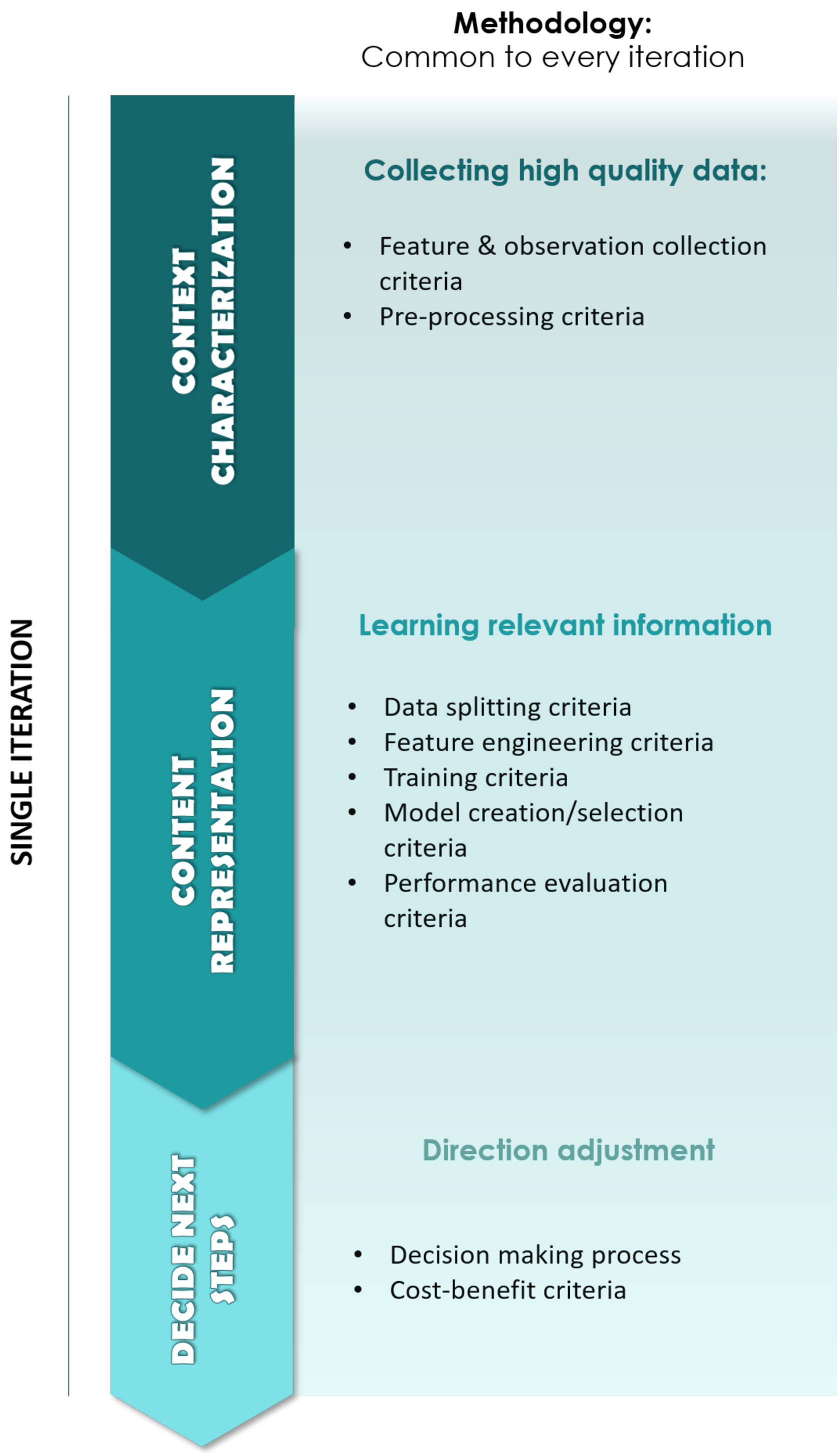


Figure 3.22: Methodology: Context characterization, content representation and decide next steps

that proper randomization is attained during data collection so that the chance to get features with spurious relations minimized.

As represented in Figure 3.23, we first decide how many and which features and observations are collected according to the needs and specific objectives of each iteration. The data is then collected according to the experimental protocol. Finally, a pre-processing is made for detection and cleaning of invalid data (see Figure 3.14), for getting overall knowledge of the information collected and adjusting the *satisficing* errors.

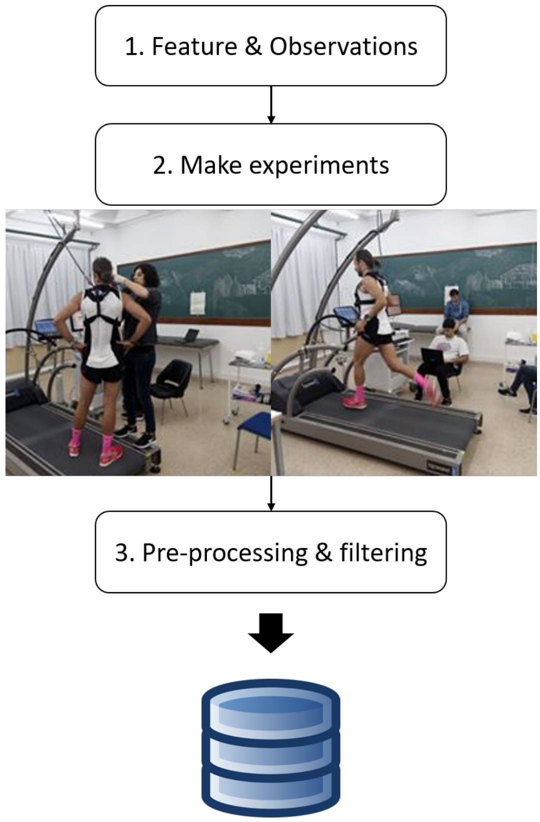


Figure 3.23: Context characterization: Steps for collecting quality data

Criteria for database creation: Features and observations

An inherent difficulty when creating a database is that the collected data always contain a combination of relevant information and noise. The purpose of the database is to gather and organize the input-output features with their corresponding observations with a minimized noise. Given the target population is clearly defined, the noise minimization is dependent on 1) collecting features that are relevant for the LT and 2) ensuring a proper randomization of the collected observations using appropriate sampling and re-sampling techniques.

Indiscriminate collection of multiple available features may seem reasonable a priory, since the more features we collect the more variability of the outcome could be potentially explained. However, it is known that this kind of approaches that do not use any criteria may lead to dimensionality problems [71] and end up finding spurious relations between the input and output features that do not generalize outside the sample. One of the solutions to minimize the risk of falling on dimensionality problems is using expert knowledge as a filter of features. In our case, it is known that multiple features such as HR related ones, perceived effort, diet, physical condition, age, athlete level etc are related with lactate production or LT [52]. A first filtering was already done when creating the experimental methodology in section 3.2. However, further filtering may be necessary at iteration level and this knowledge is thus used also as the additional filter of features.

As already explained, the features must be accompanied with the appropriate amount of observations to avoid to violate the fundamental sampling laws such as law of large numbers and central limit theorem [72; 73]. If these laws are not respected, the observed variability of the input features can end up being due to randomness and may interfere and make the inference engine misrepresent the information [71] and be unable to generalize to the recreational runners population.

But, how many observations are enough for this noise to be minimized? So far, several heuristics such as the ”one in ten rule” (10 observations per feature), the commonly used 30 observations for independent variables etc have been proposed in these endeavour. However, these heuristics are based on multiple assumptions of ideal conditions such as independence, normality of distribution... and shall be treated as a rule of thumb [74]. Moreover, it is well known that, in order to obtain a statistically sound and reliable result, the amount of data needed to support the result often grows exponentially with the dimensionality [71]. Therefore, despite these heuristics are not able to ensure how many observations are sufficient, they serve as a valid reference to determine a lower threshold of observations.

In any case, the number of observations needed is highly dependent on the problem being solved and thus the iterative approach here proposed goes hand-in-hand with this inherent uncertainty, since the iterative strategy allows to incrementally test the sample size by a posterior evaluation of the ML system performance for increasingly robust conclusions. Therefore, the number of observations will be determined in each iteration taking into account these criteria.

Moreover, as explained in section 3.2 and represented in Figures 3.15 and 3.21, the observations gathered in the database are susceptible to be invalid for both acceptable error calculation and designing purposes. To maximize the quality of the data, these invalid data is to be minimized.

To do so, first a detailed protocol is created to systematically capture, digitalize and organize data. This data acquisition protocol can be divided into: preparation, calibration & start-up of the necessary tools, confirmation of pre-requisite compliance, formatting, static feature collection, time-series feature collection, failure case protocol and digitalization. The detail of the data acquisition protocol can be found in annex B. The pre-processing step helps to detect the remaining incorrect data at sampling and data perspectives (see Figures 3.15 and 3.21) and filter out the corresponding invalid data.

Pre-processing

Given the endeavour of maximizing the quality of our data, pre-processing relates with all the initial analysis that allow us to get an overview of the sample, make early detection of invalid data and filter it out. Moreover, pre-processing also serves the purpose of getting additional knowledge of our problem by making a preliminary descriptive analysis of our sample.

This process starts with the format merging, on which each experimental tests and each results are scrutinized in detail. To do so, every experimental test is independently and one-by-one analysed to look for flagrant errors and inconsistencies between the redundant formats (pictures, excel files and handwritten documents) that were used to gather the data.

With regards of detecting and filtering out the corresponding invalid data, this pre-processing follows the same sequence than the data collection. Looking it from the samplimg perspective (see Figures 3.14), the two sources of error are:

1. Major contingencies during the experimental test: The most evident source of error and one that completely invalidates the collected data comes from any major contingency that may happen during the test. Some examples are unfinished tests, flagrantly sub-maximal tests...

2. Non-recreational runner population: The remaining data, despite being collected by following the pre-requisites defined in section 3.2, as represented in Figure 3.24, may still contain non-recreational runners. More precisely, there is always certain subjectivity in the selection of recreational runners participating in the experiments, in this case concerning the athlete level. As already defined in section 3.2, the athlete level can only be estimated a priory and thus, it may show a different performance than expected in the experiments (Vpeak below 14.5 kilometers/hour or above 20.5 kilometers/hour).

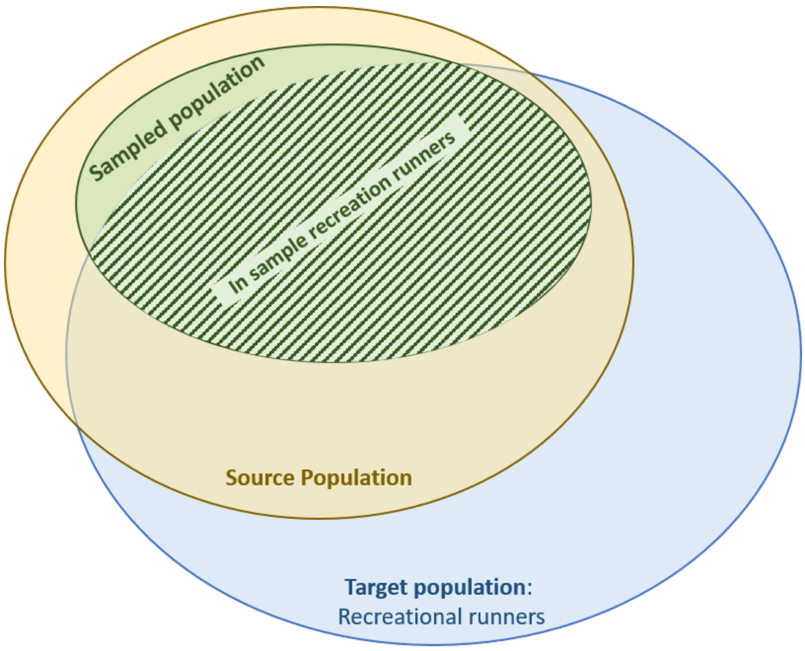


Figure 3.24: From target population to sampled recreational runners

The incorrect experiments detected in these two steps are invalid for both error calculation (see Figure 3.15) and designing purposes (see Figure 3.21), therefore are to be filtered.

Once that the invalid data coming from sampling errors has been filtered, the focus is placed in the sources of error from the data perspective. Despite the big efforts made to maximize the data collection quality, there are unavoidable errors that we divide this source error into (see Figure 3.14):

1. Primary measurement error: These errors are related to any measurement error done to acquire the raw features that are gathered in table 3.2. As already mentioned, the output incorrect data coming from measurement errors (i.e. incorrect lactate values) are still valid for acceptable error calculation (see Figure 3.15) and are reserved for this purpose. The rest of the incorrect data is invalid for both acceptable error calculation and designing purposes (see Figures 3.15 and 3.21) and are thus discarded.

2. Secondary transformation error: These errors are related to any transformation error done in the process to convert the raw features gathered in table 3.2 into the features gathered in table 3.3. The incorrect lactate curves and posterior incorrect LT are among these features. As already mentioned in section 3.2, this output data remains valid for *satisficing error* adjustments and thus it is reserved for this purpose. Notice that at this level, "incorrect" is understood in terms of the Dmax LT determination method: those experiments are considered incorrect because the Dmax LT determination method does not suit them as long as the lactate points do no exhibit a convex behaviour.

To detect and identify the incorrect data, apart from a one-by-one analysis of the data, in each iteration, an uni-variate descriptive analysis of the data is performed to illustrate the distribution of each feature. This allows to have an overview of each feature, facilitating the detection of the incorrect data that stands out. Then, using the criteria above mentioned, the invalid data is discarded. The descriptive analysis is enriched with graphical tools that allow to look at the data from other perspectives and elucidate further characteristics.

Additionally, the aforementioned uni-variate statistical descriptive analysis serves to get additional knowledge about our problem by looking to the distributions of the collected features. For instance, looking to the mean and standard deviation of the features may give additional information about the characteristics of the recreational runner population. With the same purpose of acquiring additional knowledge, an exploratory bi-variate analysis is also done. This analysis seeks to throw some light on how the collected features are interrelated by analysing the cross-correlations between them. Furthermore, these correlations are to be placed in the same map to facilitate its interpretation. This information may be valuable in future steps to make a preliminary idea of the context that we are working on in aspects such as the redundancy of the information gathered in the features.

These analysis are done according to the specific needs of each iteration, and therefore is specified in the design phase (see Chapter 4).

3.3.2 Content representation: Learning relevant information

Once the context characterization phase is finished, we have a finite amount of content collected in our database. In spite of doing great efforts in the data collection step, the database always contains variability that cannot be explained with the collected independent features. In other words, our database contains at the same time relevant information (signal) and random or spurious information (noise). Therefore, the purpose of this content representation phase is to make the most from the data we already collected by representing only the relevant part while filtering out the spurious.

To do so, ML tries to match the relevant degrees of freedom contained in the database (in form of relevant features) with the degrees of freedom of the algorithm to be used to infer the underlying relationship. In our case, aiming for robust learning, we put great efforts into every step that reduces the chances of overshooting the degrees of freedom of the algorithm and consequently creating an over-fitted model. To do so, introducing diversity in every step of the learning is fundamental [75]. As represented in Figure 3.25, this content representation can be divided into five highly intertwined parts.

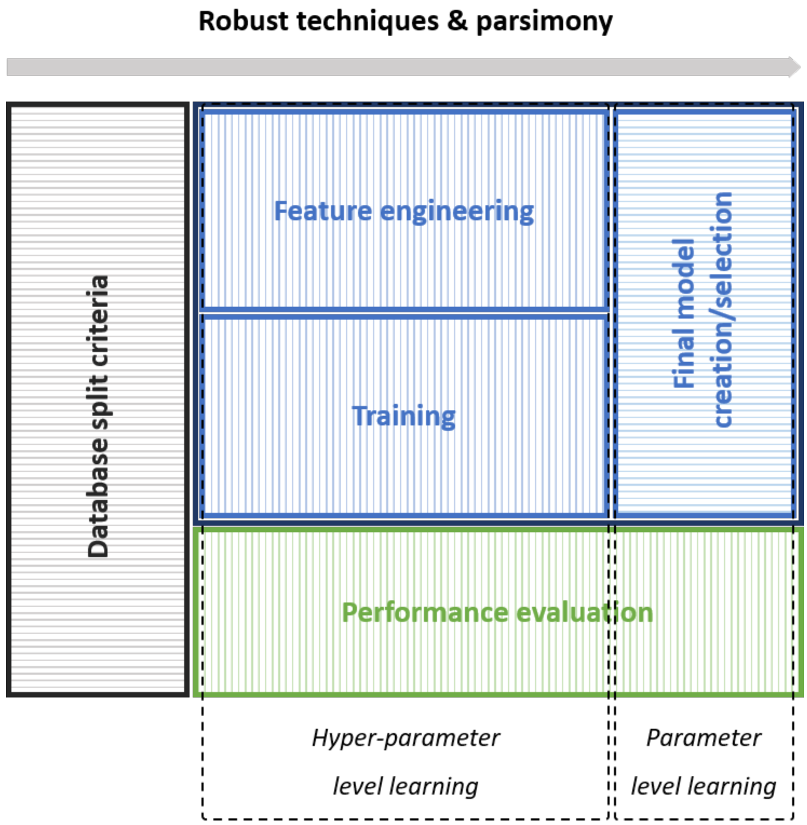


Figure 3.25: Content representation's sub-parts

Splitting the database into different non-overlapping chunks is the foundation on which machine learning stands. A model directly trained in the whole data sample would be unable to estimate the model performance on unseen data, and the chances of creating over-fitted models would significantly increase due to the lack of data diversity. Therefore, the data base splitting serves two purposes: robust learning and performance evaluation. Moreover, as represented in Figure 3.25, it is done to serve learning and evaluation in different layers, at hyper-parameter and parameter levels more precisely.

Making use of the split database, training, feature engineering and final model creation/selection are the three inferences that a ML methodology makes.

Feature engineering and training are two sides of the same coin (matching the relevant degrees of freedom of the data with the degrees of freedom of the algorithm). On the one hand, feature engineering tries to do it by selecting the relevant degrees of freedom of the feature space. To do so, features are created, transformed and combined to maximize its relevance with respect to the outcome of interest (LT in this case). On the other hand, training plays with the degrees of freedom of the algorithm (hereafter model hyper-parameters) and fit the *virtual LT sensor's* parameters to the input-output data. Finally, the last inference relates to how a model is created or selected from the entire set of models created in the previous step.

Performance evaluation goes in parallel to the different inferences providing feedback of the performance of the system at different levels (model hyper-parameter and parameter). First, the performance is assessed through the bias-variance of the different model hyper-parameters combination. Second, the appropriate hyper-parameters are selected for the final model creation or selection. Third, the final model is evaluated by comparing the performance in the data split used so far (training set) with a test set that remained outside the learning.

Finally, as mentioned in section 3.1, maximizing the robustness of this learning methodology is a necessary condition for observing the real *system's error* and consequently evaluating the applicability of the *virtual LT sensor*. In this regard, the principle of *parsimony* says that, in equal conditions, the most parsimonious approach reduces variance of the ML system, making the solution more robust to unseen data compared to other more complex ones [76]. Moreover, a parsimonious system is much easier to understand and adapt due to its simplicity and consequently low computational cost. Thus, *parsimony* is to be applied in every step of the content representation.

Since every step of the content representation is intimately intertwined to each other, the selection of the techniques to be used correspond to the design phase, when the appropriate techniques are to be decided in parallel. Therefore, the following sections make a high level analysis to define the perspectives that are going to be followed in the design phase in Chapter 4.

Database splitting criteria: Foundation for robust learning and evaluation

To make the two layer learning and performance evaluation possible (at hyper-parameter and parameter levels), the database splitting is done in two steps.

As illustrated in Figure 3.26, the first database splitting is done to separate the data set used for model training and leave the remaining data unseen for its use to test the performance of the final model.

In order to make this separation robustly, a common approach in ML is to start by randomly splitting 80% for training purposes and 20% for testing purposes which is to be used as our first design approach.

The second split relates to the hyper-parameter level learning. As represented in Figure 3.27, using only the data previously reserved for training purposes, ”n” models (*virtual LT sensors* in our case) are created with ”n” combinations for one specific set of hyper-parameters. The ”n” parameter combinations can come from different initializations of the learning algorithm and/or from different re-samples of the training data.

Similar to the context characterization step, one of the main ideas behind making multiple models from different re-samples is increasing the randomization, so that the effect of the noise contained in the data is minimized and the robustness of the learning and evaluation increased.

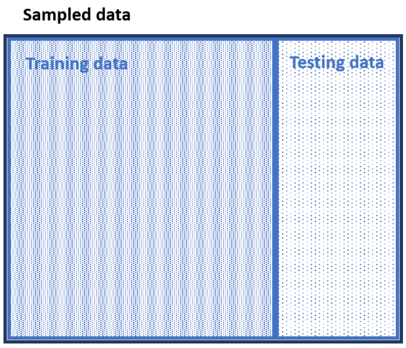


Figure 3.26: Separation of sample data into training and testing sets

Furthermore, the models are also created using 'm' number of hyper-parameters so that the appropriate trade-off between bias-variance can be found according to them. Figure 3.27 shows how these two kind of parameters are fitted during the training, illustrated by the example of a polynomial model. More precisely, in the example "n" sets of parameters are fitted for each hyper-parameter combination using the corresponding training set splits for a resulting m x n number of models.

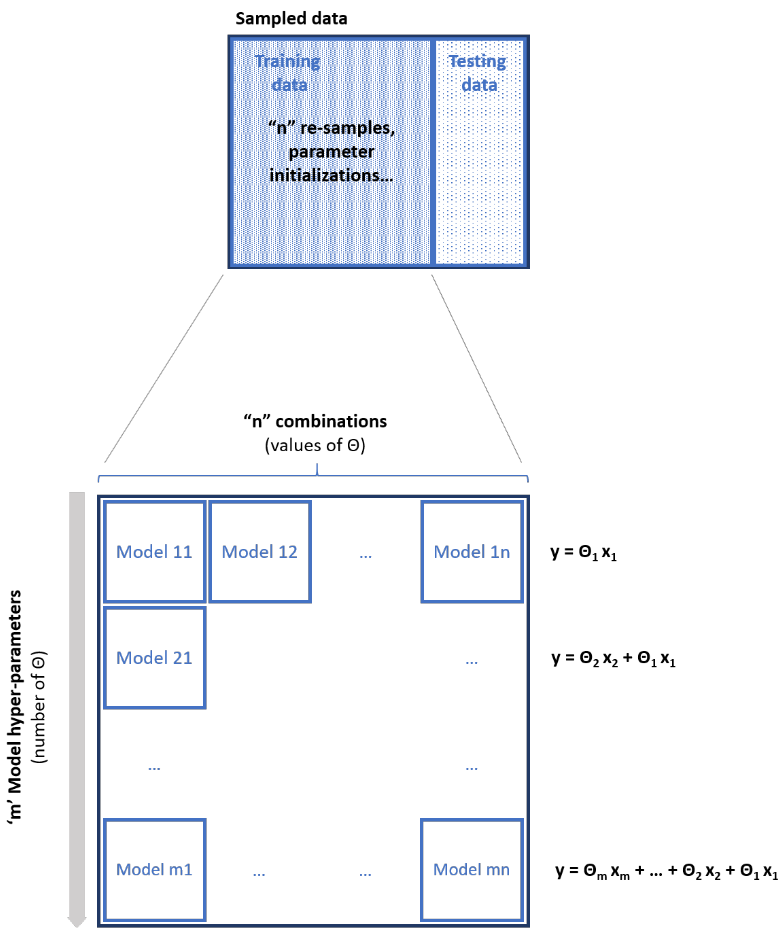


Figure 3.27: Data base splitting for hyper-parameter learning and evaluation

In order to maximize the robustness, following the principle of parsimony, the lower the number of hyper-parameters the better. Therefore, the training set split must focus on providing enough significant splits to enable the appropriate evaluation of the bias-variance corresponding to each hyper-parameter combination.

As the selection of the database splitting is highly dependent on the learning approach the selection of the exact approach is to be done in the design phase in Chapter 4.

Feature engineering: Finding the relevant features for LT

The purpose of feature engineering, i.e. the creation, transformation, reduction, selection and combination of the available features, is to make a proper representation of the feature vector. It is well known that most ML performance is heavily dependent on this representation [77] and that, if properly done, can greatly simplify the efforts needed in the training step.

Proper feature engineering is closely related with finding the features that are relevant for the output of interest (LT in this case) and extract the complementary information that these features provide. In other words, it tries to find the relevant and use only the necessary degrees of freedom, while discarding the rest. So far, multiple methods have been used in this endeavour.

Filtering feature engineering methods on their part can be used to make an explicit analysis of the feature relevance and subsequently to train the model. This approach includes techniques such as information gain, correlation coefficient, mutual information... and are independent on the training [78]. Wrapper or embedded methods such as recursive feature elimination or addition, lasso regularization... make the feature relevance implicit to the model training step.

As already said, when selecting the features from available data we are also making an inference from our limited sample, with its associated over-fitting risk. Thus, in line with the criteria for the selection of features in section 3.3.1, in the present work we use again expert knowledge to make a sub-selection of features and add robustness to the conclusions derived from the rest feature engineering methods. Additionally, the parsimony criteria is also to be applied here for maximized robustness which means that the lesser features needed the better.

Training a model: Learning the relation between input and output features

The training step deals with the selection of the approach used to learn the relationship between the input features and the output.

In this regard, there are multiple types of learning approaches such as the connectionists (neural networks, reservoir based... Figure 3.28), instance based (k nearest neighbours in Figure 3.29), statistical (Bayesian approaches, generalized linear models...Figure 3.30) etc.

Each learning approach provides with a particular set of complexity and characteristics which can be more or less appropriate for the problem in hand according to the structural characteristics of the relevant information collected in data. For instance, artificial neural networks (ANN) are universal approximators, which means that a simple neural network can represent a wide variety of functions when given appropriate parameters. Instance based learning approaches, instead of performing explicit generalization, compares new problem instances with instances seen in training, and stored in memory. Therefore, the suitability of the learning approach depends on the problem characteristics and is highly experimental.

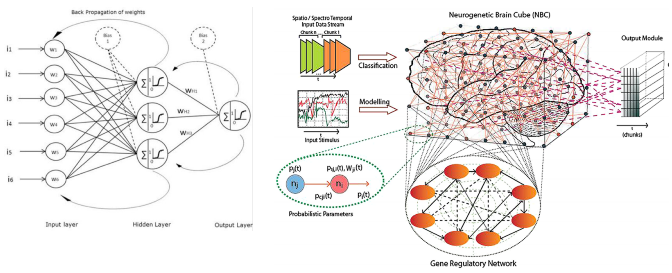


Figure 3.28: Connectionist learning approach example: Artificial Neural Networks and NeuCube

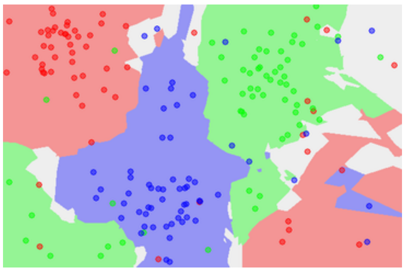


Figure 3.29: Instance based learning approach example: k Nearest Neighbors

The principle of parsimony also applies here to minimize the number of parameters that need to be fitted with the data and thus to create a more robust system. For most ML systems, the quantification of the parsimony of our system is done by counting the number of internal parameters of the ML system, the fewer the number of parameters the more parsimonious the ML system. In Figure 3.31, we show an example of how the parsimony of a linear model would be calculated according to its 'n' number of parameters. The higher the 'n' the less parsimonious model.

Therefore, a parsimonious learning could be achieved by selecting a sufficient low number of hyper-parameters of the learning algorithm. Additionally, regularization methods are also useful for this purpose, which introduce a penalty for exploring certain regions of the function space. Early stopping, regularizing for sparsity and other implicit regularization like Lasso are well known.

Performance evaluation at hyper-parameter level: Bias-variance of the methodology

Similar to the previous inferences, the performance evaluation is done at every level, i.e. at hyper-parameter and parameter level.

Usually the error analysis is done in raw error terms such as mean square error or any other metric detached from the application. However, to quantify the performance from the applicability perspective, the *satisficing* error perspective must be introduced into this analysis (see Figure 3.7 in section 3.1). Therefore, prior to making any evaluation, we introduce the method by which the raw error of the models are to be transformed.

To do so, the estimation error on each athlete (hereafter *individual error*) is calculated comparing its real lactate threshold and the estimated lactate threshold (5).

individual error = real individual lactate threshold - estimated individual lactate threshold (5)

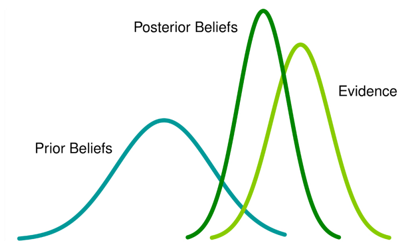


Figure 3.30: Statistical learning approach example: Bayesian inference

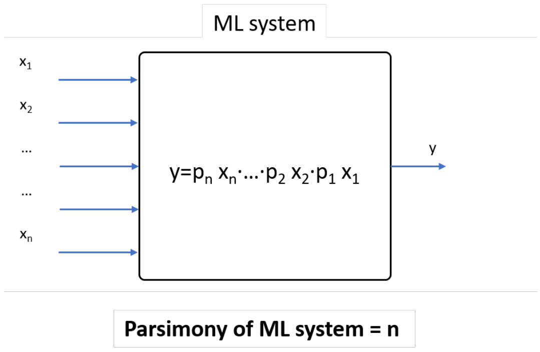


Figure 3.31: Parsimony quantification: An illustrative example using the number of parameters of a linear model

As illustrated in Figure 3.9, if the *individual error* is above the *individual acceptable error* this estimation is considered non-acceptable. Having all the athlete's estimations been classified as acceptable or non-acceptable, the *system's error* is determined calculating the number of athletes which accomplish the *individual acceptable error* and represented as a % of the total athletes in the database. This percentage is considered the *system's error* of our system. Algorithm 2 formalizes this computation and Figure 3.32 illustrates it with an example.

Algorithm 2 Compute system's error

```
for N number of athletes do
    if individual error > individual acceptable error then
        non-acceptable estimation = non-acceptable estimation + 1
    end if
end for
System's error = (N - non-acceptable estimation) / N * 100
```

With this transformation the application perspective is intrinsic to the metric to evaluate the performance and we are able to jump into the bias-variance estimation of the methodology. Despite there are multiple methods (learning curves, cross-validation, Akaike information criterion...) to evaluate the bias and variance of the learning methodology, most of them rely on analyzing the combined errors of the 'n' created models (see Figure 3.27). More precisely, the variance is obtained by observing how the error varies across the 'n' models (for instance, measuring its standard deviation) and the bias by computing the mean error of the 'n' models. Let Figure 3.33 serve as an illustration of how the bias-variance evolves according to the number of hyper-parameters.

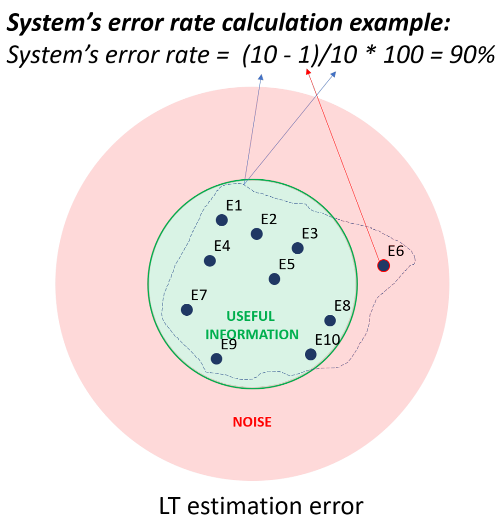


Figure 3.32: *System's error* calculation

Abbreviations: LT, lactate threshold; Ex, estimation error on athlete x

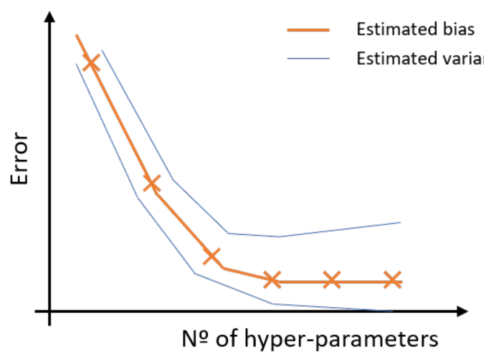


Figure 3.33: Example of bias-variance estimation for each hyper-parameter set

Hence, the diversity we introduced in the database splitting step by creating multiple models allows to make the evaluation of each hyper-parameter combination from the bias-variance perspective. The robustness of this estimation depends mainly on how well the diversification was done, the higher the re-sampling and initializations the better. Therefore, it places a greater emphasize in its importance.

Model creation or selection: Hyper-parameter selection and final model creation

Once the hyper-parameter level learning is finished (i.e. the feature engineering and the training) and the performance of the methodology estimated, the next inference is to select the model hyper-parameters and create or select the final model from it.

A parsimony based hyper-parameter selection is a common practice for increasing the robustness of this inference. This selection is usually done according to the point of diminishing error reduction with respect to the number of hyper-parameters of the model. To do so, expert knowledge or approaches such us the elbow method are commonly used. Figure 3.34 represents the application of the elbow method to find this point.

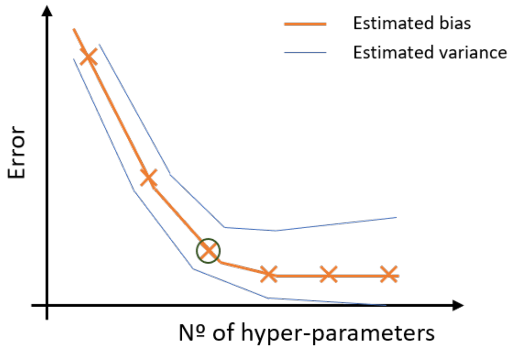


Figure 3.34: Hyper-parameter selection according using the elbow method

However, in our case, consistent with the preference of an under-fitted model as explained in section 3.1 (see Figure 3.7), we introduce an additional criteria for the hyper-parameters selection to maximize the robustness of the system. To do so, the *system's acceptable error* is used as reference and the most parsimonious hyper-parameter combination that fulfils it is selected (see Figure 3.35).

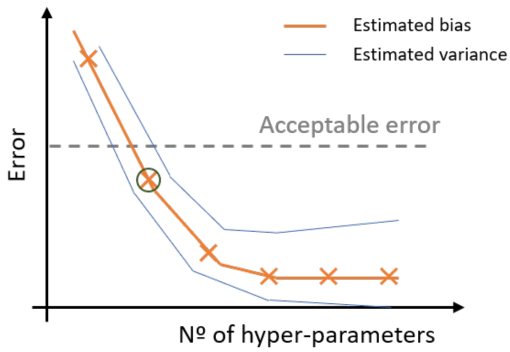


Figure 3.35: Evaluation of the bias-variance of a polynomial regression according to different number of hyper-parameters

This concept extends also to other methods such as ensembling that may be used to create the final model fusing multiple ones. The hyper-parameters of the ensembling approach are also to be taken into account during the design phase to maximize the robustness of the final model.

Performance evaluation at parameter level: final model testing

The final model performance is evaluated from the comparison of the *system's error* in the training set with respect to the *system's error* in the test set.

Then, for comparison purposes, two considerations are to be made. If the *system's error* estimated in the training set differs greatly from the *system's error* estimated in the test set it can be concluded that the variance of the system is too high, and thus the generalization of the *virtual LT sensor* is not good. On the contrary, if both errors are similar, the variance term is considered small and thus the observed error and real error may be considered similar.

Therefore, given that the robustness of the methodology has been maximized, this training-test set performance comparison places our final model in the bias-variance continuum. These conclusions are fed to the next section where the interpretation of the final performance is made to decide what direction to follow next.

3.3.3 Next step: Criteria for direction adjustment

After a design iteration is finished, the next question that arises is: what do we do next? There are three possible answers to this question: (1) Accept the solution, (2) stop because continuing is non-viable or (3) make a direction adjustment and another design iteration. As illustrated in Figure 3.37, this section deals with the process that leads to one of these decisions.

The traditional next-step decision making process for a ML model starts from the evaluated performance. If the variance of the final model is non-negligible or doubtful, the *observed error* of the *virtual LT sensor* is not close to the real *system's error* and thus the solution cannot be properly evaluated and is insufficient. Therefore, the next step to be made comes from increasing the robustness which, depending on the characteristics of the learning approach, may be obtained by decreasing the complexity of the learning and/or increasing the sample size for testing purposes. The former may include multiple approaches such as: decrease the number of features, decrease the training algorithm complexity, increase the regularization term... On the contrary, if the variance is negligible, the magnitude of the bias error is compared with the *system's acceptable error* defined in section 3.1 (see Figure 3.7).

Then, if the bias error is below the *satisficing error*, the ML system is usually accepted in the traditional next step decision making processes (see Figure 3.36). In this work we go beyond that and introduce an additional methodological perspective (see Figure 3.37). As already stated in section 3.1, in problems such as the the LT where the heterogeneity of the population is considerable, the maximization of robustness is fundamental. Therefore, in line with what is represented in Figure 3.7 we pose and additional question: can be the robustness of the methodology and ML system further improved? If the answer is positive the next iteration is headed towards further increasing the robustness of the system. If the answer is negative, the solution is accepted.

On the contrary, if the bias is not below the *satisficing error*, further increased optimization may be sought. This is usually achieved both through an increased learning complexity and/or increasing the resources for modelling. To do so, the number of features may be increased, the training algorithm complexity increased or the regularization term decreased. However, it is important to note that any of these approaches also tend to require to increase the resources for modelling (e.g. increasing sample size and/or computational power).

To summarize, if the ML system is not accepted as final solution, one of two improvement directions are possible, increasing robustness or increasing optimization. Making the corresponding improvements are subjected to a cost-benefit trade-off which may give way to the next step or decide that is unfeasible. More precisely, the cost of taking another iteration is evaluated from the economical, temporal, technical, material... perspectives and compared to the closeness to achieving a sufficient solution and the comparative value that provides.

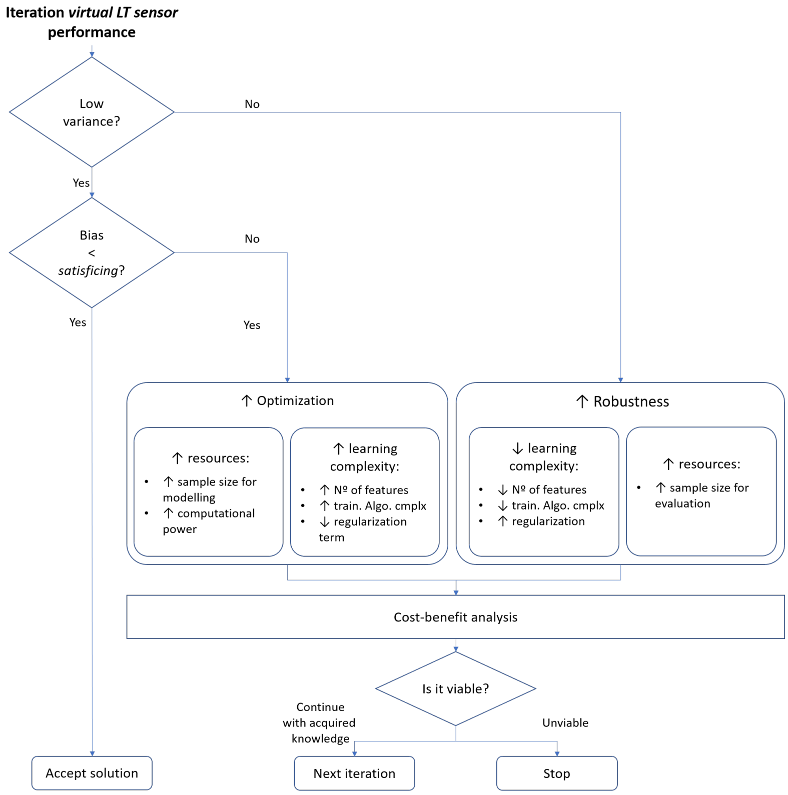


Figure 3.36: Traditional machine learning next step decision making process

Abbreviations: train, training; algo, algorithm; cmplx, complexity

3.4 Strategical and methodological conclusions

This chapter has served to, starting from the higher level of abstraction, define a set of strategical principles that, after combining with the common traits in supervised learning problems, grew into a detailed iterative methodology that is to be used in the design of the *virtual lactate threshold sensor*.

From a general analysis of the problem, the inherent difficulties of ML complex problems were identified as: (1) the problem boundary discovery and (2) defining the appropriate performance perspective. We proposed an iterative strategy to deal with the former and, for the latter, we set an approach to achieve a *satisficing* accuracy.

Then, with these strategical considerations in mind, a design methodology was developed. This methodology formalizes the common traits that are found in supervised learning and applies it to the *virtual LT sensor*, detailing the steps to be followed intra-iteration. More precisely, it is divided in three steps: context characterization, which deals with ensuring that the quality of the collected data is maximized; content representation, dealing with the approach for learning only the relevant information; and next step selection, which guides the decision making process for the next iteration. Here it is important to note that, despite this traditional next step decision making process is well known in practice [57], to the best of our knowledge, the formalization done in this work is a contribution. Moreover, in this work, we go beyond evaluating the final ML system and introduce an additional methodological perspective to the traditional next step decision making

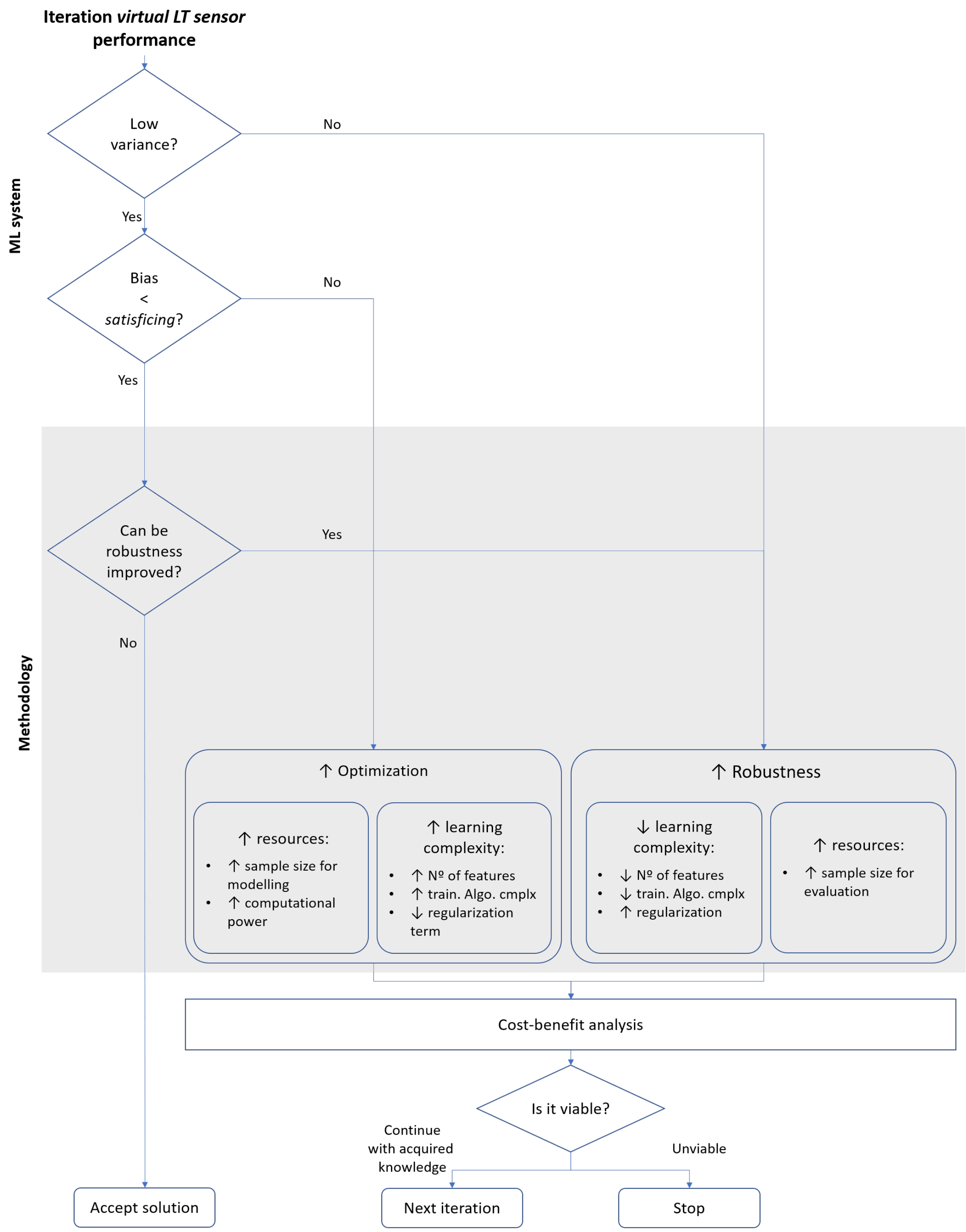


Figure 3.37: Novel machine learning next step decision making process

Abbreviations: train, training; algo, algorithm; cmplx, complexity process.

As a general conclusion, this chapter served to show the importance of evaluating the strength and weaknesses of the tools to be used (ML in this case), evaluate the characteristics of the systems we are working with (lactate metabolism) and include the application perspective towards a successful solution. Despite this perspective is not novel and is common practice in systems engineering, we brought this concepts to the field of ML where it was not done so far. Moreover, this lack of methodological rigour made its appearance even in prior works that tried to use ML for lactate estimation [54; 55].

Therefore, in this chapter we formalized a methodology and made as much as possible assumptions explicit so that we created a rigorous methodology that is easily reproducible, falsifiable, adjustable and transferable to other similar problems.

Next chapter will focus on designing the *virtual lactate threshold sensor* according to the methodology here defined.

Chapter 4

Design and development of the virtual lactate sensor

A designer knows he has achieved perfection not when there is nothing left to add, but when there is nothing left to take away - Antoine de Saint-Exupery

Figure 4.1: Walking the path towards the goal

This chapter presents the design and development of the *virtual lactate threshold sensor* based on the methodology defined in Chapter 3.

As illustrated in Figure 4.2, the methodology that was proposed in Chapter 3 unfolds into several specific steps that are to be followed in this section.

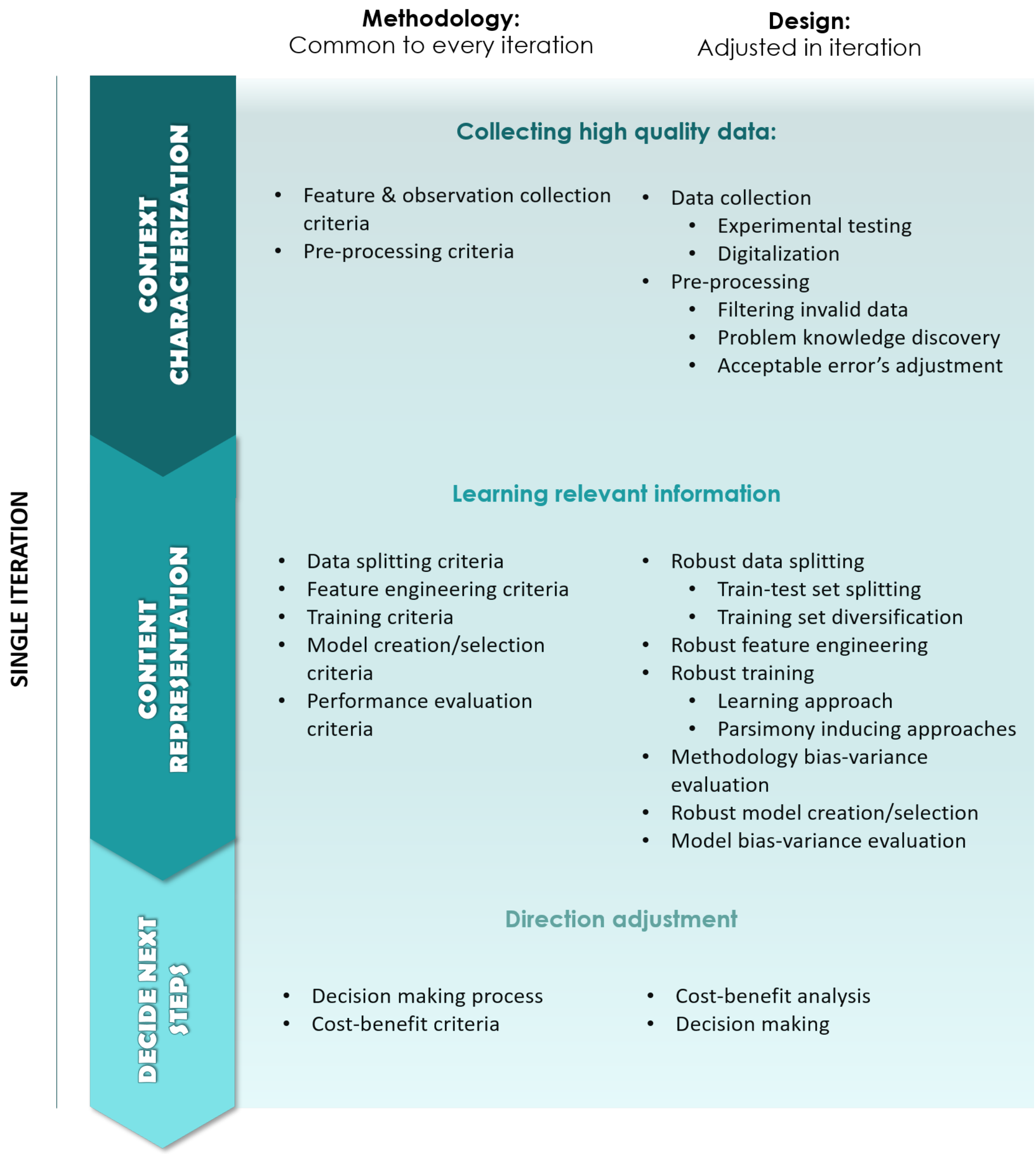


Figure 4.2: Design iteration: Unfolding of methodology at iteration level

More precisely, the context characterization, content representation and the decision of the next step contain the following steps:

- Context characterization starts with data collection, which is done according to the experimental methodology defined in Chapter 3 and afterwards formatted into a digital database. Then, the raw data is pre-processed in order to filter out the invalid data, get some knowledge about it and adjust the *satisficing* errors that are to be used as our *virtual LT sensor* performance metrics.

- Content representation starts by defining the ways in which the data is going to be split.

First, to set aside the testing set, and second, to define the re-sampling criteria that is to be used in the training set if any. It continues creating robust feature engineering and training approaches. The intrinsic bias-variance characteristics grow from these approaches which are now to be evaluated. Using this evaluation, a final model is created or selected using a robust approach. The last step concerns with the bias-variance evaluation of the final model.

- Decide next step: From the conclusions derived in the content representation, if further improvement is considered necessary, a cost-benefit analysis is done to decide which step to take next.

Altogether, and as shown in Figure 4.3, the design of the *virtual LT sensor* consists of two iterations.

The first iteration starts by identifying the LT estimation problem as a non-linear dynamic problem. Then, from the analysis of the state-of-the-art and expert knowledge of previous works, a Recurrent Neural Network (RNN) is proposed as approach to model the lactate curve. Despite the created *first iteration virtual LT sensor* (hereafter *initial virtual LT sensor*) shows a bias below the acceptable error, it is considered that further evaluation and robustification would be desirable. Consequently, a second iteration based on a heuristic as robust estimator is carried out which finally provides a robust *second iteration virtual LT sensor* (hereafter *calibrated virtual LT sensor*) for recreational runners.

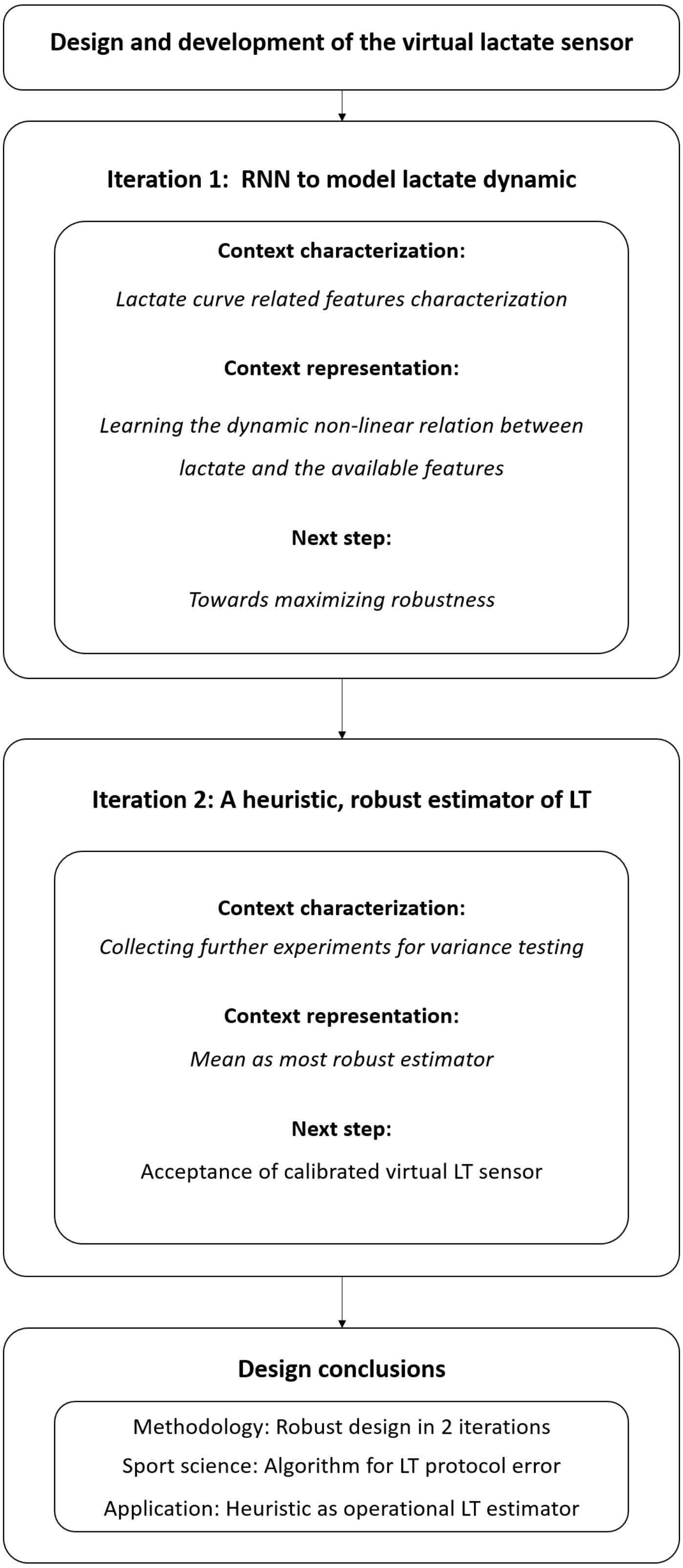


Figure 4.3: Overview of Chapter 4: Design iterations

4.1 Iteration 1: Recurrent neural networks to model lactate dynamic

This iteration entails the design of the first *virtual lactate threshold sensor*. To do so, first the relationships of the lactate and LT are analysed with respect to the available input features.

The complexity of this problem lies on the non-linear dynamic behaviour of the blood lactate appearance [79] and the multiple features involved. The individual Dmax LT is precisely an effort to characterize (part of) the dynamic behavior by focusing on its break-point. Hence, the two main characteristics of our problem can be summarized as: (1) the blood lactate concentration has a dynamic behaviour which is dependent not only on actual inputs but also on previous ones; and (2) the relation of the blood lactate concentration with the input features is non-linear.

Therefore, our first iteration starts from the hypothesis that the dynamic of the blood lactate concentration and the rest of the input features posses relevant information to be learned from. Usually the dynamic of the output feature itself (evolution of blood lactate concentration) intrinsically contains information that may not be found through other input features. This is very common in real time dynamic problems [80], specially when the output variable is dependent on multiple not easily measurable or unknown features as in our case.

Based on the literature review of Chapter 2 and of previous experiences [56], in this first iteration an ANN based approach is used to create the first *virtual LT sensor*. Figure 4.4 illustrates the structure of this iteration. The design of the *virtual LT sensor* is further divided into a context representation phase that collects and pre-processes a set of features followed by the content characterization that is created around the ANN learning based approach, plus the third *next step decision making* process.

More precisely, a RNN is used to learn the non-linear and dynamic behavior of blood lactate concentration and ties it with several easily measurable physiological features assessed during the tests.

4.1.1 Context characterization: Lactate curve related features characterization

This context characterization phase deals with the collection and pre-processing of the relevant data for its subsequent use in the content representation phase. In this first iteration, a total of 142 recreational runners participated in the data collection.

To participate in the experimental test first athletes must comply with the target population prerequisites. In these experimental tests, the features defined in Chapter 3 were collected according to the collection protocol detailed in annex B.

Once that the experiments were performed, first the experiments with sampling errors, i.e. those with major contingencies or non-target population (see Figure 3.14 in Chapter 3 section 3.2.3) were detected and discarded.

Then the remaining experiments were analysed in terms of their validity for acceptable error calculation (see Figure 3.15 in Chapter 3 section 3.2.3). In this regard, as incorrect lactate measurements and LT estimations were still valid for *satisficing error* calculations, all of them were used for this purpose. The acceptable error calculation shows that approximately the 4% of lactate

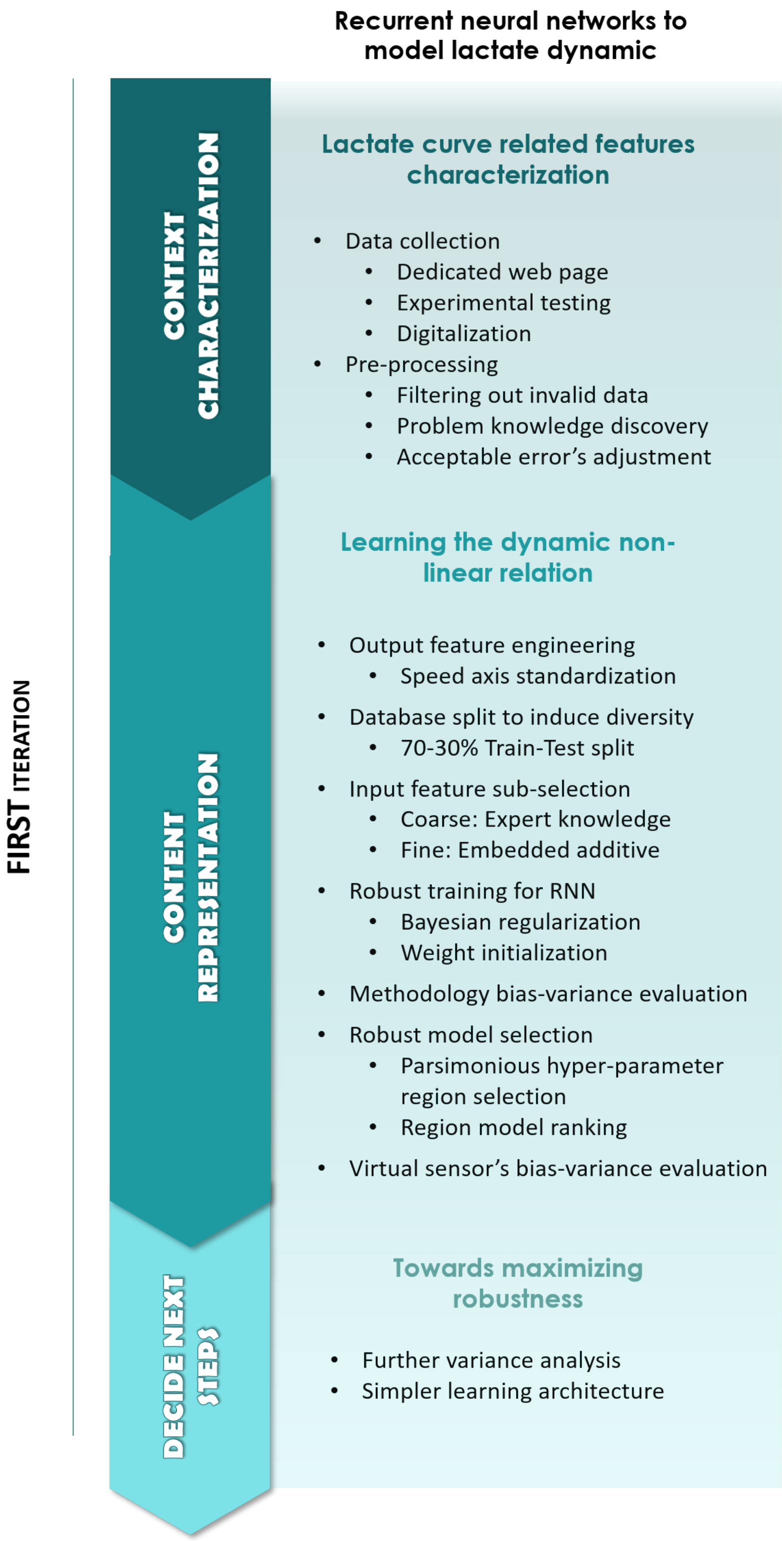


Figure 4.4: Structure of first iteration design steps

Table 4.1: Iteration 1, from collected experiments to valid ones

	Iteration 1
Selected for experiments	**142**
of which:	
No pre-requisite compliance	2
Performed experiments	**140**
of which:	
Sampling error: Major contingencies	3
Sampling error: Non-target-population	22
Valid for *acceptable error* calculation	**115**
of which:	
Data error: Incorrect lactate curves	4
Data error: Incorrect HR measurements	9
Valid for *designing*	**102**

Abbreviations: LT, lactate threshold

curves obtained with the Dmax protocol are flagrantly incorrect due to their lack of convexity and consequently, so are their LTs.

Regarding the validity for designing (see Figure 3.21 in Chapter 3 section 3.2.3), the rest of the potentially relevant features were analyzed in pursuit of flagrantly incorrect measurements or transformations that could invalid the entire experiment. After discarding them (most due to HR measurement errors), 102 test remain valid and are gathered in the final database to be used in this first iteration.

Table 4.1 makes a summary of the data collection and pre-processing and the following sections get into the details of it.

Additionally, the pre-processing is also used to get additional knowledge about the problem in hand. The following sections go into the details of these *context characterization* steps.

Data Collection

A dedicated web page was created to support part of the work to be done in this thesis [81]. One of the purposes of the web page, as represented in Figure 4.5, is to serve as platform to gather volunteers for the study.

To do so, the web page informs to the potential volunteers about these requisites defined in Chapter 3 section 3.2.2. These pre-requisites are here again stated for convenience:

- Endurance athletes training for and participating in races from 5 km upwards
- Currently running at least 3 days a week and competing in recreational endurance races
- A running experience of at least 1 year.

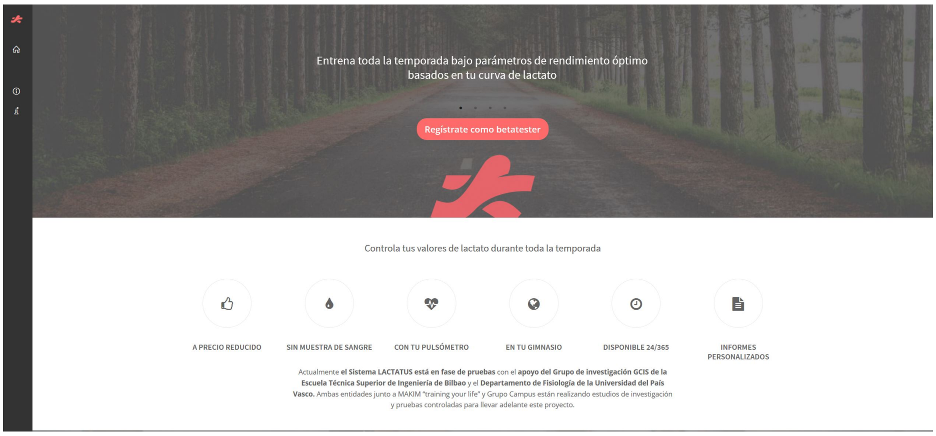


Figure 4.5: Lactatus web's registration page: Gathering volunteers

- Estimated athletic level according to the maximum running stage reached in the test herein assessed between 14.5-20.5 kilometers/hour.

Additionally, the volunteers were also informed about the additional requisites that are necessary to be able to perform the test. These requisites are here repeated for convenience. The athletic-health requisites that each athlete must fulfil are:

- Be well rested and to abstain from hard training sessions and competition for 24 hours before testing.
- Abstain from eating for 3 hours before testing.
- Abstain from taking stimulant substances before testing, including coffee or tea.
- Be familiarized with running on a treadmill.
- Being healthy and lacking of infections.

The safety and legal requisites that must also be fulfilled are:

- To be federated in their respective disciplines.
- Provide a medical certificate that ensures that they are able to perform the test.

The ethical requisites that must also be fulfilled are:

- The participant has read the information sheet (see annex A).
- Provide written informed consent acknowledging that they have been informed about the possible risks of the tests and giving their consent (see annex A).

Using the web platform, and under short notice, 803 local athletes volunteered for the study. This response again confirms the high demand that estimation of LT arises among this population. In this first iteration, the amount of observations to be collected was decided by maximizing the number of experiments that could be performed with the economical and technical resources available at the moment. In particular, as represented in table 4.1, from the list of 803 athlete's that volunteered, 142 athletes were randomly sampled and selected to participate in the experiments.

From the 142 sampled athletes, two were rejected due to not fulfilling the health requisites, while the rest 140 athletes were considered able to perform the tests.

As explained in Chapter 3, the features, together with any issue observed during the tests, were collected in both paper and digital format. HR evolution is an exception since it is collected by the HR monitor in its dedicated software (Garmin Connect, George Town, Cayman Islands). After-cross checking the three formats (paper, digital and HR monitor software), the final results were gathered in a digital database in excel format (Microsoft, Redmond, Washington, USA).

Pre-processing

As already mentioned in Chapter 3, the pre-processing serves three main purposes, filtering out invalid data, knowledge discovery and adjusting the *satisficing* errors. In this section, the pre-processing results are described.

Filtering out invalid data

As defined in Chapter 3 section 3.3 and represented in table 4.1, the filtering is done in three steps.

First of all, the sampling errors were analysed and the incorrect experiments discarded. Major contingencies detected during the tests are among these errors. These major issues may include unfinished tests, flagrantly sub-maximal test... From the 140 athletes that performed the tests, three athletes had a major problem during the test. Additionally, although the characteristics for being considered target recreational runner population are requisites for volunteering to the study, a cross-check was done after performing the tests. This allows to detect and remove any non-target-population athlete that participated. For instance, the requisite of being able to reach to a maximum step between 14.5-20.5 kilometers/hour is one of the most problematic characteristics since, without making the test, only a rough prior estimation about their performance is available. After the analysis, among the remaining 137 athletes 22 are found not to fulfil the target recreational runner population characteristics. Therefore, after cleaning the invalid data coming from sampling errors, prior to looking for incorrect data, 115 remain valid for both *acceptable error* and *virtual LT sensor* designing purposes.

Once that the invalid data coming from sampling errors has been filtered, the focus is placed in the sources of error from the data perspective. In this regard, despite the great efforts put into the data collection protocol, the possibility of gathering incorrect data is always present.

More precisely, the analysis from the data perspective (see Figure 3.14) leads to the second step of filtering. This step deals with detecting, classifying and reserving the data that is valid for *acceptable error* adjustment. As previously mentioned, both the incorrect lactate measurements and curves are valid for this purpose and thus are reserved. Figure 4.6 represents all the 115 lactate curves and show that 4 of them are flagrantly incorrect. All these lactate curves and LT are reserved for the *system's acceptable error* adjustment.

The third and last pre-processing step deals with detecting and separating the experiments that are valid for *virtual LT sensor* designing purposes. In this regard, apart from the 4 experiments

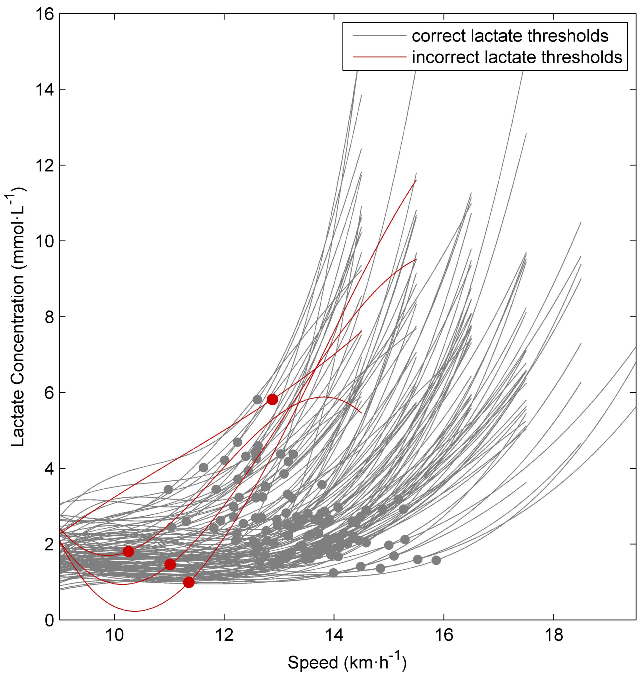


Figure 4.6: Iteration 1 incorrect lactate curves

with incorrect LT curves, the input features are also scrutinized for incorrect data. HR is among the most important input features to be used for designing purposes and also among the most probable to create incorrect measurements due to its acquisition procedure. More precisely, the HR values are measured through a HR monitoring band that is attached to the athlete chest with, due to the movements created while running, it may lose contact creating missing and/or incorrect data. In this regard, the one by one analysis among the remaining 111 athletes 9 are detected with flagrantly incorrect HR data. Therefore, 102 remained valid for designing purposes. Table 4.2 summarizes the performed uni-variate analysis presenting the mean and standard deviations of the remaining numerical features. The high Coefficients of Variation of "IAAF scores" and "training years" are an example of how heterogeneous this sampled population is it was expected to be according to the population we targeted.

Therefore, the *initial virtual LT sensor* is to be created from a database of 102 athletes.

Problem's knowledge discovery

Apart from the already explained uni-variate analysis which gives descriptive information about our sample, a bi-variate statistical analysis is also done [82]. This analysis tries to look to the interrelations between features, expecting to be of value for feature engineering purposes or to discover new knowledge that could be useful. Figure 4.7 illustrates the interrelations between all the features. For the sake of clarity, only strong correlations above R = 0.75 are illustrated. On a brown to yellow range, the closer to yellow the higher the correlation value.

In this graph several clusters of features are observed. As expected, % of maximum HR values, HR related features, anthropometric features, Lactate related features and HRR related features

Table 4.2: Iteration 1 numerical features mean,standard deviations and coefficient of variation

Feature	Mean ± SD	CoV	sample size
Lactate threshold [km/h]	13.4 ± 0.9	7.0	102
Personal best [IAAF points]	334.4 ± 224.2	67.1	50
Vpeak [km/h]	16.6 ± 1.3	7.8	102
Years train [years]	5.5 ± 5.8	105.2	95
Age [years]	36.2 ± 7	19.4	102
Height [cm]	175.2 ± 7.0	3.6	102
Weight [cm]	72.3 ± 8.8	12.2	102
Body mass index [kg/m^2]	23.5 ± 2.1	8.9	102
Abdominal diameter [cm]	80.8 ± 6.4	7.9	102
Hip diameter [cm]	92.2 ± 5.1	5.5	102
Body fat percentage [%]	15.2 ± 4.2	27.9	102
Water percentage [%]	62.0 ± 4.1	6.6	102
Resting HR [bpm]	79.0 ± 15.7	19.9	102
Maximum HR [bpm]	183.6 ± 11.5	6.3	102
%HRmax at 13.5 km/h [%]	89.0 ± 5.2	5.8	102
%HRmax at 14.5 km/h [%]	92.7 ± 5.0	5.4	102
%HRmax at 15.5 km/h [%]	95.1 ± 4.0	4.2	79
%HRmax at 16.5 km/h [%]	97.0 ± 3.2	3.3	58
Heart rate deflection point [km/h]	13.0 ± 2.1	16.2	79
HRR threshold [km/h]	13.8 ± 1.5	10.9	93
Resting Lactate [mmol/l]	1.3 ± 0.2	17.5	102
Maximum Lactate [mmol/l]	9.4 ± 2.7	28.9	102
Lactate value at 13.5 km/h [mmol/l]	3.2 ± 1.8	55.7	102
Lactate value at 14.5 km/h [mmol/l]	4.8 ± 3.2	65.8	102
Lactate value at 15.5 km/h [mmol/l]	5.3 ± 2.7	50.5	79
Lactate value at 16.5 km/h [mmol/l]	6.2 ± 2.4	38.6	58
Maximum muscular Borg [Borg scale]	7.8 ± 2.1	26.5	101
Maximum respiratory Borg [Borg scale]	9.1 ± 1.5	17.1	102

Abbreviations: LT, lactate threshold; CoV, Coefficient of Variation

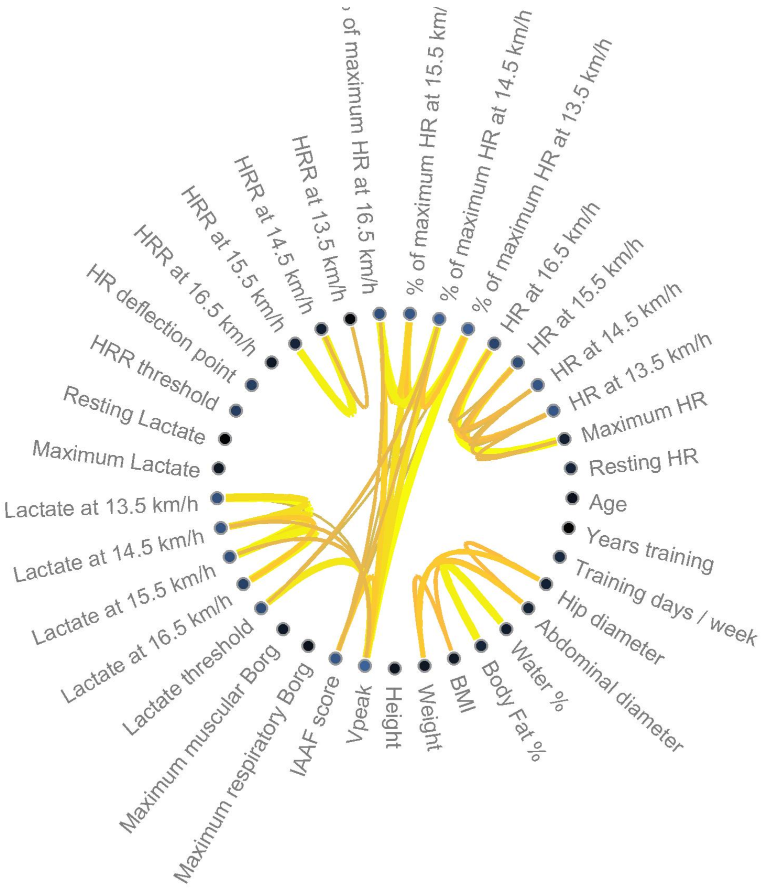


Figure 4.7: Iteration 1 network correlation analysis

form five clear clusters of features. Among them, HR related features, anthropometric features and HRR related features are seen as independent in terms of any strong relevance (R > 0.75). Lactate related features on their part appear somehow related to the performance related features (IAAF score and Vpeak) which is consistent with what is known in the literature [83; 84]. Finally, the strongest inter-cluster relations appear between the %HR max related, lactate threshold and performance related features, showing that it may be an interesting connection between them.

Acceptable errors' adjustment

As already mentioned in Chapter 3, there are two acceptable errors to be adjusted, *individual acceptable error* and *system's acceptable error rate*. For the *individual acceptable error*, 111 LT points are available (115 valid for *acceptable error* calculation - 4 Incorrect lactate curves). Using them, algorithm 1 is computed. In our particular case, different "W" numbers of bootstrap re-samples (10, 20 and 100) have been used on the 111 athletes. Different "Z" number of random samples (10, 20 and 100) have also been used for each of the blood lactate measurements. In both cases the results from 100 random do not significantly differ from those obtained with 20, so a higher number of random samples is not considered necessary. The results are represented in Figure 4.8 for a given athlete.

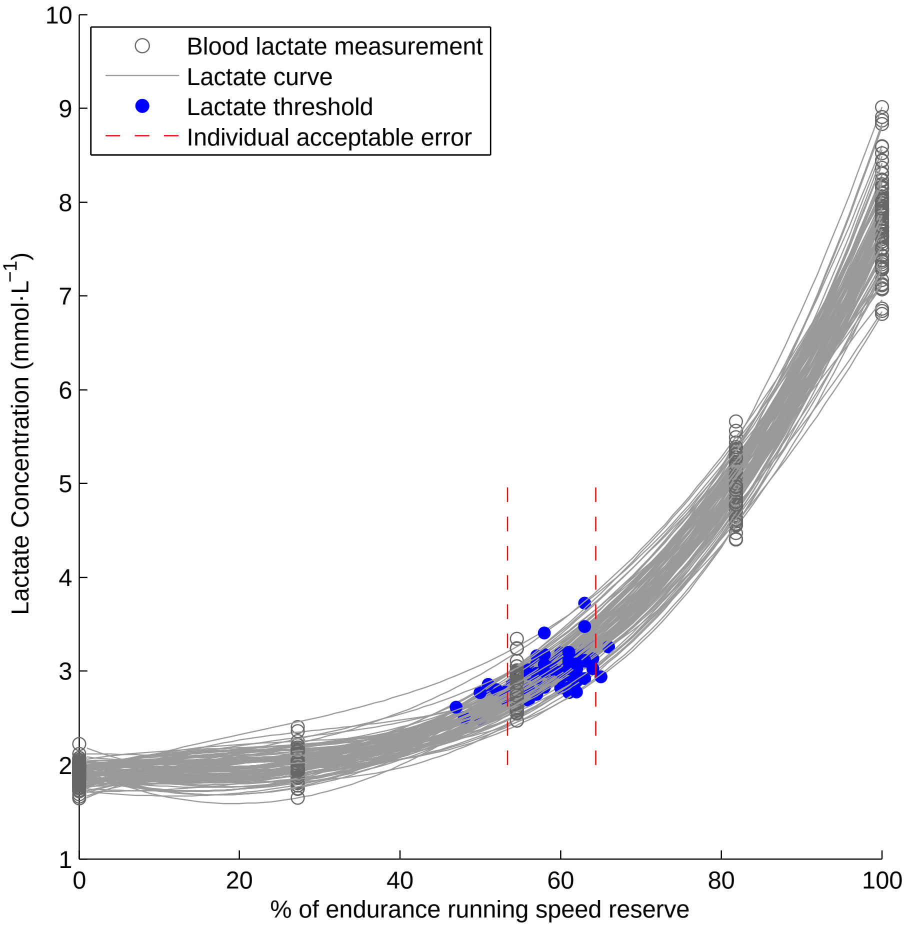


Figure 4.8: Dmax lactate threshold variation for a given athlete caused by blood lactate measurement precision error

From the computation of algorithm 1, we have estimated the standard deviations of the *Dmax LT method precision error* and illustrated in table 4.3, classified by the ”Y” number of blood lactate points taken during the treadmill speed test (Lactate Points column, table 4.3). These results show how the *Dmax LT method precision error* improves with the number of blood lactate measurements taken from the athletes, which is consistent with what is known in the literature and practice [34].

To further analyze how the *Dmax LT method precision error* can affect our estimator, we compare it to the *individual acceptable error* defined in table 3.1 in Chapter 3 in section 3.2.3 and represent the residual errors in Figure 4.9. In this Figure, we observe that the *Dmax LT method precision error* can significantly affect the final error, showing that the 98.8% of the plausible Dmax LTs are within the range of the *individual acceptable error*. In other words, from this analysis the *virtual LT sensor* could at best achieve a *ceiling accuracy* of 98.8%.

Therefore, as represented in Figure 4.9, the *LT method precision* errors derived from the physiology equipment and methodology is sometimes even higher than the *individual acceptable error* determined by expert knowledge. This fact reinforces that this *satisficing* threshold is a safe and robust metric to determine the validity of our *initial virtual LT sensor*.

Regarding the *system's acceptable error*, using the *LT method precision* we derived that there

Table 4.3: *Dmax lactate threshold method precision error* according to number of lactate points

Y Lactate points	*Dmax LT method precision error* real LT (SD)
5	8.1
6	5.6
7	6.2
8	5.3
9	3.4
10	2.5

Abbreviations: LT, Lactate threshold ; SD, standard deviation; It, iteration

is at least a ceiling accuracy of 98.8%. Moreover, as previously shown, there are 4 incorrect lactate curves among the total measured 115. This means that we can consider that the Dmax protocol fails for 5 % of the athletes (4% due to the incorrect LT plus 1% due to the *Dmax LT method precision error*) which gives an additional perspective of how strict we can be for our *virtual LT sensor*. Therefore, it seem clear that a *system's acceptable error rate* 90−95% *satifices* the objectives of our *initial virtual LT sensor*.

4.1.2 Content representation: Learning the dynamic non-linear relation between lactate and the available features

The evolution of blood lactate concentration is a dynamic problem where the output variable is dependent on itself and on multiple not easily measurable or unknown features. As already explained, this first iteration starts from the hypothesis that the dynamic of the blood lactate concentration and the related input features possess the relevant information needed to create a *virtual LT sensor*.

In this regard, dynamic modelling is a very active area that deals with this kind of problems. Particularly, future event prediction based on past information or time-series forecasting is a dynamic modelling field that has been highly researched [85]. However, as shown in Figure 4.10, dynamic modelling involves many other approaches different than time-series forecasting.

Among these other dynamic approaches, time-series forecasting with exogenous features is an extension of the well-known time-series forecasting area. However, there is another less known research area that focus on modelling the dynamic behavior of a system in certain specific conditions in which there are no real current and/or past values for forecasting the future ones. As represented in the third row of Figure 4.10, this means that the time-series of interest has to be completely estimated from the input features. Following the classification of Figure 4.10, the lactate dynamic modelling is of the third kind, as it tries to estimate a complete time series from exogenous features.

With regards to the solutions available to dynamic problems, the review done by Makridakis et. al. [86] showed that, for time-series forecasting, simple statistical models obtain better out-

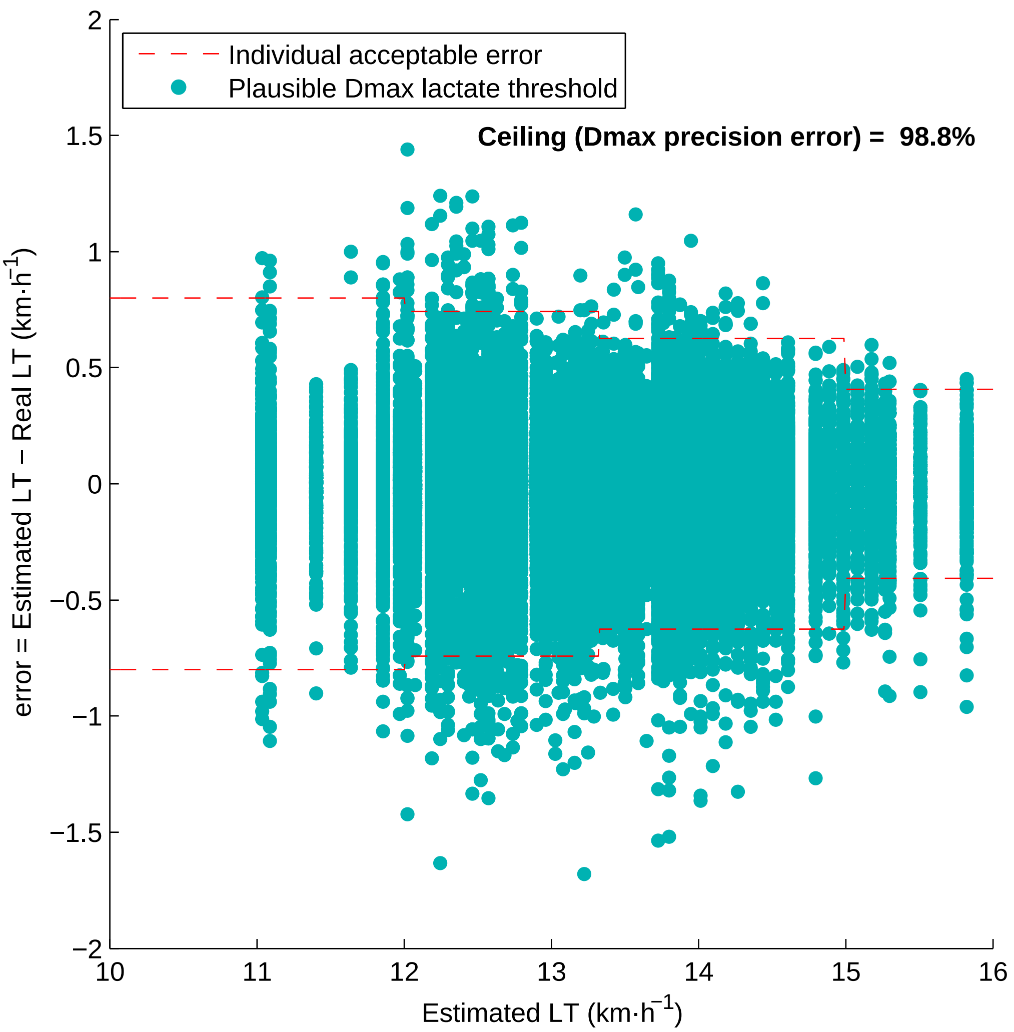


Figure 4.9: Individual acceptable errors created by initial *Dmax LT method precision error*

of-sample accuracy than more complex ML methods. However, Makridakis et. al. also stated in their work [86] that the findings may be different if nonlinear components are present, or if the data is being dominated by other factors. In such cases, they suggested that the highly flexible ML methods could offer significant advantage over statistical ones. Among the multiple techniques available, ANN presents as a good candidate compared to other statistical approaches to solve the lactate problem, since they have the ability to reproduce complex non-linear processes.

Moreover, ML has already been used to address this kind of problems. For instance, in [80] a RNN was used to model dynamic non-linear systems of a gas turbine for simulation of its start-up operation. A NARX model without current time step data was used in this case and the final models were validated against other three experimental data sets. The lack of required initial output values make Elman based RNN also suitable to the characteristics of estimating complete lactate curves. In this regard, Arriandiaga et. al. [56], used a Elman based RNN to estimate complete time series in a specific time interval from multiple and distinct input time-series.

Therefore, based on the literature and the knowledge of previous experiences, selecting RNNs to create a *initial virtual LT sensor* shows the characteristics of a robust first approach. Thus, the

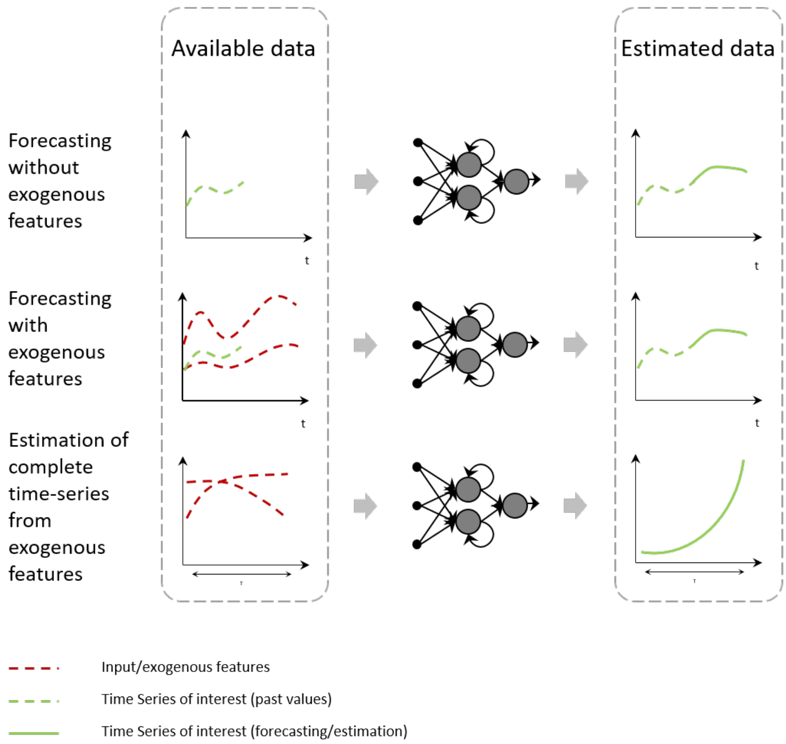


Figure 4.10: Different dynamic modelling approaches

methodology here proposed models the dynamic of the blood lactate appearance in the aforementioned incremental treadmill speed test conditions using a RNN.

As already mentioned in Chapter 3, there are multiple decisions that are to be taken in parallel in the context representation phase about training set split, feature engineering, training... Therefore, the selection of RNN as the training algorithm sets the first stone that influences the rest of the decisions that are to be made to build a robust context representation step according to the methodology explained in Chapter 3. Figure 4.11 gives an overview of these decisions by placing them in their corresponding context representation step.

Robust output feature engineering: standardization of lactate curves x-axis

An important decision that affects the consequent context representation steps is the format in which the output feature, i.e. the LT is to be modeled. In our case, the interest of the LT resides in its x-axis component, i.e. in the dimension related to the LT exercise intensity (speed in this case).

Related to it, as represented in Figure 4.12, the raw lactate curves have different lengths depending on the peak running speed that the athlete obtained during the incremental treadmill test (hereafter Vpeak). As more fitted athletes can maintain higher running speeds, the duration of each test depends on the individual fitness level of each athlete. As shown in Figure 4.12, as recreational athletes have very diverse levels, the tests are also very diverse in length and the LT intensities vary highly.

If these raw time-series were used to train the RNN it is presumable that the longer ones would have more relevance in the learning process due to their greater number of data points and

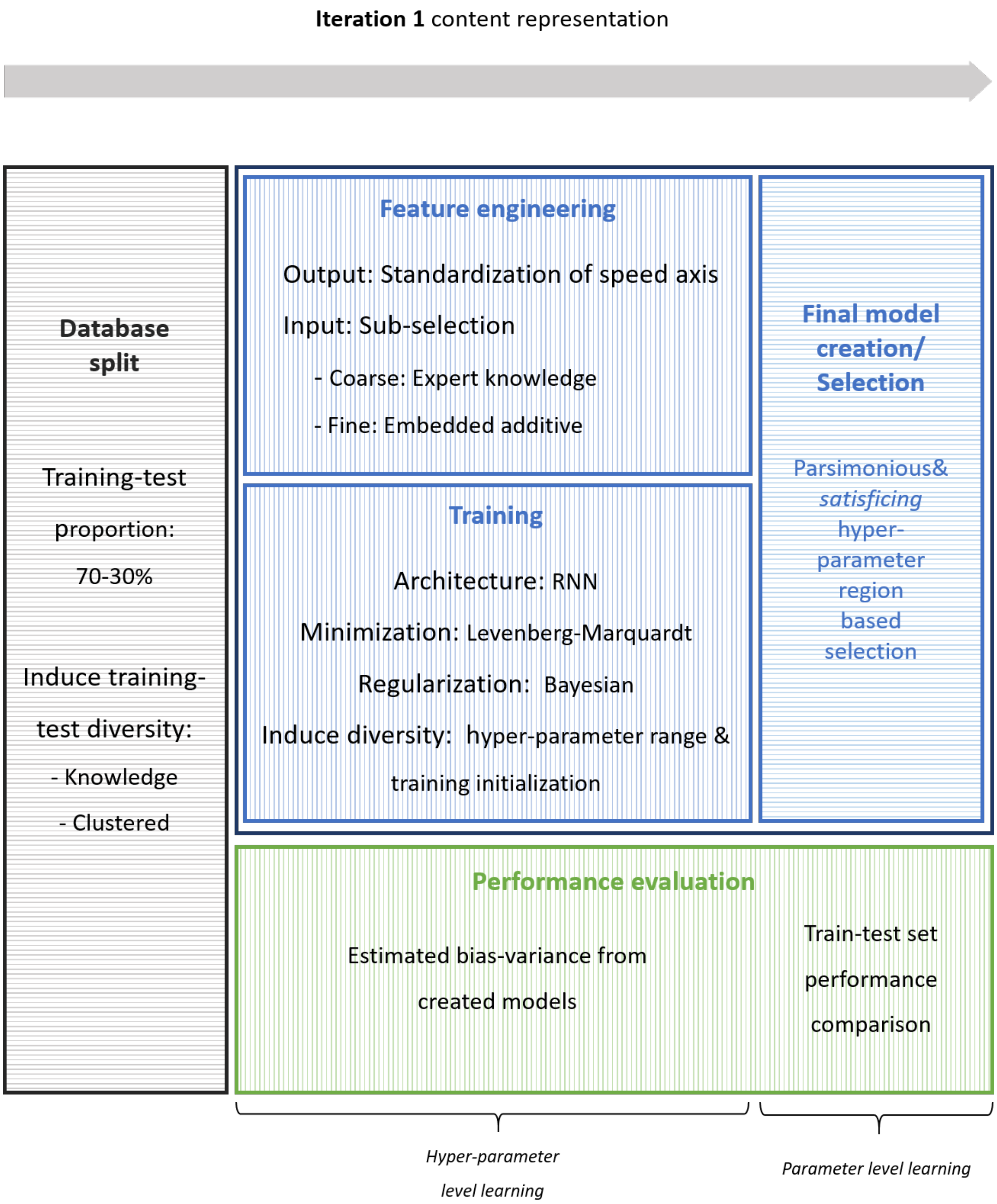


Figure 4.11: Iteration 1 content representation for dynamic modelling

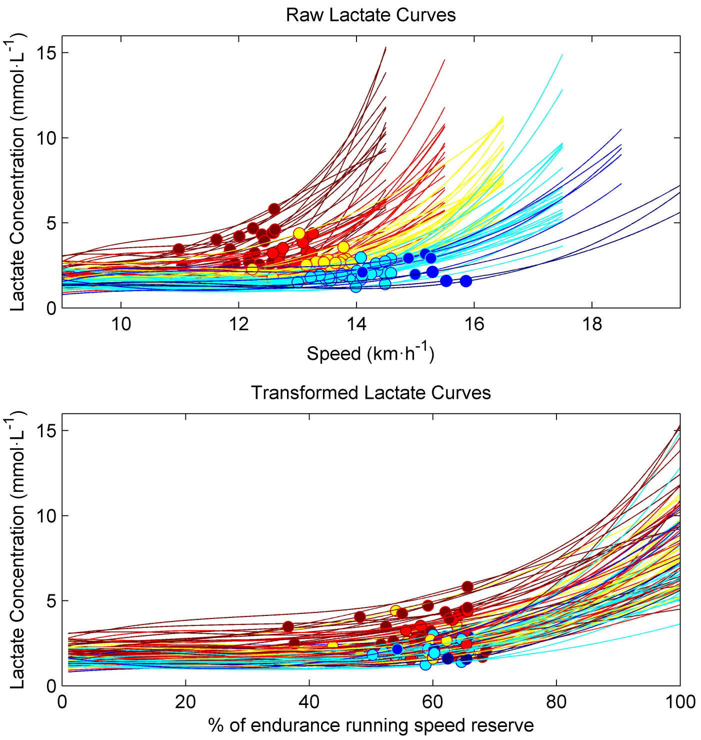


Figure 4.12: Raw & standardized lactate curves

Abbreviations: LSF, last step finished

impacting on the validity of the model for athletes with lower Vpeak. Moreover, the individual lactate threshold is not related with an absolute exercise intensity but with the tipping point of the curve.

Therefore, a standardization of the x-axis of the lactate curves is proposed to mitigate this problem. More precisely, the two main purposes of standardizing the lactate curves are to (1) unify the lengths of all time-series so they have equal relevance in the learning process and (2) to concentrate the lactate threshold of all athletes in the same region so that the learning process is simplified. This way, as the running speed is directly related with the exercise intensity, each test features are standardized with respect to the maximum intensity of each athlete. In this work we define the difference between the Vpeak and the initial running speed of the maximum incremental running test as endurance running speed reserve (6).

endurance running speed reserve [km/h] = Vpeak [km/h] - initial running speed [km/h] (6)

Then, as shown in Figure 4.12, the lactate curves are represented in percentage of each individual endurance speed reserve, in order to group athletes with different fitness levels and treat them as a single group. This approach is useful when, as in our case, the dynamic of the system is much more relevant for the final result than the absolute values related to it. What is more, once the LT is dis-attached from the effect of the Vpeak, all the lactate curves have equal relevance; the variability of the LT is highly reduced and the problem is simplified.

Database split: Train-test split and inducing diversity

As already mentioned, a proper database splitting is fundamental to induce diversity into the modelling process and be able to create a robust ML system with a minimized *generalization error*. Moreover, there are different levels and parameters for which different splits or criteria can be followed.

Training and test data set separation enables to leave part of the data unseen for final model performance evaluation. In this regard, two considerations are to be made: (1) how much is left for testing purposes and (2) how to ensure that it is properly split.

Regarding the proportion of the data splitting, as represented in Figure 4.13, in this iteration a 70-30% split has been selected for the training and testing sets respectively, a common criteria in ML [76].

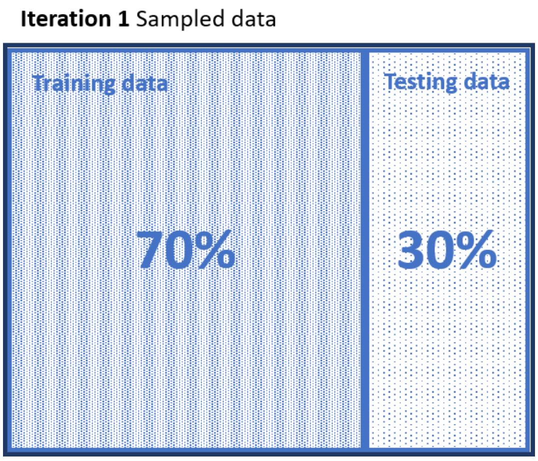


Figure 4.13: Iteration 1 training and testing set split proportions

Regarding the way of splitting, usually, the training and test examples selection is done by pure random sampling. However, this approach is suitable only if the available database is big enough so that random selection maximizes the diversity in both data sets. In other words, the aim is to make the database split so the target population's (i.e. recreational runners) diversity is characterized in both sets. Since in the present work the sample is not huge, two different training and test set selection methods have been followed and compared in this pursue.

First, expert knowledge is used to make the separation so that the diversity of lactate curves is maintained in both sets. In this approach, the selection of the train and test examples has been made based on the knowledge of the physiology experts. The training and test examples have been selected taking into account the diversity of lactate curve shapes so that all type of shapes were present both in the training and test sets. In this regard, several characteristics of the curves such as the maximum/minimum lactate values, last step reached, decreasing/increasing patterns, non-exponential shapes and other characteristics have been considered. Figure 4.14, shows the expert knowledge based training test set split.

The other splitting approach proposed in the present work is a variation of a stratified random sampling. In this method, the whole population is classified into mutually exclusive and more homogeneous groups called strata. Then, a simple random sampling is made from each stratum

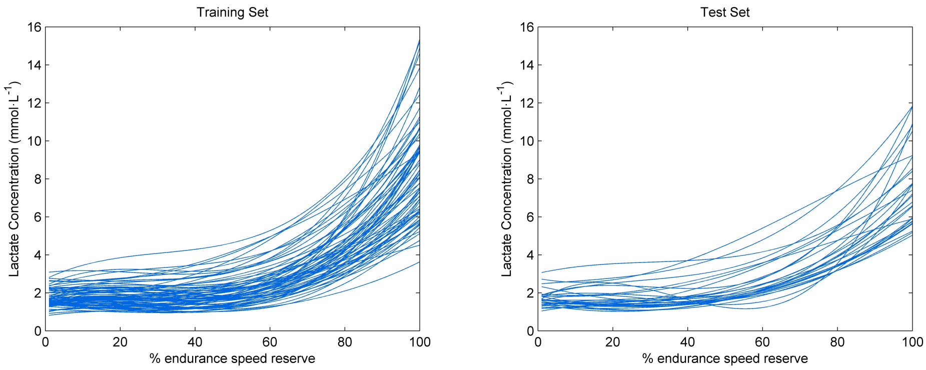


Figure 4.14: Knowledge based sampled data

so that a heterogeneous sample is created containing examples of all the sub-populations. In our case, this method is used to create training and test set splits representative of the heterogeneity of the complete database.

Usually, the classification of the whole population in several stratum is made according to one or more static parameters. However, in our case the parameter which is considered most

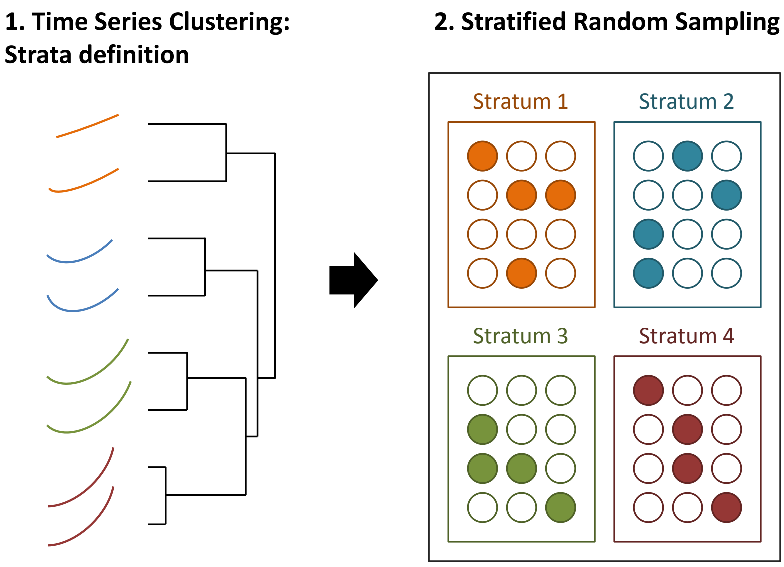


Figure 4.15: Esquema of modified stratified random sampling process

In this case, a hierarchical time-series clustering algorithm is used to create the strata as it allows not to pre-select the number of clusters and is independent on the form of the clusters. More precisely, using the hierarchical clustering technique, 10 different sub-populations are found as the strata. Then, the 30% (from the 70-30% split) of the examples of each stratum are randomly selected and included in the test set. The remaining examples correspond to the training set. Figure 4.16, shows the training and test set selection using the modified stratified sampling method.

In Chapter 3 we state the possibility of that further randomization can be made with the training set for hyper-parameter level learning. However, instead of following an approach that requires

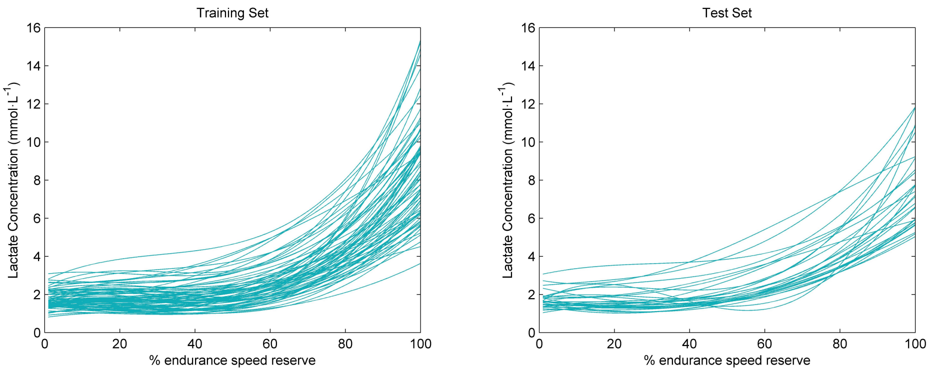


Figure 4.16: Modified stratified sampled data

from further training set splitting such as early stopping, strong randomization and regularization methods are to be used in the training step for *generalization error* reduction [87].

Robust input feature engineering: sub-selection from candidate input features

Regarding the input feature engineering, as already mentioned, after using expert knowledge to select a set of candidate input features for training, an additional sub-selection is done.

The purpose of this sub-selection is to minimize the amount of features to be used so that the robustness of the learning methodology is maximized. This sub-selection can be further divided into two steps: a coarse sub-selection of features using the knowledge of experts in physiology and a fine tuning of the features using an embedded additive selection.

For coarse feature selection, inputs which contain inter-individual and highly relevant information of the athletes are considered among the collected candidate features. Among them, features such as HR-derived ones, RPE, and age appear to be interesting. More precisely, it is known that HR-derived parameters and its evolution is related with lactate values and threshold [52; 48]. In addition, it is also known that the evolution of the RPE expressed as Borg scale [51] is related with the evolution of the lactate production.

The following list ranks in decreasing order the input features according to the potential relevance according to expert knowledge.

1. HR evolution
2. HRR evolution
3. RPE evolution
4. %HR max 14.5
5. Age
6. Gender

The final features are then to be selected from this list using the an additive strategy. The additive strategy works as follows: the simplest option (i.e. zero input features) is used to create RNN models. Then, the process is repeated adding the next most relevant input feature and creating another set of RNN models. Then the performance of both approaches is measured and compared and the best one selected. This procedure is then repeated until the addition of a new input feature no longer improves the quality of the previous approach. The reasoning of using an additive selection lies on that, intrinsically, this approach allows to go from most parsimonious (less features - lower model complexity) to more complex (more features - higher model complexity) approach, so that the final solution is always the most parsimonious that *satisfices*. This strategy falls into the category embedded methods, which means that the final selection of the features is done after the performance of the models is evaluated.

Robust training based on Bayesian regularization and weight initialization

As already mentioned, an Elman based RNN is selected as the learning architecture of this first iteration. The main characteristic of this type of network is the feedback from the hidden layer to the input layer, which is essential for problems where previous real values are not available as in our case. More precisely, as represented in Figure 4.17, a layer-recurrent neural network (LRNN) architecture has been selected. The LRNN is an Elman-inspired recurrent neural network which has flexibility to configure the number of hidden layers and the transfer function of each layer [88].

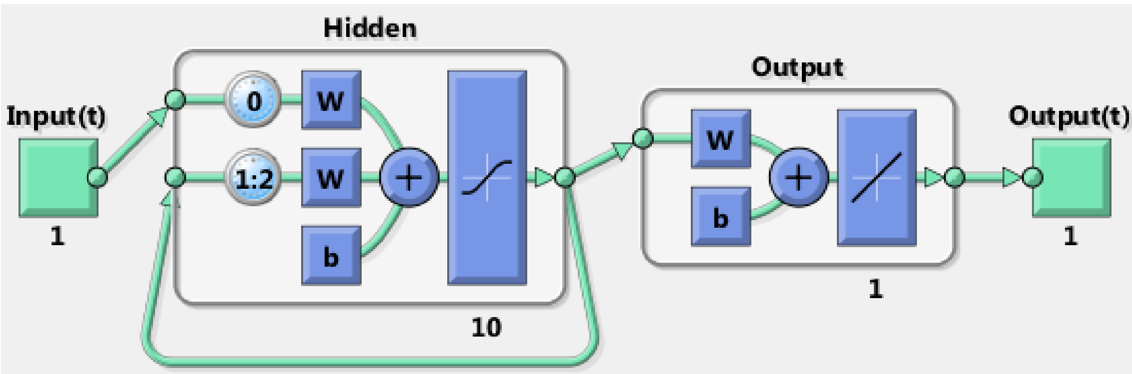


Figure 4.17: Layer-recurrent neural network

Reprinted from [89]

This kind of architecture requires from several configuration decisions that would define how the learning is done. The minimization function to be used is one of them, and in this case, the Levenberg-Marquardt minimization function is used to fit the model parameters as it has a faster convergence and lower error rate than other widely used minimization algorithms [90; 56].

Additionally, there are a couple of training configurations to be done to ensure that the training is robust. As explained in Chapter 3, early stopping and regularization are two of the ways in which generalization can be obtained. In this iteration a regularization approach is used which, contrary to a early stopping method, does not require from dividing the database in training, validation and test sets. More precisely, Bayesian regularization is used which may help achieve higher generalization capabilities than early stopping [87; 91].

As already mentioned, regardless the technique used, increasing the diversity of the multiple created models is another way of improving the robustness of the learning methodology. As represented in Figure 4.18, this diversity is to be achieved at two levels, hyper-parameter and parameter levels. The diversity introduced in the models that are created for each hyper-parameter configuration allows to evaluate the methodology bias-variance at hyper-parameter level.

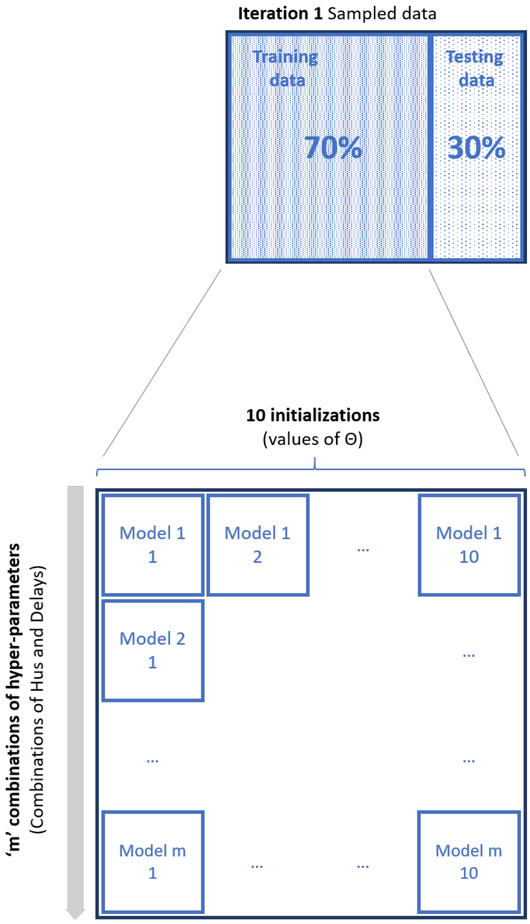


Figure 4.18: Iteration 1 training for hyper-parameter and model parameter diversity

The diversity created for each set of hyper-parameter set is usually done supported by a prior training data set split. In this case, the function minimum found by the training algorithm depends on the initial weight values of the model. Therefore, in the present work, each neural network hyper-parameter set is trained ten times with different weight initialization so that we can ensure that the performance evaluated from the set of models appropriately reproduces the bias-variance corresponding to the hyper-parameter set. In this case, the algorithm proposed by Nguyen and Widrow [92; 56] has been applied because it reduces training time over other weight initialization methods such as the layer-by-layer or purely random initialization.

Regarding the range of hyper-parameters to be used, in our LRNN case, the model complexity and structure is configured with two types of them: hidden units (HU) and delays. In order to minimize the training process time, it is desirable to limit the range of hyper-parameters while ensuring that the final hyper-parameter set (i.e. the final HUs and Delay combination) is inside that range. To do so, a preliminary training analysis is performed using a small portion of the database, so that a preliminary range of hyper-parameters can be selected at no high computational cost. This range is then to be used in the actual training. This preliminary training analysis has three steps:

1. Coarse tuning: Several trainings are performed with a wide range of hyper-parameter configurations trying to cover a big range of the bias-variance spectrum. The performance of these first models is calculated.

2. Increased resolution on operation point: The configuration parameter ranges are reduced and the resolution increased in order to focus in the zone where the best results have been obtained in the first step. Several models are trained according to this approach and their performances calculated. If possible, this step is repeated and the range is further reduced according to the results obtained in the second step.

Table 4.4 shows the results of each of the steps of the preliminary training. These preliminary training included data of 14 athletes that completed the 17,5 km/h stage. These athletes were selected as they are medium level recreational athletes which presumably have characteristics of the high and low level athletes.

Table 4.4: Training algorithm configuration parameters range definition using preliminary trainings

Step	Hidden Units	Delay
1	[1, 5, 10]	[1, 3, 5, 8, 10]
2	[1, 2, 3, 4]	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
3	[1, 2, 3, 4]	[5, 6, 7, 8, 9, 10, 11]

Therefore, the ranges of hyper-parameters to be used are, 1 to 4 HUs and 5-11 Delays, with 10 models for each combination of hyper-parameters for a total of 280 models.

In addition, as illustrated in Figure 4.19, the training process is repeated for each input feature combinations (i.e. 280 models for each combination) using previously mentioned constructive criteria. More precisely, the set of features trained are 1) none; 2) HR evolution; HR evolution & HRR evolution; and 3) HR evolution, HRR evolution & RPE.

Finally, as illustrated in Figure 4.20, the training process is performed for both previously established splitting criteria: knowledge based and clustering based.

The training is done using Matlab R2013b software Neural Network and Statistics and Machine Learning Toolboxes (MathWorks, Natick, Massachusetts, USA).

Methodology bias-variance evaluation

As stated in Chapter 3 and represented in Figure 3.7, the variance of the methodology is to be estimated to select the most robust hyper-parameter set. However, prior to selecting the final hyper-parameters, we first evaluate the most robust input features and data split combinations according to their overall estimated variance (see Figure 4.21). As explained in Chapter 3 section 3.3, the variance is estimated analysing the difference of the training and test errors of all the

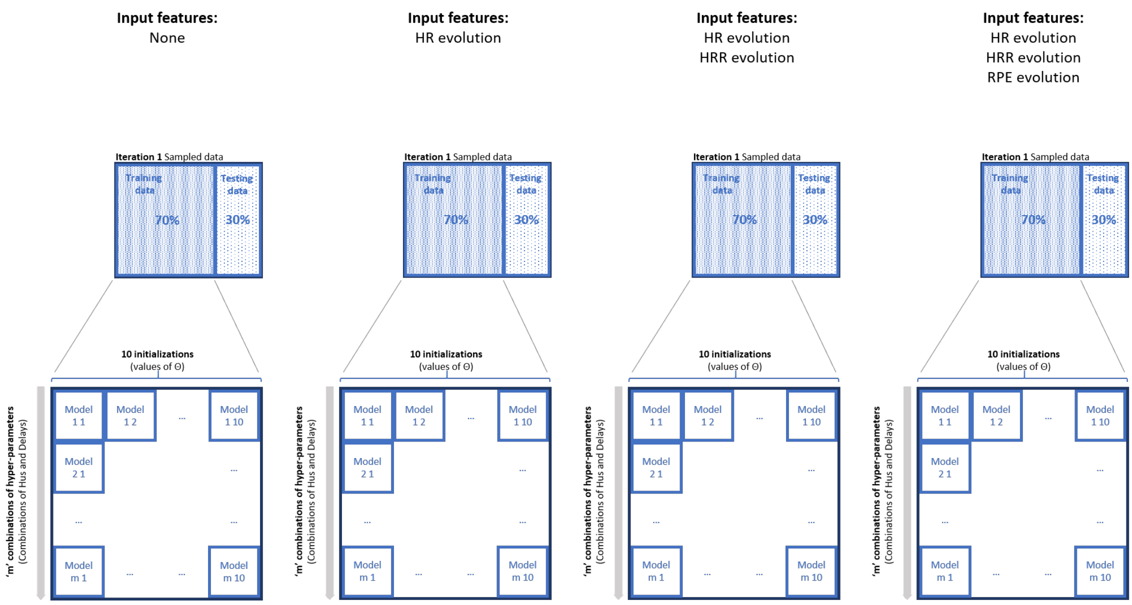


Figure 4.19: Repeating the training process for feature selection

hyper-parameter combinations across the 10 different initializations that were made. In this case, the standard deviation of the errors corresponding to each hyper-parameter set is used to estimate the variance while the bias is estimated with the mean of the models' error.

Additionally, with the purpose of evaluating the closeness to the *system's acceptable error*, the bias term is also calculated and represented in Figure 4.22. In this step, the bias evaluation only serves to discard approaches that could be far from achieving the *system's acceptable error*. As explained in Chapter 3 section 3.3, the bias term is calculated according to the performance in the test data set.

Regarding the input feature selection, as seen in Figure 4.21, the variance of the zero-feature approach has the smallest overall variance, independent of the data split method. Additionally, its bias term (see Figure 4.22) shows that the zero-feature approach (i.e. the most parsimonious one) is able to reach to the *system's acceptable error of* 10% and that it is close to more complex combinations. Therefore, according to the robustness criteria, among the four combinations of input features the zero input features is considered the best.

Regarding the train-test set splitting method, both approaches yield almost identical results which gives higher validity to the performance conclusions derived. Among them, the clustering based splitting is selected as it automatizes the laborious process of splitting and may be prone to less human errors in future steps that this approach could be repeated.

Once that the feature selection and data splitting method selection is done, the methodology performance evaluation focuses on the relation between the hyper-parameters (HU and Delays) and the performance in terms of bias and variance. To do so, as explained in section 3.1 and represented in Figure 3.7, two considerations are made: (1) the estimated low variance zones and (2) the parsimony (i.e. the smallest amount of hyper-parameters). Regarding the variance estimation, Figure 4.23 represent a contour map with the estimated bias-variance of the methodology according to the different sets of hyper-parameters. Additionally, the same Figure 4.23 also represents

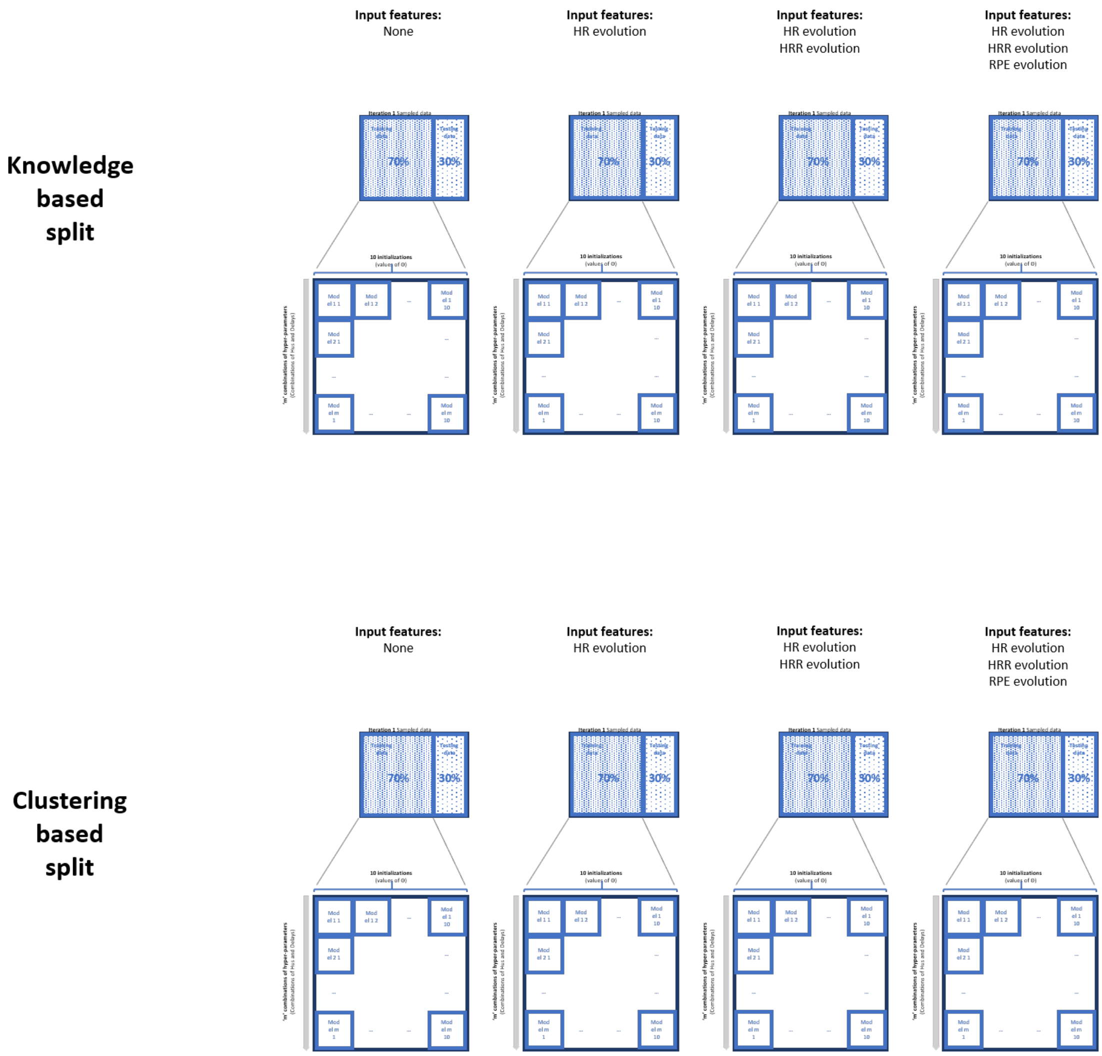


Figure 4.20: Repeating the training process for data split method selection

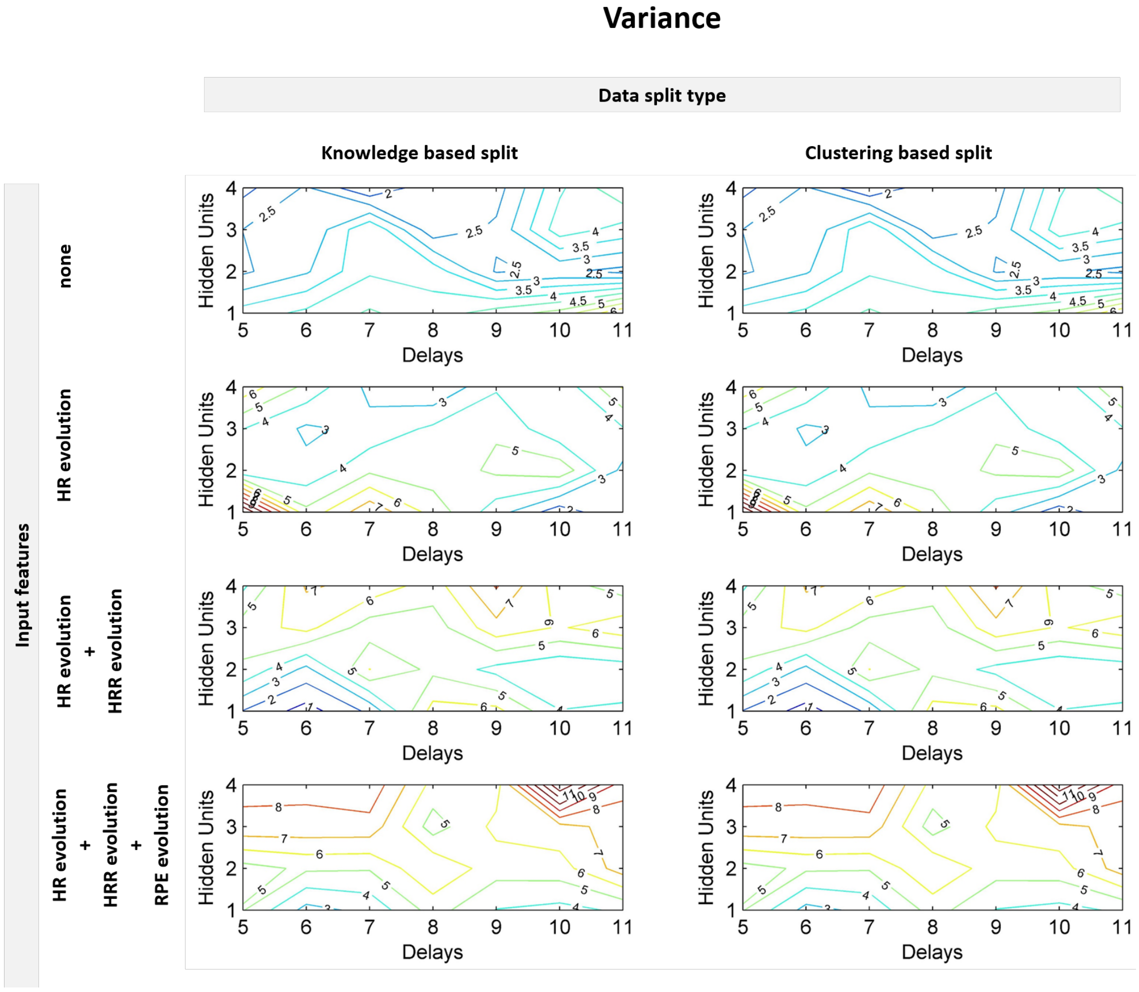


Figure 4.21: Variance of feature and data split method combinations

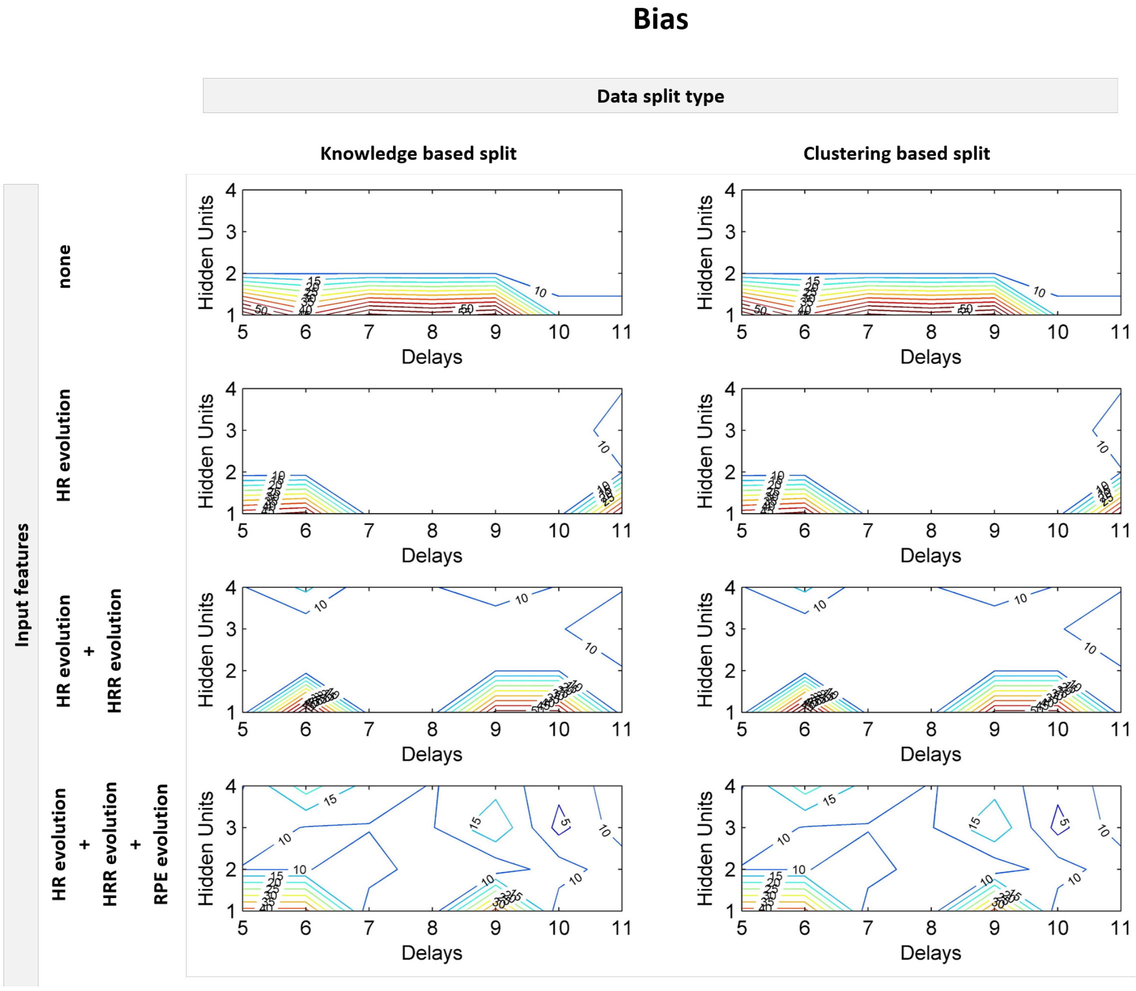


Figure 4.22: Bias of feature and data split method combinations

gradient map pointing towards the most parsimonious approach.

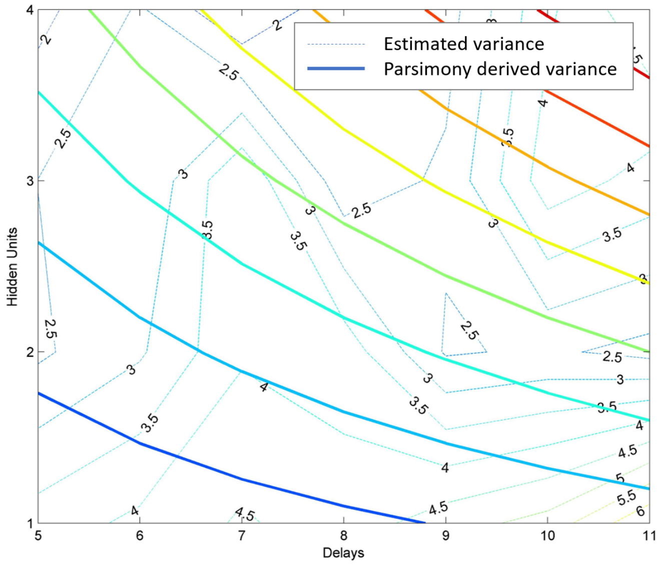


Figure 4.23: Variance estimation of zero features and clustering based split methodology

The zones that consistently find low variance and *satisficing* bias are the most interesting regions on which to find our final model.

Robust final model creation/selection

In this iteration, the final model is to be selected in the hyper-parameter zone that shows lowest variance while achieving a *satisficing* bias. Figure 4.24 represents the lowest variance zone as the best to look for the final model.

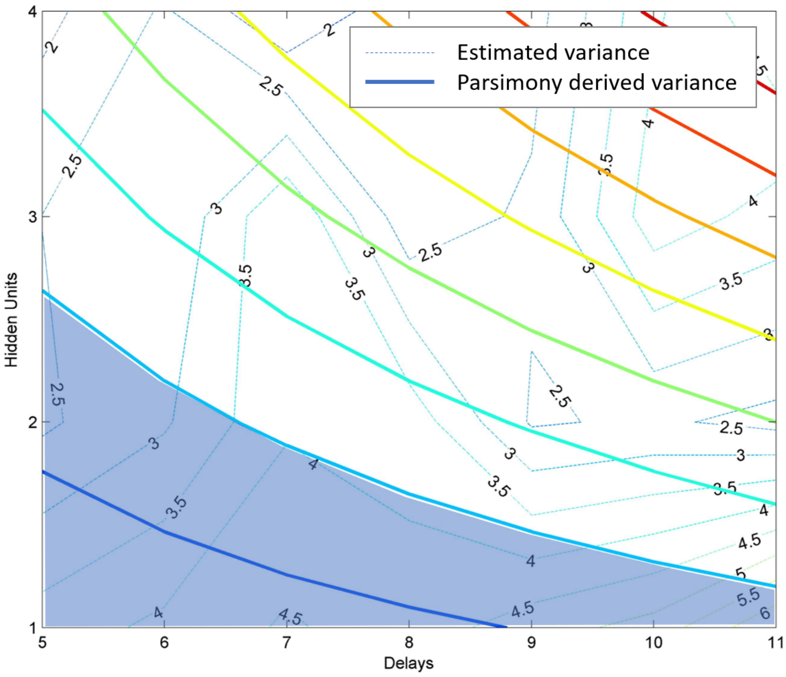


Figure 4.24: Selected hyper-parameter zone

Therefore, once that the best low variance zone is identified, all the models trained with the clustering data-splitting method are ranked in decreasing performance order. Table 4.5 represents the 10 best models among the zone selected in Figure 4.24.

Table 4.5: Iteration 1 best 10 models ranked by increasing *system's error*

Rnk	Hyper-parameters		*System's error*
	HU	Del	Test set error %
1.	2	11	6.4
2.	2	10	6.4
3.	2	11	6.4
4.	2	10	9.7
5.	2	11	9.7
6.	2	5	9.7
7.	2	10	9.7
8.	2	5	9.7
9.	2	11	9.7
10.	2	5	9.7

Abbreviations: Rnk, Rank; HU, hidden units; Del, delays; Inv, Invalid LT estimations due to error above *satisficing*.

The final model selected is the one ranked in 6^{th} position.

Final model performance evaluation

As explained in Chapter 3, the final model performance is estimated from the comparison of the performance in the training set with respect to the performance in the test set. In this regard, table 4.6 gathers the *system's error* on both training and test sets.

Table 4.6: Iteration 1 final model's performance

Perf. Ind.	Training set performance	Test set performance
System's error (%)	10.8	9.7
System's accuracy **(%)**	**89.2**	**90.3**

Abbreviations: Perf, Performance; Ind. , indicators; Train, Training set; Test, Test set.

In order to better observe the performance of the final model on the unseen data, Figure 4.25 shows the residuals analysis done for the tests set with the selected final model. Additionally, Figure 4.26 illustrates some estimated lactate curves and threshold comparing them with the real ones.

Comparing the train and test set *system's error* it is observed that they are almost equal, which suggests that the variance of the final RNN of this *initial virtual LT sensor* is very small, something fundamental for robustness.

Regarding the bias term, these results show that the estimated bias of the final model is approximately % 10, which is within the range of the *system's acceptable error*. Therefore, both

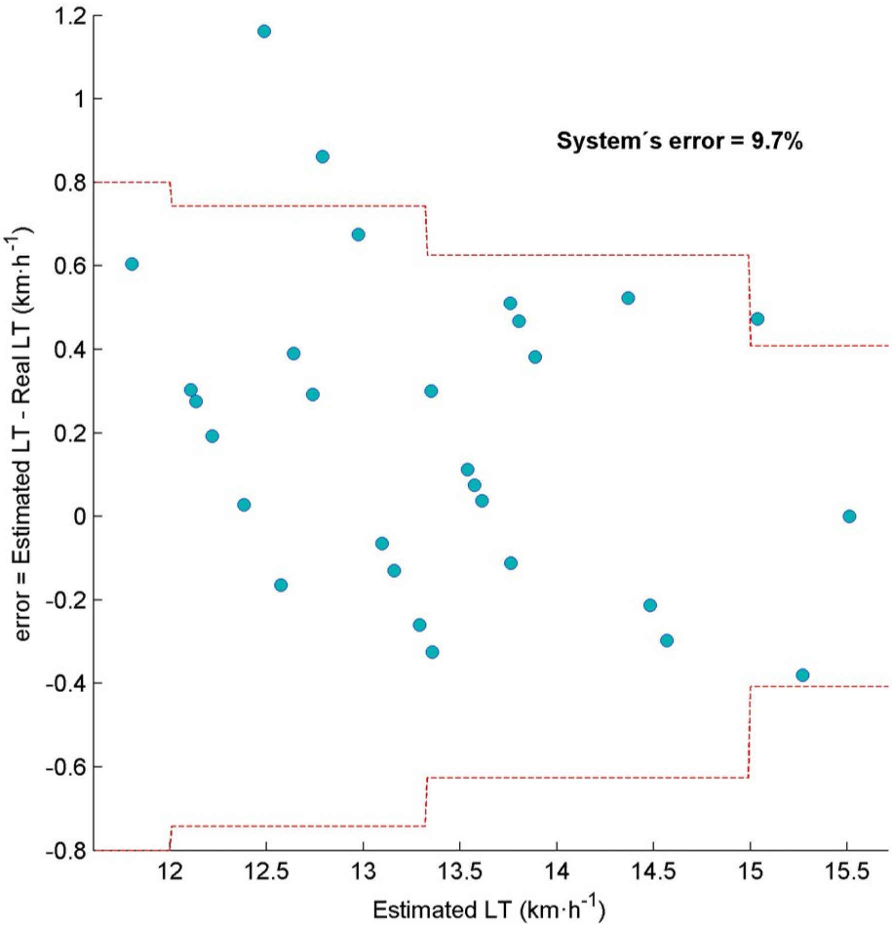


Figure 4.25: Residuals of iteration 1 final model

Abbreviations: Lac, Lactate; Test, tested; Est, estimated; LT, lactate threshold

performance criteria are fulfilled.

4.1.3 Decide next step: Towards maximizing robustness

The next-step decision making process starts from the performance evaluated in the previous section. The performance of this *initial virtual LT sensor* showed both low variance and *satisficing* bias. Moreover, the ML system bases on an output feature transformation (x-axis standardization with respect to Vpeak) and does not require from input features making it more robust. Therefore, in this iteration we have tested the hypothesis that supervised learning is possible to be used to create an operational *virtual LT sensor*.

So, after fulfilling the first two requisites for acceptance, the next question is: can the robustness of the methodology be improved? The lack of requirement of input features suggests that a simpler learning architecture may be more valid and robust to design a robust *virtual LT sensor*. Therefore, following the principles described in Chapter 3, there is room for creating a more robust *calibrated virtual LT sensor*.

Therefore, as shown in Figure 4.27, we decide that another iteration is to be made to increase

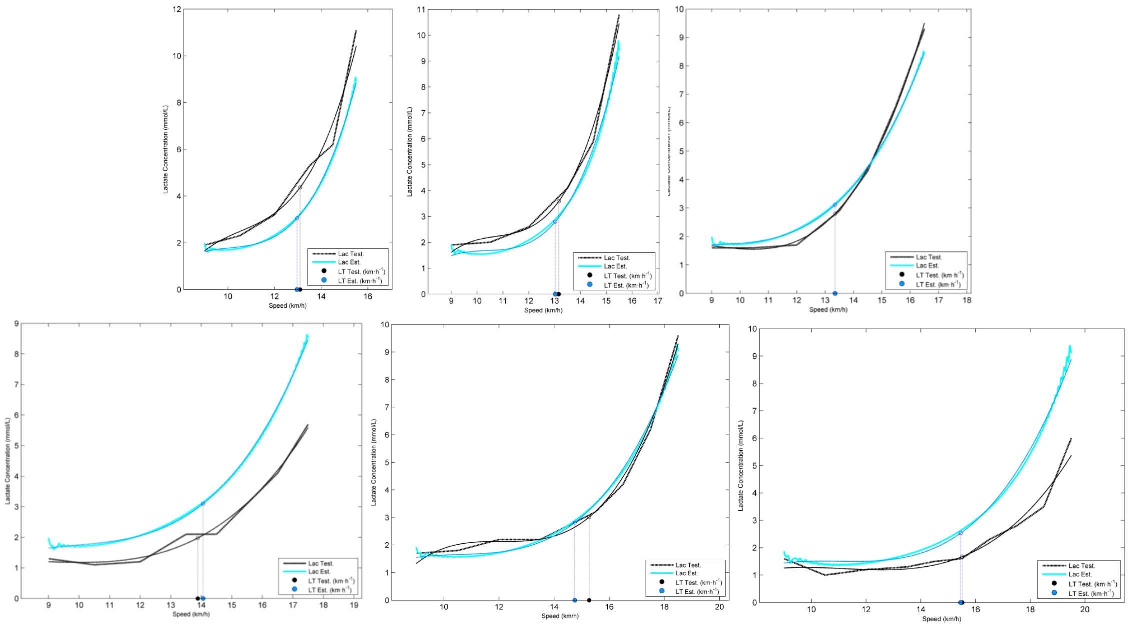


Figure 4.26: Examples of correct lactate curve and threshold estimations

: Lac, Lactate; Test, tested; Est, estimated; LT, lactate threshold

the robustness of the methodology and consequently of the *calibrated virtual LT sensor*. In particular both increasing the sample size solely for testing purposes and minimization of the learning approach complexity are seen as interesting.

As already mentioned, making these improvements are subjected to a cost-benefit trade-off. The cost of taking another iteration is evaluated from the economical, temporal, technical, material... perspectives and compared to the closeness to achieving a sufficient solution and the comparative value that provides. Increasing sample size for testing purposes may greatly benefit to further test the *virtual LT sensors* robustness, specially because the whole data sample would be entirely used for testing purposes. This involves taking more experiments which is not always economically viable. Additionally, the possible benefit of creating an even simpler model is high since it would be more robust against the unseen, more operational and also much easier to reproduce and falsify, both key to ease future attempts to extend this work to wider populations and methods.

Thus, since making the costs of acquiring new experiments and creating a more robust *calibrated virtual LT sensor* is acceptable, we decide to make another iteration to work towards further robustifying the *virtual LT sensor*.

4.1.4 Conclusions of iteration 1

In the present work, a first operational *virtual LT sensor* has been designed.

Following the methodology presented in Chapter 3, a system based on a previously consolidated ML architecture (LRNN) was created. The methodology was developed in detail using several ad hoc applied methods. A web page was created to improve the sampling diversity and

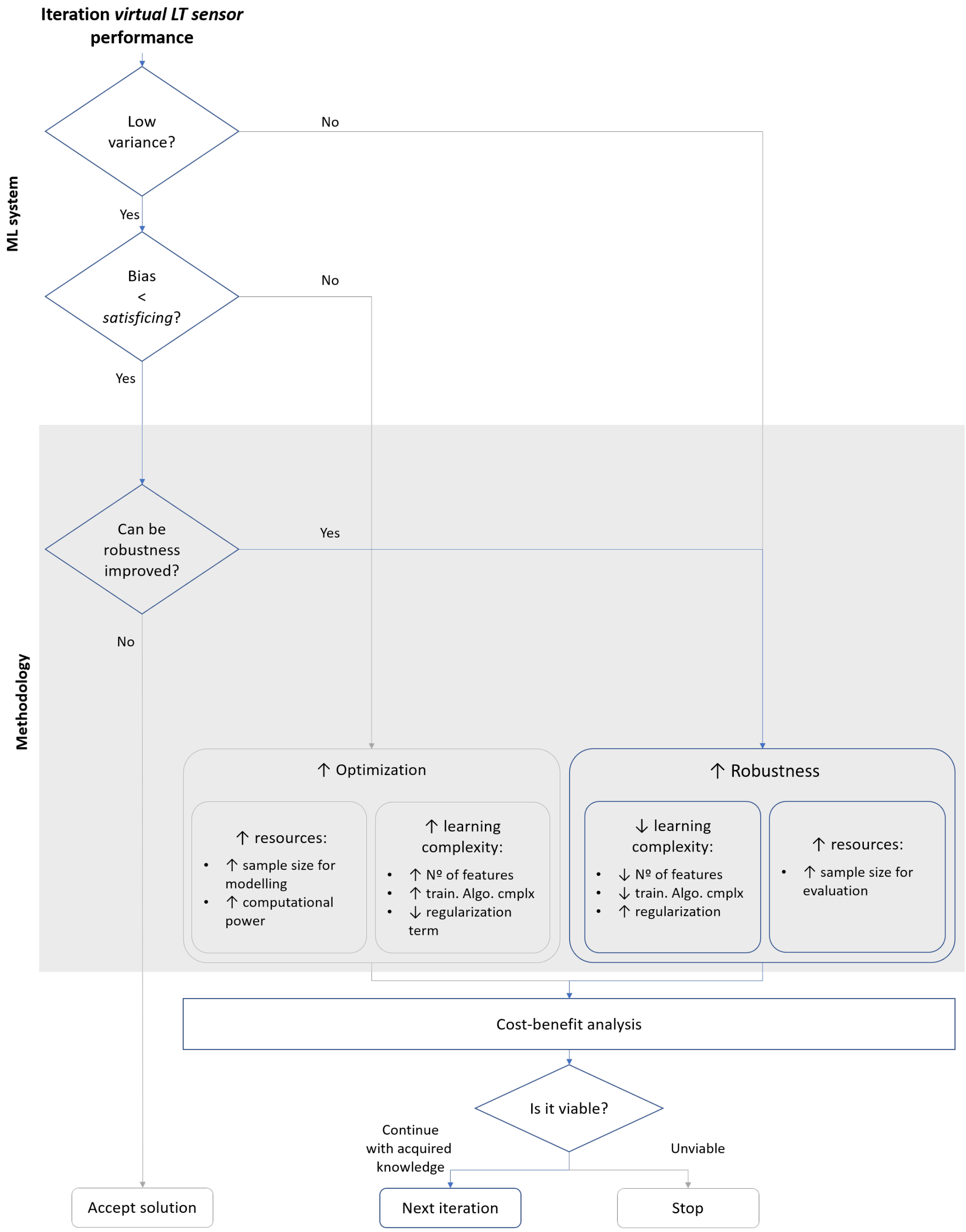


Figure 4.27: Iteration 1 next step decision making process

quality. Additionally, in order to homogenize the output feature, a standardization of the temporal axis was used. Furthermore, a combination of two database splitting methods (knowledge based and a novel modification of the stratified sampling method) were used to explore and achieve the right diversity in both data sets. Regardless whether this first iteration achieved the desired *virtual LT sensor*, a robust methodological conclusion of this iteration is that there is room for creating a supervised ML system to estimate LT if proper methodologies are followed.

Additionally, this design iteration helped to get additional knowledge about the *problem complexity* of creating an operational *virtual LT sensor* and also threw some light to its relation to other features. Regarding the LT estimation, it has been observed that the transformation of LT by means of the Vpeak diminishes most of the variability of LT. On a side note, it has also been observed that there is a strong relation between %HRmax, LT and performance related features, which may be interesting for future work.

From the application perspective, a ML system capable of successfully estimating the lactate threshold for 90.3% of the study population has been created. This is within the range of the acceptable error that we defined as *satisficing*.

Anyway, as stated in Chapter 3, despite making big efforts from the methodological point of view to maximize the robustness of the *virtual LT sensor*, we consider that there is further room for improvement. Therefore, in the next iteration the variance of the ML system is to be further evaluated and a simpler learning approach proposed to create a more robust *calibrated virtual LT sensor*.

4.2 Iteration 2: A heuristic as robust estimator of LT

This second iteration works from the *initial virtual LT sensor* towards creating a more robust *calibrated virtual LT sensor*. To do so, two approaches are followed: (1) increasing the sample size solely for testing purposes and minimization of the complexity of the learning approach to create a more robust *calibrated virtual LT sensor* by following a simpler one.

In the first iteration, we hypothesized that the lactate curves in combination with additional input features contained the necessary information to properly estimate the LT. However, we saw that a ML system with no input features was enough to achieve a *satisficing* accuracy. This means that somehow averaged the lactate curves for recreational runners. The success of this method was largely influenced by the correlation between Vpeak and LT that enabled to unify the lactate curves by the standardization of the speed axis.

Therefore, our second iteration hypothesises that the information that Vpeak contains about LT is high enough so that it can be used to estimate a LT within the *satisficing* accuracy, providing a more robust ML system. Based on what it was observed in the first iteration, the mean value of the standardized LTs arises as the simplest and most robust approach that could be used to calibrate the *initial virtual LT sensor*. Moreover, in order to better test and compare the robustness of both the *initial virtual LT sensor* and *calibrated virtual LT sensor*, additional experiments are collected. Figure 4.28 illustrates the structure of this iteration. The iteration includes a first part where *calibrated virtual LT sensor* is designed, plus the next step decision making process. The design of the *calibrated virtual LT sensor* is further divided into a context representation phase that collects and pre-processes more experiments for testing purposes. Then content characterization follows, which deals with using the mean as the most robust estimator that could be used to design the *calibrated virtual LT sensor*.

This way, a heuristic is obtained as a simplified, easier to implement and more robust solution that only requires from athletes Vpeak value obtained during the tests to estimate its LT.

4.2.1 Context characterization: Collecting further experiments for variance testing purposes

As already mentioned, to further test both the *initial virtual LT sensor* and the *calibrated virtual LT sensor* new experimental tests were made with 91 recreational athletes.

As in the first iteration, to be able to participate in the experimental test, the athletes must comply with the target population pre-requisites. Despite the multiple features defined in Chapter 3 were not to be used in this iteration for modelling purposes, they were still collected for knowledge extraction and for possible future uses. As expected, to ensure the compatibility between first and second iteration, the data collection and pre-processing followed the exact same steps of the first iteration, i.e. the methodology defined in Chapter 3.

Additionally, iteration 1 showed some interesting relations between Vpeak, LT and %HRmax with the proposed performance proxy (i.e. IAAF score), which may inform about an athlete's performance qualities. Therefore, we took advantage of this experimental design to measure a more realistic performance indicator and further explore these relationships. Concretely, 40 ath-

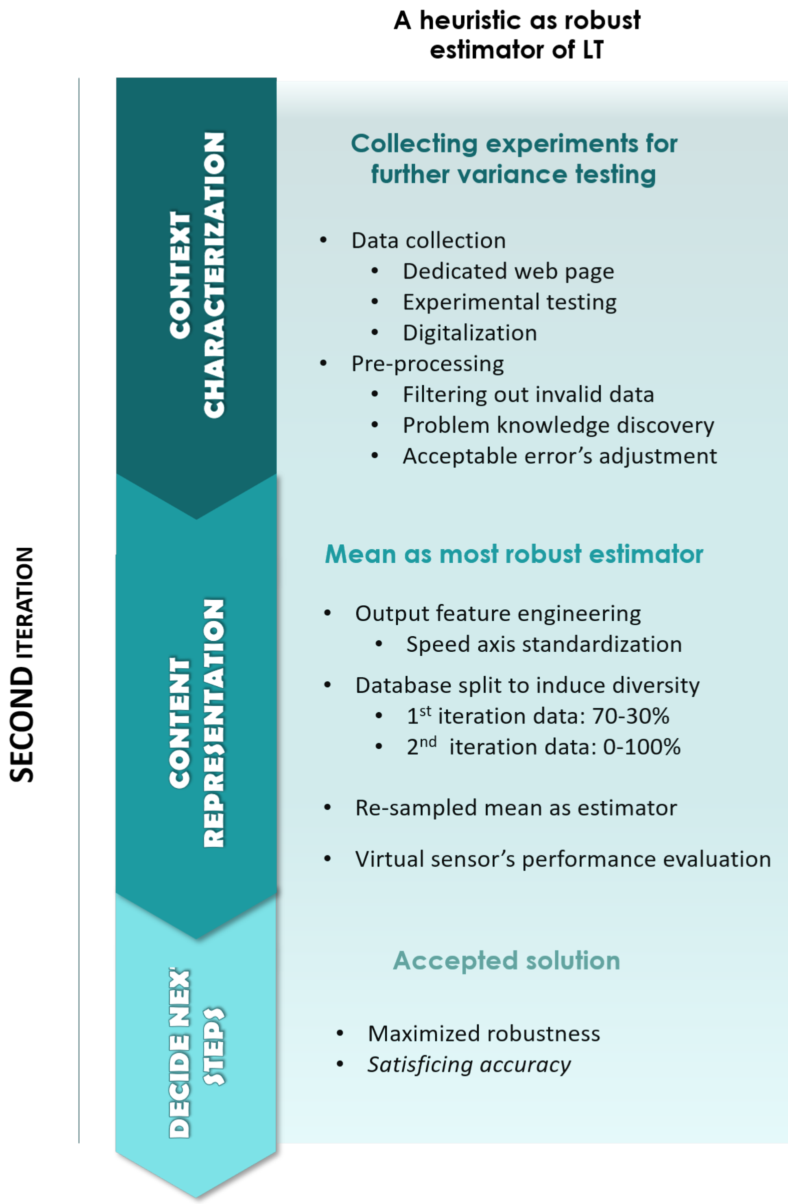


Figure 4.28: Structure of second iteration design steps

Table 4.7: Iteration 2, from collected experiments to valid ones

	Iteration 2
Selected for experiments	**91**
of which:	
No pre-requisite compliance	1
Performed experiments	**90**
of which:	
Sampling error: Major contingencies	1
Sampling error: Non-target population	10
Valid for *acceptable error* calculation	**79**
of which:	
Data error: Incorrect lactate curves	1
Retrievable from iteration 1	9
Valid for *designing*	**87**

Abbreviations: LT, lactate threshold

letes participated in an official 10K race so that the race time can be used as an accurate indicator of performance.

After performing the experiments, those with sampling errors, i.e. those with major contingencies or non-target population runners (see Figure 3.14 in Chapter 3 section 3.2.3) were detected and discarded.

Then the remaining experiments were analysed in terms of their validity for acceptable error calculation (see Figure 3.15 in Chapter 3 section 3.2.3). Using both iteration data valid for acceptable error calculation, it is estimated that approximately the 3% of the lactate curves obtained with the Dmax protocol were flagrantly incorrect and consequently, so are its LTs.

Regarding the validity for designing (see Figure 3.21 in Chapter 3 section 3.2.3), since no input feature is to be used, all the experiments with correct lactate curves are valid for this purpose. Moreover, this fact allowed us to recover 9 experiments of iteration 1 that were discarded for incorrect input feature measurements. Therefore, 87 remained valid for designing. Table 4.7 makes a summary of the data collection and pre-processing. The following sections get into the details of it.

Data Collection

The data collection followed the same approach that iteration one. From the list of 803 athlete's that already volunteered for the study, 91 athletes were randomly selected without replacement. From the sampled athletes one was rejected due to not fulfilling the pre-requisites, while the rest 90 athletes were considered able to perform the experiments.

From the 90 athletes that performed the tests, 40 athletes participated in a official 10K running race to measure their real performance level (see Figure 4.29). This race took place at most 45 days after the experimental tests in Donostia-San Sebastian under the name ”XV Carrera de Primavera

Donostiarrak" [1]. The race was recorded by the officials of the Local Athletic Federation using a chip at the start and end the of the race.

Figure 4.29: Recreational runners participants in "XV Carrera de Primavera Donostiarrak" [1].

In the treadmill speed test, as in the previous iteration, HR evolution was collected in its dedicated software. The rest of the features, together with any issue observed during the tests, were collected in both paper an digital format. After-cross checking the three formats (paper, digital and HR monitor software), the final results were gathered in a digital database in excel format (Microsoft, Redmond, Washington, USA).

Pre-processing

As already mentioned, the pre-processing serves three main purposes, filtering out invalid data, knowledge discovery and adjusting the *satisficing* errors. These are done according to the methodology stated in Chapter 3 and consequently in the same manner than in the first iteration.

Filtering invalid data

As in the previous iteration, first of all, the sampling errors are analysed and the incorrect experiments discarded. From the 90 athletes that performed the tests one of them had a major problem during the test. Additionally, among the remaining 89 athletes, 10 are found not to fulfil the target recreational runner population characteristics due to not arriving to the minimum 14.5 km/h running stage. Therefore, after cleaning the invalid data coming from sampling errors, prior to looking for incorrect data, 79 remained valid for both *acceptable error* and *virtual LT sensor* designing.

Next the the analysis was done from the data perspective (see Figure 3.14). As in the first iteration, this analysis first deals with detecting, classifying and reserving the data that is valid for *acceptable error* adjustment. As previously mentioned, both the incorrect lactate measurements and curves are valid for this purpose. Figure 4.30 represents all the 79 lactate curves and show that one of them is flagrantly incorrect. All these lactate curves and LT were reserved for the *system's acceptable error* adjustment.

The third and last pre-processing step deals with detecting and separating the experiments that are valid for *virtual LT sensor* designing purposes. In this regard, apart from the experiment

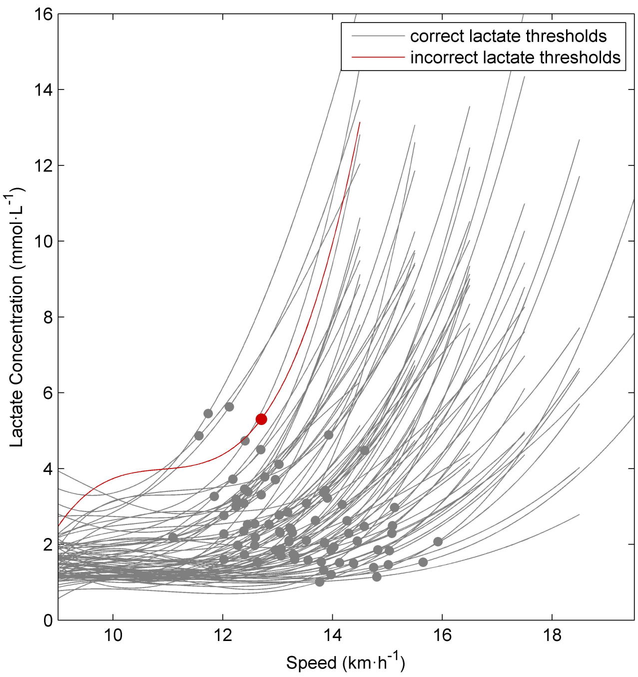


Figure 4.30: Iteration 2 incorrect lactate curves

with incorrect LT curve, the input features were also scrutinized for incorrect data. Although no input feature is used nor for *satisficing error* calculation nor for modelling and an analysis of the input features seems unnecessary, they are still performed as they may uncover incorrect data that indicates flagrant experimental error, specially about the maximality of the test. Once the incorrect LT experiment discarded, 78 experiments remain valid for designing purposes. Table 4.8 gathers the mean,standard deviations and coefficient of variation of the numerical features. Here again we can see the heterogeneity of the population highlighted in the high coefficients of variation of features such as: ”IAAF scores” and ”training years”. Additionally and despite this heterogeneity it is also observed that the Coefficient of Variation of LT is relatively low.

Finally, since no input features are needed for designing purposes, we were able to recover 9 experiments of iteration 1 which had good lactate curves and but were previously discarded for some flagrant incorrect input feature. Therefore, 87 were finally valid for designing purposes in this second iteration.

Problem’s knowledge discovery

As in the first iteration, a bi-variate statistical analysis is also done to look to the interrelations between features, expecting to be of value for feature engineering purposes or to discover new knowledge that could be useful. Additionally, unlike the first iteration, using the 10K race times collected allow us to make a deeper analysis of the several features related to running performance. Figure 4.31 illustrates the interrelations between all these features. For the sake of clarity and comparability with the first iteration, only strong correlations above R=0.75 are illustrated. On a brown to yellow range, the closer to yellow the higher the correlation value.

In the first iteration, we discovered a potentially interesting connection between Vpeak, LT

Table 4.8: Iteration 2 numerical features mean, standard deviations and coefficient of variation

Feature	Mean ± SD	CoV	sample size
Lactate threshold [km/h]	13.4 ± 1.0	7.7	78
10K time [min:sec]	38:38 ± 2 min 58 seg	7.7	40
Personal best [IAAF points]	399.4 ± 193.2	48.4	58
Vpeak [km/h]	16.7 ± 1.4	8.5	78
Years train [years]	9.4 ± 7.7	81.6	78
Age [years]	38.5 ± 11.7	30.4	78
Height [cm]	171.7 ± 15.0	8.8	78
Weight [cm]	67.4 ± 13.0	19.3	78
Body mass index [kg/m^2]	22.5 ± 2.1	9.5	78
Abdominal diameter [cm]	79.9 ± 13.6	17.0	78
Hip diameter [cm]	93.1 ± 4.8	5.1	78
Body fat percentage [%]	16.0 ± 5.8	36.0	73
Water percentage [%]	61.7 ± 4.5	7.3	73
Resting HR [bpm]	58.0 ± 10.1	17.4	78
Maximum HR [bpm]	182.6 ± 10.2	5.6	78
%HRmax at 13.5 km/h [%]	87.3 ± 5.9	6.8	49
%HRmax at 14.5 km/h [%]	91.4 ± 5.5	6.0	49
%HRmax at 15.5 km/h [%]	94.3 ± 4.4	4.6	45
%HRmax at 16.5 km/h [%]	96.2 ± 3.7	3.8	33
Heart rate deflection point [km/h]	13.4 ± 2.1	16.0	37
HRR threshold [km/h]	13.9 ± 1.2	8.6	44
Resting Lactate [mmol/l]	1.3 ± 0.3	21.2	78
Maximum Lactate [mmol/l]	10.1 ± 2.9	28.3	78
Lactate value at 13.5 km/h [mmol/l]	3.2 ± 2.1	66.4	78
Lactate value at 14.5 km/h [mmol/l]	4.9 ± 3.4	69.5	78
Lactate value at 15.5 km/h [mmol/l]	5.3 ± 2.9	55.0	62
Lactate value at 16.5 km/h [mmol/l]	6.3 ± 3.1	49.0	44
Maximum muscular Borg [Borg scale]	8.4 ± 1.8	21.4	78
Maximum respiratory Borg [Borg scale]	8.8 ± 1.9	21.2	78

Abbreviations: LT, lactate threshold; CoV, coefficient of variation

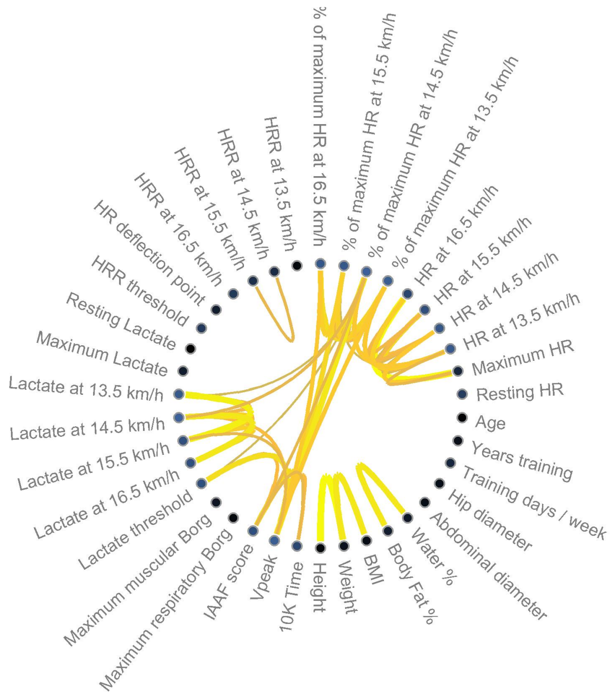


Figure 4.31: Iteration 2 network correlation analysis

and %HRmax and IAAF score used as performance proxy. In this second iteration we decided to put special attention on it and included a 10 km race time as a more robust performance proxy. As represented in Figure 4.31, the associations between these features are still strong as in iteration 1 and 10K race time also falls into the same cluster. This strengthens the previous hypothesis that the Vpeak, LT and %HRmax are interesting performance indicators of endurance running, especially in 10K distance.

Acceptable errors' adjustment

The robustness of the estimation of the acceptable errors' grow together with the sample size. Thus, both iteration observations are combined to calculate it.

For the *individual acceptable error*, including the first iteration points, 189 LT points are available (194 valid for *acceptable error* calculation - 5 incorrect lactate curves). Using them, algorithm 1 is computed. As in the first iteration, different ”W” numbers of bootstrap re-samples (10, 20 and 100) are used on the 189 athletes. Different ”Z” number of random samples (10, 20 and 100) have also been used for each of the blood lactate measurements. In both cases the results from 100 random do not significantly differ from those obtained with 20, so a higher number of random samples is not considered necessary.

From the computation of algorithm 1, we have estimated the standard deviations of the *Dmax LT method precision error* and illustrated in table 4.9, classified by the ”Y” number of blood lactate points taken during the treadmill speed test (Lactate Points column, table 4.9). As in iteration 1, these adjusted results show how the *Dmax LT method precision error* improves with the number of blood lactate measurements taken from the athletes, which is consistent with what is known in the literature and practice ([34]). Moreover, the increased sample potentiated with the re-sampling

Table 4.9: *Adjusted Dmax lactate threshold method precision error* according to number of lactate points

Y Lactate points	*Dmax LT method precision error* real LT (SD)
5	4.7
6	3.2
7	2.6
8	2.4
9	1.7
10	1.7

Abbreviations: LT, Lactate threshold ; SD, standard deviation; It, iteration

throws non-linearly improved results. Table 4.9 shows how the *Dmax LT method precision error* consistently improves for each number of lactate points measured.

To further analyze how the *Dmax LT method precision error* can affect our estimator, we compare it to the *individual acceptable error* defined in table 3.1 in Chapter 3 in section 3.2.3 and represent the residual errors in Figure 4.9. In this Figure, we observe that, as in iteration 1, the 98.8% of the plausible Dmax LTs are within the range of the *individual acceptable error*. In other words, from this analysis the *virtual LT sensor* could at best achieve a *ceiling accuracy* of 98.8%.

Therefore, as represented in Figure 4.32 and also concluded in iteration 1, the *LT method precision* errors derived from the physiology equipment and methodology is sometimes even higher than the *individual acceptable error* determined by expert knowledge. This fact reinforces that this *satisficing* threshold is a safe and robust metric to determine the validity of our *initial virtual LT sensor*.

Regarding the *system's acceptable error rate*, using the *LT method precision* we derived that the adjusted ceiling accuracy is at least of 98.8%. Moreover, as previously shown, there are 5 incorrect lactate curves among the total measured. This means that we can consider that the Dmax protocol fails for 4% of the athletes (3% due to the incorrect LT plus 1% due to the *Dmax LT method precision error*) which gives an additional perspective of how strict we can be for our *virtual LT sensor*. Therefore, it seem clear that a *system's acceptable error rate* 90−95% *satisfices* the objectives of our *initial virtual LT sensor*.

4.2.2 Content representation: the mean as the most robust estimator

This second iteration focuses on creating a more robust *calibrated virtual LT sensor*. This design starts from the hypothesis that the mean value of the LT defined in standardized terms (i.e. with respect to Vpeak) may be able to create very robust *virtual LT sensor* with *satisficing* accuracy. Then, both iteration *virtual LT sensors* are compared to determine which one is most desirable in terms of *satisficing accuracy* and robustness.

Based on the robustness criteria, the methodology here proposed estimates the LT determined in the aforementioned incremental treadmill speed test conditions using the mean of the standard-

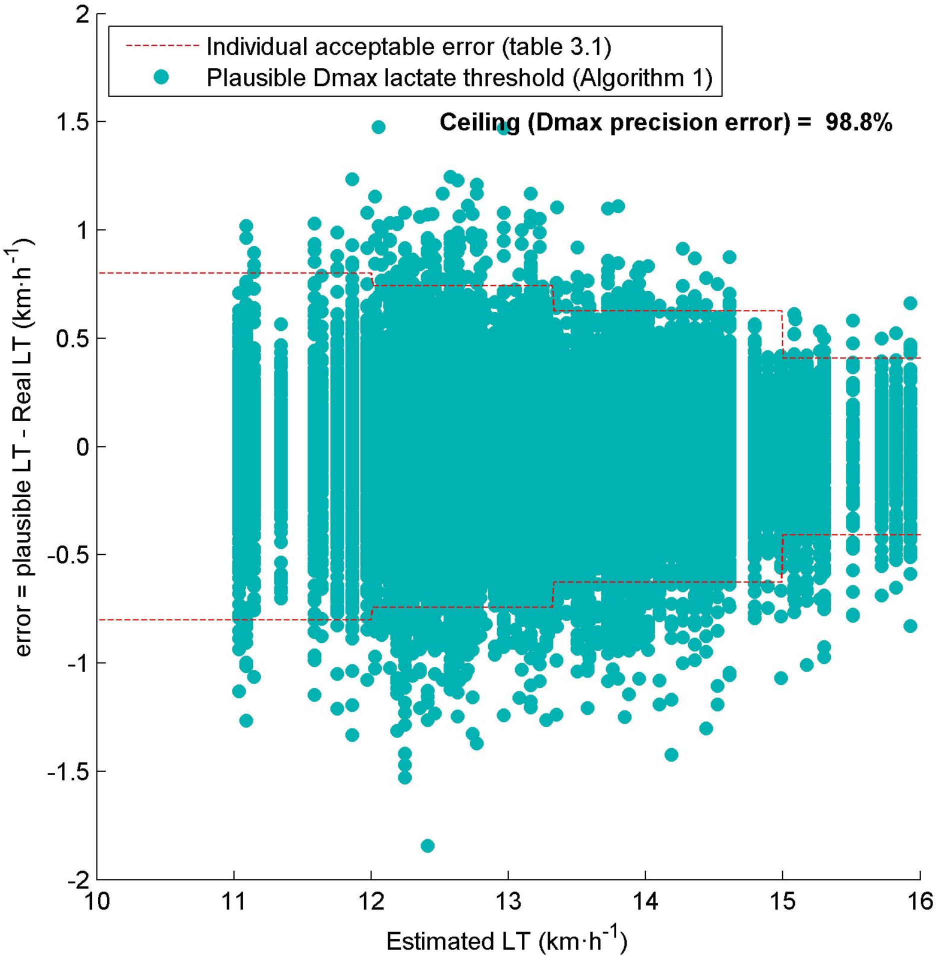


Figure 4.32: *Adjusted Dmax lactate threshold method precision error* VS Individual acceptable error

ized LT with Vpeak as the only feature.

The different database split, feature engineering, training... steps ensure that this general approach is robustly materialized. Figure 4.33 gives an overview of these steps.

Robust output feature engineering: Standardization of lactate curves x-axis

The standardization of the speed axis is the only feature transformation on which this context representation approach relies on.

Figure 4.34 shows the raw and standardized curves of both iteration 1 and 2 experiments. As in the first iteration, it is observed that once we dis-attach the LT from the effect of the Vpeak, the variability of the LT is highly reduced with respect to the x-axis and thus the problem is simplified.

Therefore, this feature transformation represents each LT in relative terms whit respect to the individual *endurance running speed reserve* (6):

LT [%] = (LT [km/h] - initial running speed [km/h]) / endurance running speed reserve [km/h] ×

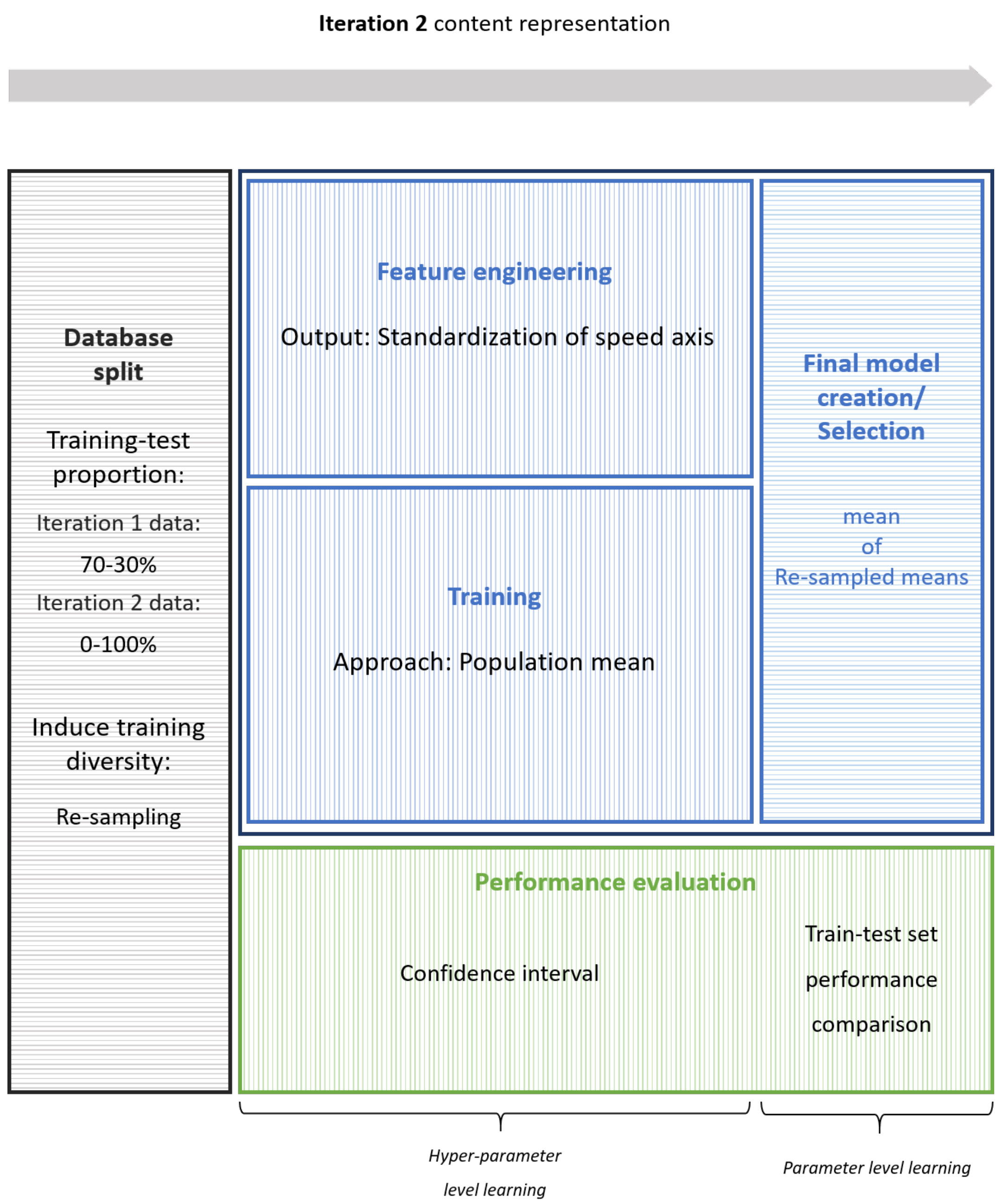


Figure 4.33: Content representation for calibrated *virtual lactate threshold sensor*

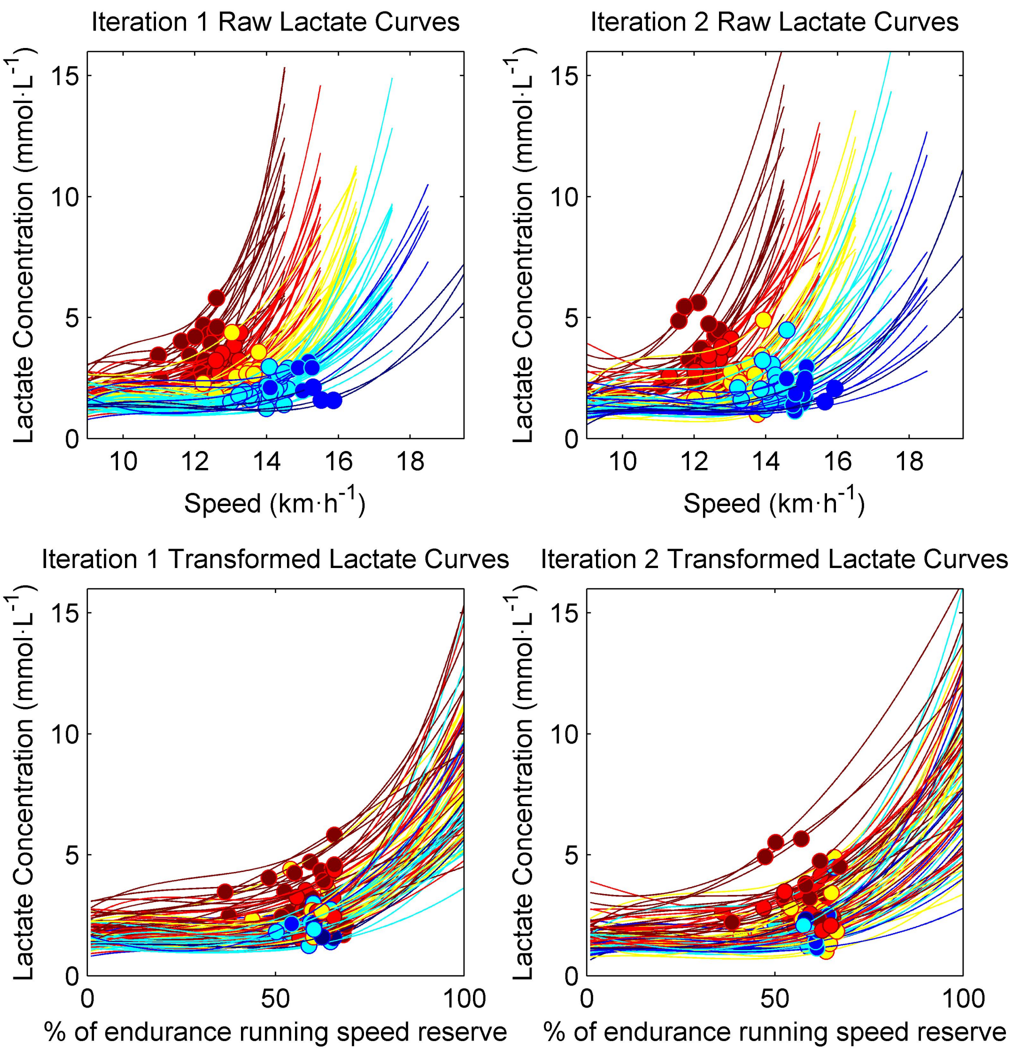


Figure 4.34: Iteration 1 and 2 raw and standardized lactate curves

Abbreviations: LSF, last step finished

$$100 \text{ (6)}$$

This transformation allows to put LT in relative terms using the initial running speed (fixed for the proposed protocol) and Vpeak. This means that a single feature is needed for this transformation and thus it is the most parsimonious approach that could be used.

Database split to induce randomization

As represented in Figure 4.35, in the first iteration we used the available data to create the first *virtual LT sensor*. Now, in this second iteration, we combine the data collected in both iterations for a higher available data.

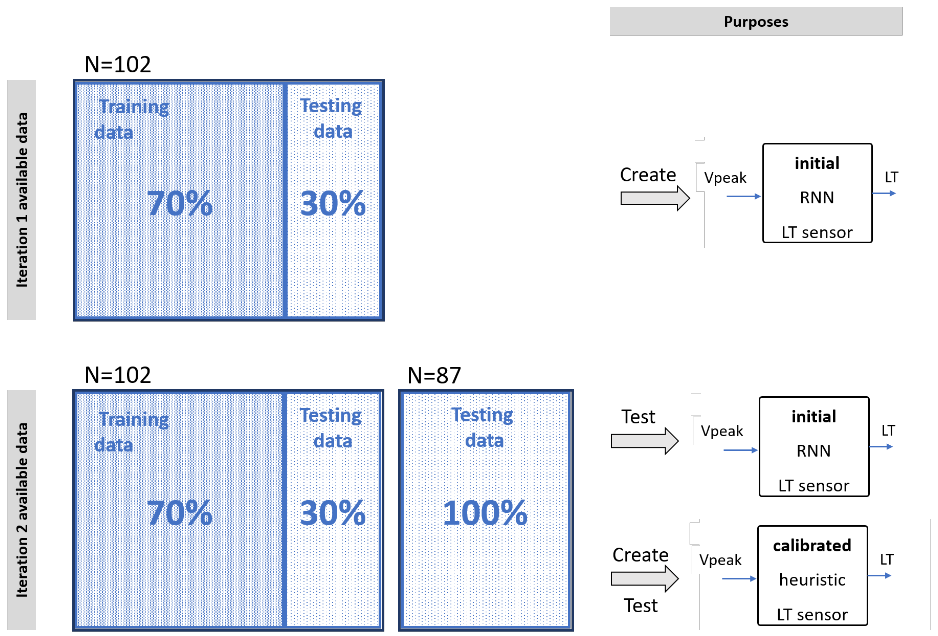


Figure 4.35: Training and testing set split comparison between iterations

The purpose of this iteration is to create a more robust *calibrated virtual LT sensor*. To be able to ensure that this *calibrated virtual LT sensor* is more robust, it is fundamental to evaluate it in the same terms of the *initial virtual LT sensor*. Therefore, as shown in Figure 4.35, the same data splitting is used, i.e. the data-set used in the first iteration is employed to create this calibrated model, and the rest of the data is used to evaluate it.

Regarding the creation of a more robust *calibrated virtual LT sensor*, a re-sampled mean is calculated using the 71 samples of the training data (n=102 x 70%). 100 re-samples are taken since more re-sampled did not provide significant improvements to the results.

Robust training: final estimator creation with re-sampled mean

As already mentioned, the transformed population's LT mean (with respect to Vpeak) is to be used as estimator of the LT. The mean is arguably the most used estimator as a simple measure of the central tendency of a probability distribution. Therefore, the training done in this step is so simple that it lacks from hyper-parameters and intermediate evaluation steps.

Then, using the % of *endurance running speed reserve* corresponding to the LT as estimand, we use the mean as an extremely *parsimonious* estimator. It is important to note that, the mean, by definition, would produce a LT estimation with no variance (i.e. the estimation is always a fixed value which corresponds to the mean value of LT) and therefore it is the most *parsimonious* estimator that we could use.

However, the mean of a probability distribution is the long-run arithmetic average value of a random variable having that distribution. Since the data that is averaged is finite (training set data in Figure 4.35), the LT mean may slightly differ depending on the data sample. In order to estimate a mean closer to the expected mean, multiple re-sampled mean values are combined. Additionally, we address the confidence intervals (CI) of the estimated mean value using the bias corrected and accelerated percentile bootstrap method which is a non-parametric and more robust way to estimate the CI [93]. This calculation shows that the CI are narrow around the mean value (60.0% (58.8 - 61.2)) and thus that the estimation is robust.

Therefore, the 60% of the *endurance running speed reserve* is to be used as the heuristic to estimate the LT.

Virtual LT sensor's performance evaluation

Using the database created in the context characterization step, the performance of the *initial virtual LT sensor* and the *calibrated virtual LT sensor* are evaluated against the new experimental data. Table 4.10 compares each estimated performance. As expected due to their simplicity, both the *initial virtual LT sensor* and the *calibrated virtual LT sensor* show very similar bias between iteration 1 and 2 data sets, due to the robustness of the approaches. Apart from the raw estimated performance, table 4.10 also considers the 4% *intrinsic error of the Dmax LT* calculated in the pre-processing step. Since this error is an irreducible part of the output feature (LT) it is subtracted from the raw performance to obtain the total accuracy. Thus, *Total error* is the one to be used to compare it with the *system acceptable error*.

Table 4.10: Initial and calibrated *virtual lactate threshold sensors* system error

virtual LT sensor	1^{st} Iter DB Train set error	1^{st} Iter DB Test set error	2^{nd} Iter DB test set error
Initial (raw)	10.8%	9.7%	8.1%
Initial (total)	**6.8%**	**5.7%**	**4.1%**
Calibrated (raw)	9.9%	9.7%	11.5%
Calibrated (total)	**5.9%**	**5.7%**	**7.5%**

Abbreviations: Iter, Iteration; DB, Data base; perf, performance

The variance term represents how an estimator's performance, the *virtual LT sensor* in this case, varies withing sub-samples of the same population. Thus, to calculate the variance term, we look to the worst case scenario and compare the most extreme errors obtained in different data sets. This means that, for the *initial virtual LT sensor* the Train set error and 2^{nd} iteration test set error are compared which shows a 2.7% variance. For the *calibrated virtual LT sensor* the 1^{st}

iteration test set error and the 2^{nd} iteration test set error are compared, which shows a variance of 1.8% when doing the same.

Looking to the bias, the *initial virtual LT sensor* shows an smaller bias in both test sets (5.7% and 4.1%) compared to the *calibrated virtual LT sensor* (5.7% and 7.5%). In any case, both are within the range of the *system acceptable error* of 5-10% defined in Chapter 3.

In any case, these quantitative results must be taken with a grain of salt. If we look into the detail of these small performance variations, we observe that the differences may easily be due to very few incorrectly estimated LTs. More precisely, a single incorrect LT estimation in the 2^{nd} iteration test may cost around 1.1% of error. This means that, the performance differences between both *virtual LT sensors* are spurious. In other words, the only robust conclusion that may be derived from this performance analysis is that both *virtual LT sensors* are in the same order of magnitude in terms of their bias-variance, both are robust and both fulfil the *system acceptable error*.

Therefore, to select the final *virtual LT sensor*, apart from the previous quantitative performance perspective, the *initial and calibrated virtual LT sensors* are compared in terms of parsimony. In equal conditions, as defined in Chapter 3, the most parsimonious approach is preferred and since the number of parameters of the heuristics are zero, this is the most parsimonious approach.

Finally, both *virtual LT sensors* are also compared in operational terms. While the *initial virtual LT sensor* is based on a RNN that need to make external computations to estimate the LT (implemented in a web page or device), the heuristic (*calibrated virtual LT sensor*) provides a method for coaches and athletes that does not require from any tool, device nor computation, as it is the computation of the 60% of Vpeak. Therefore, the *calibrated virtual LT sensor* is selected and accepted as the best approach.

In order to better observe the performance of the selected *calibrated virtual LT sensor* on the unseen data, Figure 4.36 shows the residuals analysis of the heuristic with respect to the *individual acceptable error* and with the 1^{st} iteration data.

In Figure 4.37 we represent the performance of the *calibrated estimator* with respect to the second iteration database and the *Dmax LT intrinsic error*.

Additionally, we can further analyse the *system's error* that would correspond to the extreme values of their confidence interval. As represented in table 4.11, results show that, *system's error* corresponding for the lower and higher CI values (i.e. 58.8% and 61.2%) are the same than for the the mean 60.0% *system's error*. In every case the *system's error* is of 7.5% which means that the *calibrated virtual LT sensor* very robust.

Table 4.11: System's error corresponding to confidence interval's extremes

mean (boots CI)	*system's error* range
60.0% (58.8 - 61.2)	7.5% (7.5% - 7.5%)

Abbreviations: LT, lactate threshold; boots CI, bootstrap confidence interval

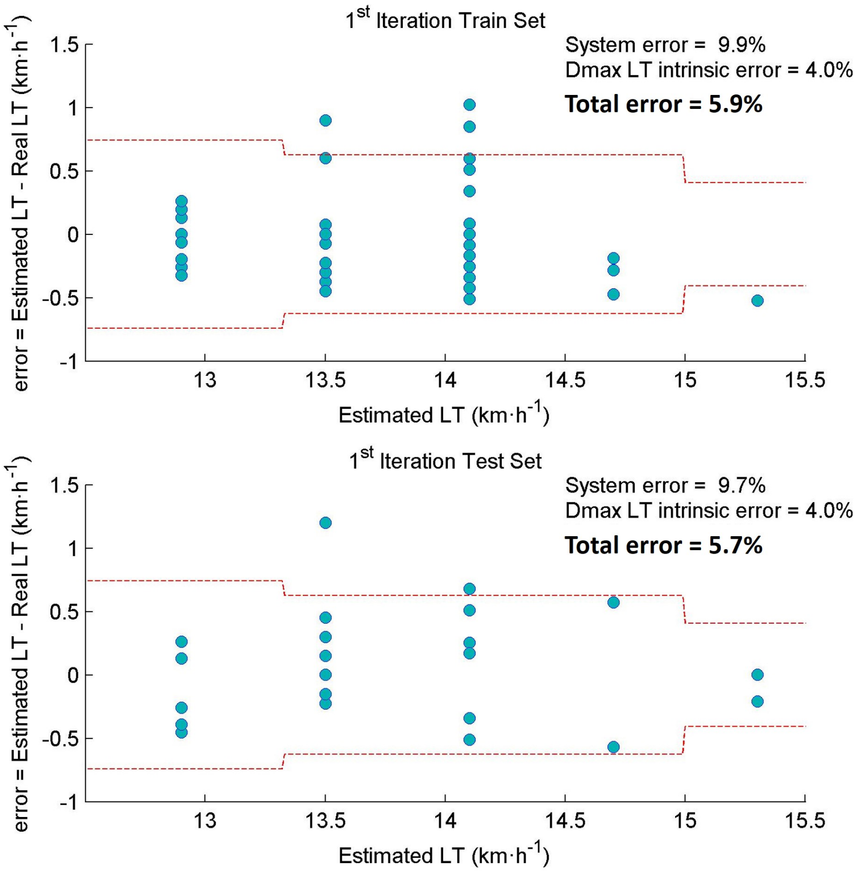


Figure 4.36: *Calibrated heuristic lactate threshold sensor's* residuals: Iteration 1 Data

4.2.3 Decide next step: accepting the *calibrated virtual LT sensor* as valid solution

In the previous content representation step, we concluded that the variance of the *calibrated virtual LT sensor* is extremely low. Thus, following the next step decision making process represented in Figure 4.38, it fulfills the first requisite towards an acceptable solution. Moreover, the bias error is also below the *satisficing acceptable error*. Finally, the robustness of the entire design has been maximized.

Therefore, and considering that the most robust possible approach has been followed, the *calibrated virtual LT sensor* is accepted as final solution.

4.2.4 Conclusions of iteration 2

In this second iteration, we created a *calibrated virtual LT sensor* based on a simple heuristic that estimated the LT using the 60% of the *endurance speed reserve* as estimand.

Following from the *initial virtual LT sensor* created in the first iteration, a more robust system was created by focusing on a simpler approach. In the first iteration, it was discovered that the standardization of the temporal axis was able to homogenize the LT of different level athletes, and that it is so mostly because of the relation that Vpeak had with the LT. This second approach brought this relationship to the center and exploited it to create a extremely robust *calibrated*

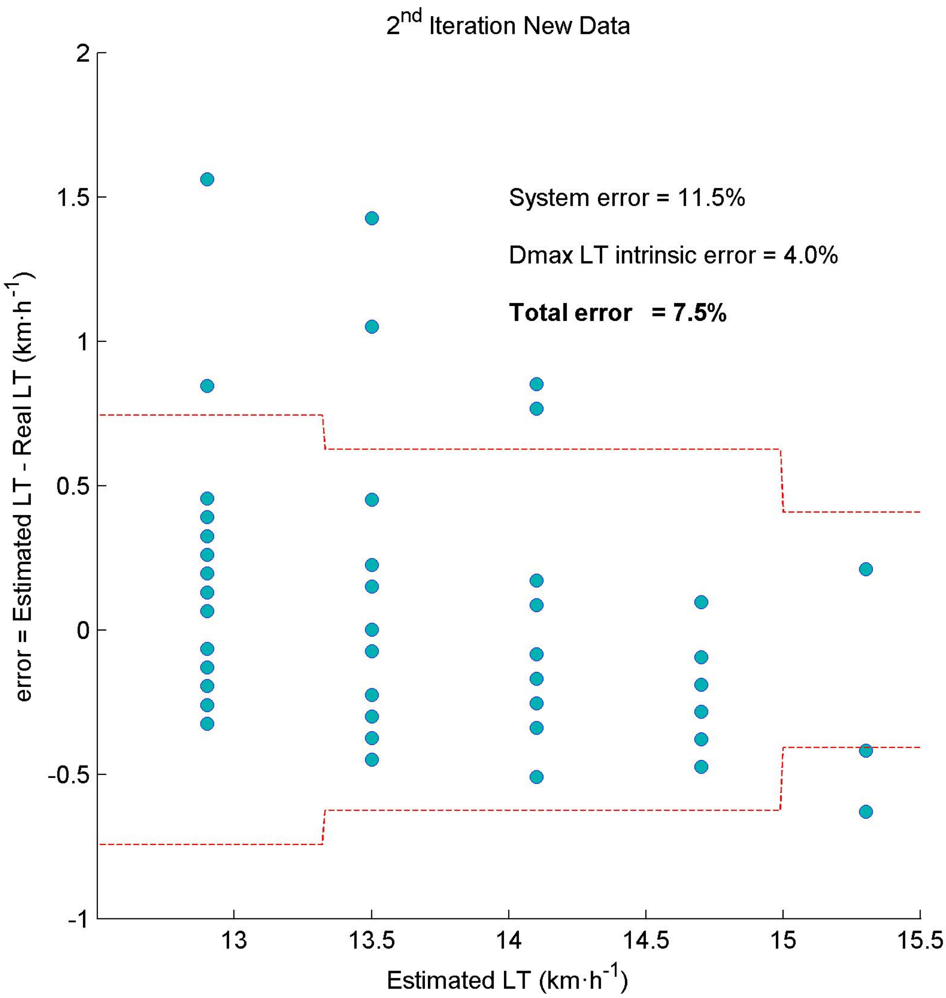


Figure 4.37: *Calibrated heuristic lactate threshold sensor's* residuals: Iteration 2 New Data

virtual LT sensor. To do so, the mean of the standardized LT was proposed as the simplest possible estimator. A methodological conclusion of this iteration is that, as it has been done in this work, when evaluating the accuracy of a supervised learning system, it is of major importance to make a deep analysis of intrinsic errors of the input features and use them as reference to compare it with the final estimation errors.

From the sport science perspective, this design iteration further expanded the knowledge about the *problem complexity* to create an operational *virtual LT sensor* and continued digging into the relation between the different features. In particular, an interesting association network was found around %HRmax at various stages, LT, Vpeak, IAAF score and 10K race times. This opens the opportunity to further evaluate other features as proxies of performance as it is done with LT.

From the application perspective, the main objective of the present work was to create an accessible method to estimate the LT and to facilitate its integration into the training process of recreational runners. We showed that a heuristic (%60 of *endurance running speed reserve*) fulfills this objective as it is as reliable as the Dmax LT protocol and covers the operational needs for a tool useful in training decision-making. More precisely, this heuristic is both robust and, considering

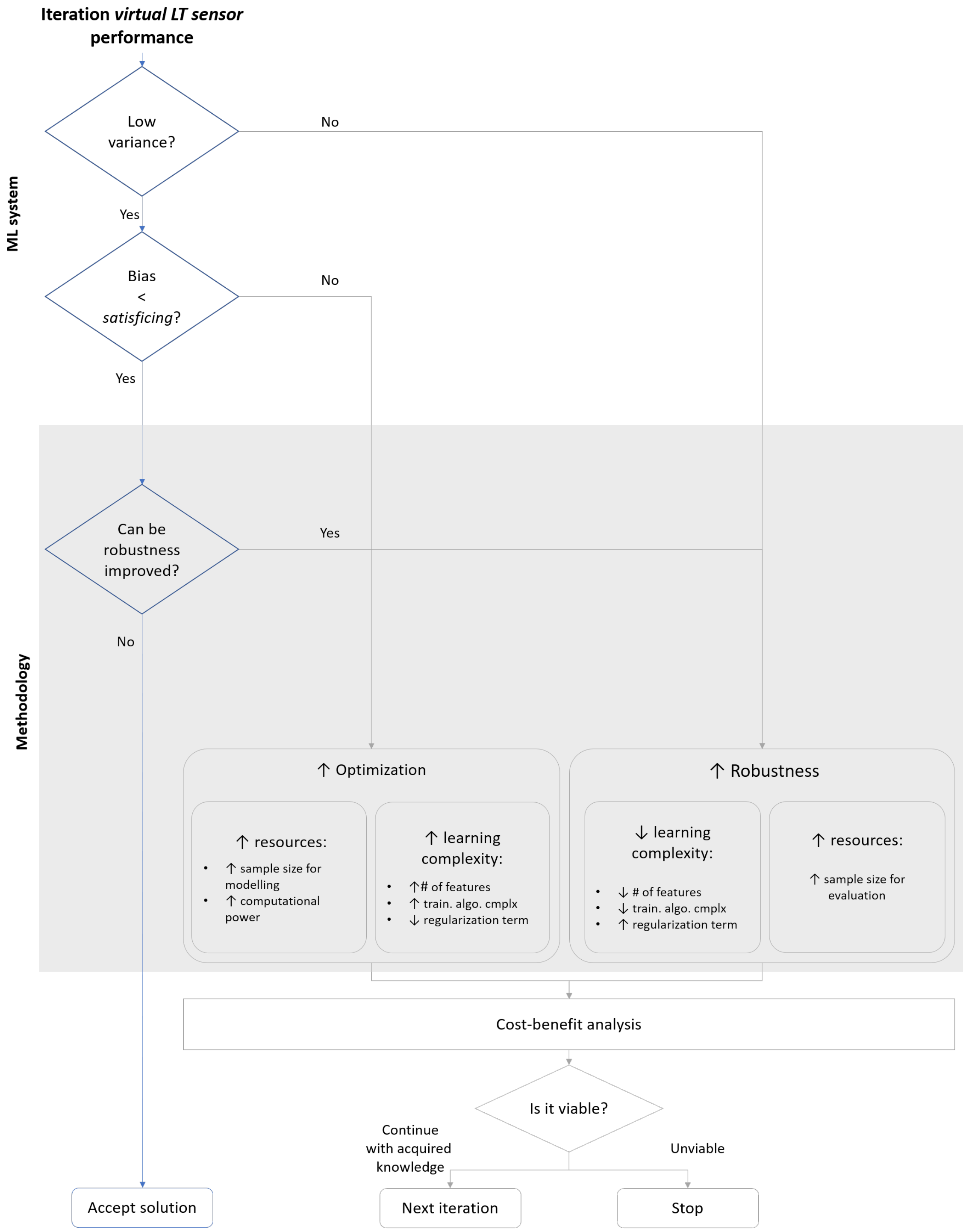


Figure 4.38: Iteration 2 next step decision making process

the *Dmax LT intrinsic error* 4%, is capable of successfully estimating the lactate threshold for 92.5% of the study population, which is within the range of the acceptable error that we defined as *satisficing*.

4.3 *Virtual lactate threshold sensor* design conclusions

In this chapter, we designed a *virtual LT sensor* that, according to the strategical and methodological criteria described in chapter 3, is able to robustly estimate the LT using Vpeak as an operational feature.

From the methodological perspective, the design consisted of two iterations. In the first iteration, a RNN was used as the ML architecture which, combined with the a rigorous data collection, a standardization of the temporal axis, a combination of two database splitting methods a robust *initial virtual LT sensor was created.* Then, further room for robustness improvement was detected and a second iteration was made in this pursue. Inspired by the relation between Vpeak and LT, the mean of the standardized LT was used to create a extremely robust *calibrated virtual LT sensor*. In both cases, the previously proposed Dmax LT intrinsic error analysis showed its importance to understand the variability of the LT with respect to the inputs. This is considered an important methodological conclusion of this chapter that demonstrates how important is to know the measurement errors.

From the sport science perspective, we have set the boundaries of the *Dmax LT method precision error* and shown that other LT protocols could also be evaluated from this perspective in order to quantitatively address their reliability. This may be very helpful to make an objective assessment of how accurate the different LT protocols are and make a comparison between them. Additionally, we got additional knowledge about the *problem complexity* of creating an operational *virtual LT sensor* and also threw some light to its relation to other features. Among them, the association network between %HRmax, LT, Vpeak, IAAF score and 10K race times is seen as interesting to explore in future works.

From the application perspective, we have shown that a simple heuristic (60% of *endurance speed reserve*) is capable of providing an estimation as good as the commonly used Dmax LT protocol for the target recreational runner population. Unlike the Dmax LT protocol, this heuristic is an operational solution that facilitates its consistent use as it relies solely on the athlete's Vpeak: an easily measurable, non-invasive and robust feature that is well established for performance evaluation [94; 95; 82].

One of the possible limitations of this heuristic is that our population is drawn from local running clubs. This means that it is possible that the target recreational runner population here characterized may not be representative of recreational runners of other culture, ethnicity or different contexts. However, one of the main advantages of providing a simple solution is that, unlike other black-box models, it is easily reproducible and adjustable, meaning that we have set a common ground for other researchers to evaluate the impact of our proposal. In the best case scenario, future experiments done in other contexts will validate that we have been capable of discovering a common characteristic of recreational runner population. In the worst case scenario, we have provided an easy to follow methodology (see Chapter 3) and an strong prior that will allow to adjust the estimator according to individual characteristics of different populations.

Chapter 5

Implementation of the *virtual lactate threshold sensor* and the extracted knowledge to help recreational runners' trainings

Knowing how people will use something is essential - Donald Norman

Figure 5.1: Plastic bags out-performing gore-tex in the snow

Chapter 6

Conclusions and future work

Every limit is a beginning as well as an ending - George Eliot, Middlemarch

Figure 6.1: Delicious chocolate tasting in the top

6.1 Research conclusions

Nowadays, there is a huge recreational runner population that wants to train for performance. These athletes have particular interest in assessing the evolution of their performance to help improve their training. In this regard, features related to the intensity of exercise at aerobic/anaerobic transition are good indicators of performance in endurance sports. Particularly, lactate threshold is probably the most used one with this purpose. In fact, current recreational runners, despite their limited resources, pay a reasonable amount of money to estimate their LT in specialised centres.

Many efforts have been made so far in order to address the interest in having operational ways to get information for training decision-making. This thesis is one of them, with the main focus on providing an operational LT estimation. From the work done the following mayor conclusions are derived:

1. The analysis of the state-of-the-art made clear that multiple efforts have been made to determine the LT. Among them, some methods mainly pay attention to accuracy while the operationality was relegated to a secondary place. From these efforts several well known consolidated LT estimation methods arose.

   The state-of-the-art also showed that, with the consolidated methods as reference, many attempts have been made focusing on improving their operationality. However, none of the proposed alternatives are able to solve the operational problem of the current LT determination methods with regards to the needs of the recreational runner population.

   In any case, all these efforts show that characterizing the LT phenomena is complex and that there is a huge interest in this matter.

2. A supervised learning based *virtual LT sensor* has the potential to provide an operational LT estimation providing information in comparable terms to the consolidated methods. This was precisely the first hypothesis of this thesis that was later confirmed.

3. The individual Dmax LT is a consolidated method that maximizes the chances of creating a robust *virtual LT sensor* and is selected as the reference to label the data.

Additionally, this thesis makes the following methodological, sport science and application contributions.

6.1.1 Strategical & methodological contributions

- A framework to determine the value of the multiple state-of-the-art LT estimation methods was created. To do so, we first established the qualities used to determine the value in the context of training decision making of recreational runners. Additionally, we created a *value map* that served not only to organize all the important LT determination methods in the same place (and have a visual big picture), but also to, in future steps, be able to place our solution and compare with the rest of the proposed approaches. This framework allowed to make a deep analysis of first, the consolidated LT determination methods available nowadays and second, the attempts that have been made to improve the operationality of these approaches.

In a more general matter, the criteria for value estimation and the *value map* sets a framework that can be easily generalized to any other problem of information integration into training decision making and is an interesting contribution of this thesis. Moreover, this concept can be easily extended to other research areas where a trade-off between some sort of accuracy and operationality is required and a visual tool may provide a bird's-eye view that can help throw some clarity to the problem-space.

- A strategy was created to help properly pose and apply ML to complex phenomena. This strategy combines the identification of the inherent difficulties that creating a supervised learning based *virtual sensor* has, with the particularities of complex phenomena.

  More precisely, the inherent difficulties of ML complex problems were identified as: (1) the problem boundary discovery and (2) defining the appropriate performance perspective. We proposed an iterative strategy to deal with the former and, for the latter, we set an approach to achieve a *satisficing* accuracy. Despite in ML these iterative approaches are commonly used [57], to the best of our knowledge, they haven't been formalized as done in the present work. This is an important contribution of this thesis which may help better understand the strengths and limitation of ML according to the area or problem in which is to be applied.

- An exhaustive experimental methodology was created and formalized to help maximize the quality of the data collected. To attain so, the experimental methodology was divided in five steps.

  First, select the candidate features according to expert knowledge. Despite all the candidate features were certainly not to be used in this work, possessing multiple relevant features allows to have enough alternatives if new or complementary research paths were discovered both as part of this thesis or for future work. Second, the protocols to collect both the static and time-series data were defined. Third, due to the heterogeneity of the target recreational runner population, we defined the term of 'what a recreational runner is' in this work. From this definition several pre-requisites for the target population were born. Fourth, additional athletic, health, legal and ethical requisites were defined to either protect the athlete and/or ensure that the experiments are performed under the appropriate conditions. Finally, to minimize the inevitable collection of invalid experiments, the validity of the observation was defined according to correctness and/or application type. More precisely, the validity of the experiments was defined according to two purposes: validity for *satisficing* error calculation and validity for *virtual LT sensor* design.

- A precise methodology was defined to design the *virtual LT sensor*. This methodology formalized the common traits that are found in supervised learning and applies it to the *virtual LT sensor*, detailing the steps to be followed in each design iteration. Three steps are considered: context characterization, which deals with ensuring that the quality of the collected data is maximized; content representation, dealing with the approach for learning only the relevant information; and next step selection, which guides the decision making process for the next iteration. Here it is important to note that, despite this traditional next step decision making process is well known in practice [57], to the best of our knowledge, the formalization done in this work is a contribution. Moreover, in this work, we go beyond evaluating the final ML system and introduce an additional methodological perspective to the traditional next step decision making process from the robustness perspective. This methodology may allow to apply virtual sensing concept to solve problems related to sports so it could be extended to other future demands of this area.

- A computational algorithm was proposed to make a Dmax LT intrinsic error analysis. Apart from the sport science implications described afterwards, the application of this algorithm showed the importance of understanding the variability of the output features with respect to the inputs. This algorithm helped to discover the inherent and irreducible noise that the input and output features may contain. Moreover, it also showed that this error may have a major impact in the conclusions that are derived about the relevant information that a certain feature may contain. So, an important methodological conclusion is that a proper feature error analysis has mayor importance for supervised learning systems.

- An *initial virtual LT sensor* was created based on a previously consolidated ML architecture (LRNN) in the first design iteration. To do so, the previously established methodology was developed in detail using several ad hoc applied methods. A web page was created to improve the sampling diversity and quality. Additionally, in order to homogenize the output feature, a standardization of the temporal axis was used. Furthermore, a combination of two database splitting methods (knowledge based and a novel modification of the stratified sampling method) were used to explore and achieve the right diversity in both data sets. Regardless whether this first iteration achieved the desired *virtual LT sensor*, a robust methodological conclusion of this iteration is that there was room for creating a supervised ML system to estimate LT if proper methodologies are followed.

- A *calibrated virtual LT sensor* was created based on a simple heuristic that estimated the LT using the 60% of the *endurance speed reserve* as estimand in the second design iteration.

  Following from the *initial virtual LT sensor* created in the first iteration, a more robust system was created by focusing on a simpler approach. In the first iteration, it was discovered that the standardization of the temporal axis was able to homogenize the LT of different level athletes, and that it is so mostly because of the relation that Vpeak had with the LT. This second approach brought this relationship to the center and exploited it to create a extremely robust *calibrated virtual LT sensor*. To do so, the mean of the standardized LT was proposed as the simplest possible estimator. A methodological conclusion of this iteration was that, as it has been done in this work, using the intrinsic errors of the features as reference to compare with the final estimation errors may be of major importance for proper evaluation of the accuracy of a supervised learning system.

- An implementation of the knowledge of this work was made according to a broad understanding of the concept of operationality. This means that the whole process of gathering information and integration into the training decision-making process is considered. This understanding of operationallity sets a holistic way of creating tools that may help to further close the gap between theoretical contributions and their application in the real world and may be interesting for future works.

6.1.2 Contributions to sport science & physiology

- The design iterations showed that the transformation of LT by means of the Vpeak diminishes most of the variability of individual Dmax LT.

- The analysis of the association between the input and output features showed that there is a strong relation between %HRmax, LT and performance related features. In particular, an interesting association network was found around %HRmax at various stages, LT, Vpeak, IAAF score and 10K race times. This opened the opportunity to further evaluate other features as proxies of performance as it is done with LT.

- The application of the computational algorithm for precision error analysis of the LT is an important contribution to sport science. This analysis is the first quantitative analysis of the precision error of the Dmax LT protocol in the literature, and was performed for different number of blood lactate measurements (5,6,7,8,9 and 10). More importantly, we have provided a computational method (Chapter 3 algorithm 1) that would be easily applicable to calculate the precision error of Dmax protocol with other parameters (regression function, number of points, initial speed etc.). It must be also noted that this computational method, with the appropriate adjustments, may be also useful to estimate the precision error of other LT estimation protocols. This would enable to make quantitative comparisons between protocols, something that, to the best of our knowledge, is not well addressed in the literature.

- We presented the idea that a higher accuracy of the *virtual LT sensor* is unnecessary and even to some extent, non-characterizable and irreducible. First, and according to the calculated LT precision error, a higher accuracy may not be achievable. Additionally, from the perspective of integration of LT in training decision making, we also presented the idea that there is an additional unknown error that arises from the application of different LT intensity indicators (speed or HR) to the real activity. Moreover, this error depends on individual characteristics that vary in a daily basis and that translates differently depending on the sport in which is to be applied. Therefore, the combination of errors may make the pursue of a higher accuracy LT irrelevant once it is applied.

6.1.3 Application contributions

- An operational *virtual LT sensor* is presented in this work for recreational runners. We showed that a heuristic (%60 of *endurance running speed reserve*) fulfills this objective as it is as reliable as the Dmax LT protocol and covers the operational needs for a tool useful in training decision-making. More precisely, this heuristic is both robust and, considering the *Dmax LT intrinsic error* 4%, is capable of successfully estimating the lactate threshold for 92.5% of the study population, which is within the range of the acceptable error that we defined as *satisficing*. Unlike the Dmax LT protocol, this heuristic is an operational solution that facilitates its consistent use as it relies solely on the athlete's Vpeak: an easily measurable, non-invasive and robust feature that is well established for performance evaluation.

- From the additional knowledge gained about other physiological features, %HRmax is presented as an operational and sub-maximum indicator to be used in synergy with the LT, so that more robust conclusions about athlete's performance can be obtained.

- The *Lactatus* SW has been created to ease the athletes' LT estimation process and implement the additional information in the training decision-making process of recreational runners. This way, the work of this thesis has been made tangible and widely available and usable to recreational runners and give guidance to integrate it into the training decision making process.

6.2 Results

From the work done in this thesis, the following journal articles have been published:

- U. Etxegarai, E. Portillo, J. Irazusta, A. Arriandiaga, and I. Cabanes. "**Estimation of lactate threshold with machine learning techniques in recreational runners**, **Q1** - *Applied Soft Computing* , vol. 63, pp. 181196, 2018.
  https://doi.org/10.1016/j.asoc.2017.11.036 [95]

- U. Etxegarai, A. Insunza, J. Larruskain, J. Santos-Concejero, S. M. Gil, E. Portillo, and J. Irazusta. **Prediction of performance by heart rate-derived parameters in recreational run-ners**, **Q1** - *Journal of Sports Sciences*, vol. 00, no. 00, pp. 19, 2018.
  https://doi.org/10.1080/02640414.2018.1442185 [82]

- U. Etxegarai, E. Portillo, J. Irazusta, L. Koefoed, and N. Kasabov. **A heuristic approach for lactate threshold estimation for training decision-making: An accessible and easy to use solution for recreational runners**, **Q1** - *European Journal of Operational Research*, 2019.
  https://doi.org/10.1016/j.ejor.2019.08.023 [103]

Additionally, the following conference papers and presentations have been made:

- **Estimation of lactate threshold using machine learning techniques** - *22nd annual Congress of the European College of Sport Science* (paper presentation)

- **An accessible lactate threshold assessment tool to support endurance athletes trainings** - *29th European Conference on Operational Research (EURO2018)* (paper presentation)

- **Un método inteligente para estimar el umbral de lactato de atletas recreacionales de manera accesible y no invasiva** - *XXXIX Jornadas de Automática* (conference paper and poster presentation)

- **Towards an adaptive lactate threshold estimation methodology: A personalized modelling approach** - *XII World Congress of Performance Analysis of Sport* (poster presentation)

- **Aplicación de técnicas de agrupamiento a corredores de resistencia para la estimación del umbral de lactato** - *XL Jornadas de Automática* (conference paper) - **Werium award to the best bio-engineering work**

The knowledge obtained in this thesis also derived in a collaboration in other application areas:

- **A Question of Trust: Statistical Characterization of Long-Term Traffic Estimations for their Improved Actionability** - *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* [104]

Finally, this work also resulted in a research stay of 6 months in the Knowledge Engineering and Discovery Research Institute, Auckland University of Technology (AUT), New Zealand, with a scholarship of the European Commissions Erasmus Mundus Action 2 PANTHER (Pacific Atlantic Net- work for Technical Higher Education and Research).

6.3 Future work

This work has enabled to detect several potentially interesting future work. The most relevant are:

- The main limitation of *virtual LT sensor* may be in its generalization beyond the local recreational runners given the diverse characteristics of the worldwide recreational runners of other culture, ethnicity or different contexts. However, one of the main advantages of providing a simple solution is that, unlike other black-box models, it is easily reproducible and adjustable, meaning that we have set a common ground for other researchers to evaluate the impact of our proposal and transfer it to other populations.

  This is precisely considered as a future interesting research line, where future experiments done in other contexts will further test our proposal in a broader sense. In the best case scenario, these experiments will validate that we have been capable of discovering a common characteristic of recreational runner population. In the worst case scenario, we have provided an easy to follow methodology (see Chapter 3) and an strong prior that will allow to adjust the estimator according to individual characteristics of different populations.

  Additionally, clustering techniques have the potential to group different populations or subpopulation of endurance runners. The application of these techniques to this problem arise as an interesting research path that may help automatize the adjusting process of the *virtual LT sensor*.

- Additionally, the methodology presented in Chapter 3 allows to extend the *virtual sensing* to other LT's that may be more interesting for other uses. This also includes the transfer of the algorithm 1 for the calculation of other LT method's precision error and make a quantitative comparison between them.

- Finally, the %HRmax has shown to be a potential sub-maximal interesting indicator for recreational runners. In this regard, the implementation of the %HRmax into the *Lactatus* SW is foreseeing as an interesting future work. Additionally, a deeper analysis of the relation between the %HRmax and LT may be interesting to find new possible combinations between. For example, the %HRmax (as it is sub-maximal) could be used in a more daily basis to get the underlying interesting information and complement it with more periodical LT estimations using the *virtual LT sensor*.

References

[1] Http://www.kirolprobak.com/competicion_ficha.php?id=1736, "XV Carrera Primavera Donostiarrak UPV-EHU."

[2] Https://www.runningusa.org/, "Running USA - Road Race Trends (accesed 12.12.2018)."

[3] C. Fiuza-Luces, N. Garatachea, N. A. Berger, and A. Lucia, "Exercise is the real polypill," *Physiology*, vol. 28, no. 5, pp. 330–358, 2013.

[4] A. J. Vickers and E. A. Vertosick, "An empirical study of race times in recreational endurance runners," *BMC Sports Science, Medicine and Rehabilitation*, vol. 8, p. 26, 2016.

[5] J. G. Pallarés, R. Morán-Navarro, J. F. Ortega, V. E. Fernández-Elíaas, and R. Mora-Rodriguez, "Validity and reliability of ventilatory and blood lactate thresholds in well-trained cyclists," *PLoS ONE*, vol. 11, pp. 1–16, 2016.

[6] P. Hofmann and G. Tschakert, "Intensity- and duration-based options to regulate endurance training," *Frontiers in Physiology*, vol. 8, p. 337, may 2017.

[7] K. Tanaka and Y. Matsuura, "Marathon performance, anaerobic threshold, and onset of blood lactate accumulation.," *Journal of applied physiology: respiratory, environmental and exercise physiology*, vol. 57, pp. 640–3, sep 1984.

[8] Lactate.com, "Lactate threshold definitions (accesed 16.09.2019)."

[9] Https://iqo2.zendesk.com/, "Lactate models to determine aerobic threshold (accesed 16.09.2019)."

[10] Https://iqo2.zendesk.com/, "Lactate models to determine anaerobic threshold (accesed 16.09.2019)."

[11] R. K. Binder, M. Wonisch, U. Corra, A. Cohen-Solal, L. Vanhees, H. Saner, and J. P. Schmid, "Methodological approach to the first and second lactate threshold in incremental cardiopulmonary exercise testing," *European Journal of Preventive Cardiology*, vol. 15, no. 6, pp. 726–734, 2008.

[12] R. M. Nicholson and G. G. Sleivert, "Indices of lactate threshold and their relationship with 10-km running velocity.," *Medicine and science in sports and exercise*, vol. 33, pp. 339–342, feb 2001.

[13] E. O. Acevedo and A. H. Goldfarb, “Increased training intensity effects on plasma lactate, ventilatory threshold, and endurance,” *Med.Sci.Sports Exerc.*, vol. 21, pp. 563–568, oct 1989.

[14] V. L. Billat, P. Sirvent, G. Py, J.-P. Koralsztein, and J. Mercier, “The Concept of Maximal Lactate Steady State,” *Sports Medicine*, vol. 33, no. 6, pp. 407–426, 2003.

[15] J. Santos-Concejero, C. Granados, J. Irazusta, I. Bidaurrazaga-Letona, J. Zabala-Lili, N. Tam, and S. M. Gil, “OBLA is a better predictor of performance than Dmax in long and middle-distance well-trained runners,” *J Sports Med Phys Fittness*, vol. 54, no. 5, pp. 553–557, 2014.

[16] J. R. Lacour, S. Padilla-Magunacelaya, J. C. Barthélémy, and D. Dormois, “The energetics of middle-distance running.,” *European journal of applied physiology and occupational physiology*, vol. 60, pp. 38–43, 1990.

[17] H. Heck, A. Mader, G. Hess, S. Mücke, R. Müller, and W. Hollmann, “Justification of the 4-mmol/l lactate threshold.,” *International Journal of Sports Medicine*, vol. 6, pp. 117–130, jun 1985.

[18] O. Storen, J. Helgerud, E. M. Stoa, and J. Hoff, “Maximal Strength Training Improves Running Economy in Distance Runners,” *Medicine and Science in Sports and Exercise*, vol. 40, no. 6, pp. 1087–1092, 2008.

[19] Http://www.wqusability.com/, “Usability (accesed 12.10.2019),” 2001.

[20] A. Mader and H. Heck, “A theory of the metabolic origin of ”anaerobic threshold”.,” *International journal of sports medicine*, vol. 7 Suppl 1, pp. 45–65, jun 1986.

[21] J. L. Chicharro, A. Carvajal, J. Pardo, M. Perez, and A. Lucia, “Physiological parameters determined at OBLA vs. a fixed heart rate of 175 beats x min-1 in an incremental test performed by amateur and professional cyclists.,” 1999.

[22] R. Beneke, “Maximal lactate steady state concentration (MLSS): experimental and modelling approaches,” *Eur J Appl Physiol*, vol. 88, no. 4-5, pp. 361–369, 2003.

[23] J. Jakobsson and C. Malm, “Maximal Lactate Steady State and Lactate Thresholds in the Cross-Country Skiing Sub-Technique Double Poling.,” *International journal of exercise science*, vol. 12, no. 2, pp. 57–68, 2019.

[24] T. R. Figueira, F. Caputo, J. G. Pelarigo, and B. S. Denadai, “Influence of exercise mode and maximal lactate-steady-state concentration on the validity of OBLA to predict maximal lactate-steady-state in active individuals,” *Journal of Science and Medicine in Sport*, vol. 11, no. 3, pp. 280–286, 2008.

[25] R. Beneke, M. Hütler, and R. M. Leithäuser, “Maximal lactate-steady-state independent of performance,” *Medicine and Science in Sports and Exercise*, vol. 32, pp. 1135–1139, jun 2000.

[26] J. Keul, G. Simon, A. Berg, H.-H. Dickhuth, I. Goerttler, and R. Kuebel, "Bestimmung der individuellen anaeroben Schwelle zur Leistungsbewertung und Trainingsgestaltung," *Deutsche Zeitschrift für Sportmedizin*, vol. 30, no. 7, 1979.

[27] H. Stegmann and W. Kindermann, "Comparison of prolonged exercise tests at the individual anaerobic threshold and the fixed anaerobic threshold of 4 mmol.l(-1) lactate.," *International journal of sports medicine*, vol. 3, pp. 105–10, may 1982.

[28] M. Anderson, W. Hopkins, A. Roberts, and D. Pyne, "Ability of test measures to predict competitive performance in elite swimmers.," *Journal of sports sciences*, vol. 26, pp. 123–30, jan 2008.

[29] B. S. Denadai, E. B. G. Gomide, and C. C. Greco, "The relationship between onset of blood lactate accumulation, critical velocity, and maximal lactate steady state in soccer players.," *Journal of strength and conditioning research / National Strength & Conditioning Association*, vol. 19, pp. 364–8, may 2005.

[30] K. M. Gilenstam, K. Thorsen, and K. B. Henriksson-Larsén, "Physiological correlates of skating performance in women's and men's ice hockey.," *Journal of strength and conditioning research / National Strength & Conditioning Association*, vol. 25, pp. 2133–42, aug 2011.

[31] P. Larsson, P. Olofsson, E. Jakobsson, L. Burlin, and K. Henriksson-Larsén, "Physiological predictors of performance in cross-country skiing from treadmill tests in male and female subjects.," *Scandinavian journal of medicine & science in sports*, vol. 12, pp. 347–53, dec 2002.

[32] S. Padilla, I. Mujika, G. Cuesta, and J. J. Goiriena, "Level ground and uphill cycling ability in professional road cycling.," *Medicine and science in sports and exercise*, vol. 31, pp. 878–85, jun 1999.

[33] K. Tanaka, "Lactate-related factors as a critical determinant of endurance.," *The Annals of physiological anthropology = Seiri Jinruigaku Kenkyukai kaishi*, vol. 9, pp. 191–202, apr 1990.

[34] J. Santos-Concejero, R. Tucker, C. Granados, J. Irazusta, I. Bidaurrazaga-Letona, J. Zabala-Lili, and S. M. Gil, "Influence of regression model and initial intensity of an incremental test on the relationship between the lactate threshold estimated by the maximal-deviation method and running performance.," *Journal of Sports Sciences*, vol. 32, pp. 853–859, 2014.

[35] N. R. Borges and M. W. Driller, "Wearable Lactate Threshold Predicting Device is Valid and Reliable in Runners," *Journal of Strength and Conditioning Research*, vol. 30, no. 8, pp. 2212–2218, 2016.

[36] B. Cheng, H. Kuipers, A. C. Snyder, H. A. Keizer, A. Jeukendrup, and M. Hesselink, "A new approach for the determination of ventilatory and lactate thresholds.," *International journal of sports medicine*, vol. 13, pp. 518–22, oct 1992.

[37] A. Zaj, Z. Wa, and K. Miko, "Lactate Threshold (D Max Method) and Maximal Lactate Steady State in Cyclists," *Journal of Human Kinetics*, vol. 21, no. April, pp. 49–56, 2009.

[38] T. Yoshida, "Effect of exercise duration during incremental exercise on the determination of anaerobic threshold and the onset of blood lactate accumulation," *European Journal of Applied Physiology and Occupational Physiology*, vol. 53, no. 3, pp. 196–199, 1984.

[39] F. A. Machado, F. Y. Nakamura, and S. M. F. D. Moraes, "Influence of regression model and incremental test protocol on the relationship between lactate threshold using the maximal-deviation method and performance in female runners.," *Journal of sports sciences*, vol. 30, no. 12, pp. 1267–74, 2012.

[40] C. Papadopoulos, J. A. Doyle, and B. D. LaBudde, "Relationship between running velocity of 2 distances and various lactate parameters.," *International journal of sports physiology and performance*, vol. 1, pp. 270–283, sep 2006.

[41] F. A. MacHado, S. M. F. De Moraes, C. S. Peserico, P. V. Mezzaroba, and W. P. Higino, "The Dmax is highly related to performance in middle-aged females," *International Journal of Sports Medicine*, vol. 32, pp. 672–676, sep 2011.

[42] D. R. Seshadri, R. T. Li, J. E. Voos, J. R. Rowbottom, C. M. Alfes, C. A. Zorman, and C. K. Drummond, "Wearable sensors for monitoring the physiological and biochemical profile of the athlete," *npj Digital Medicine*, vol. 2, no. 1, 2019.

[43] P. J. Derbyshire, H. Barr, F. Davis, and S. P. J. Higson, "Lactate in human sweat: A critical review of research to the present day," *Journal of Physiological Sciences*, vol. 62, no. 6, pp. 429–440, 2012.

[44] C. Abraham, Y. Beiderman, N. Ozana, F. Tenner, M. Schmidt, M. Sanz, J. Garcia, and Z. Zalevsky, "Photonic non-contact estimation of blood lactate level," *Biomedical Optics Express*, vol. 6, pp. 130–135, oct 2015.

[45] B. Jotta, A. B. Coutinho, A. V. Pino, and M. N. Souza, "Lactate threshold by muscle electrical impedance in professional rowers," *Review of Scientific Instruments*, vol. 88, p. 045105, apr 2017.

[46] J. Bourgois, P. Coorevits, L. Danneels, E. Witvrouw, D. Cambier, and J. Vrijens, "Validity of the heart rate deflection point as a predictor of lactate threshold concepts during cycling," *Journal of Strength and Conditioning Research*, vol. 18, pp. 498–503, jul 2004.

[47] N. R. Borges and M. W. Driller, "Wearable lactate threshold predicting device is valid and reliable in runners," *Journal of Strength and Conditioning Research*, vol. 30, no. 8, pp. 2212–2218, 2016.

[48] F. Conconi, M. Ferrari, P. G. Ziglio, P. Droghetti, and L. Codeca, "Determination of the anaerobic threshold by a noninvasive field test in runners," *Journal of Applied Physiology*, vol. 52, pp. 869–873, apr 1982.

[49] J. Bourgois, P. Coorevits, L. Danneels, E. Witvrouw, D. Cambier, and J. Vrijens, "Validity of the heart rate deflection point as a predictor of lactate threshold concepts during cycling.," *Journal of strength and conditioning research*, vol. 18, pp. 498–503, aug 2004.

[50] J. V. Ringwood, J. O. Neill, P. Tallon, N. Fleming, and B. Donne, "Non-invasive anaerobic threshold measurement using fuzzy model interpolation," *IEEE Conference on Control Applications (CCA)*, pp. 1711–1715, 2014.

[51] G. Borg, "Psychophysical bases of perceived exertion.," *Medicine and science in sports and exercise*, vol. 14, no. 5, pp. 377–381, 1982.

[52] J. Lopez Chicharro, S. Aznar Laín, A. Fernández Vaquero, L. M. López Mojares, A. Lucía Mulas, and M. Pérez Ruiz, *Transicion aerobica-anaerobica: concepto, metodologia de determinacion y aplicaciones*. Master Line & Prodigio, 1st ed., 2004.

[53] A. P. Proshin and Y. V. Solodyannikov, "Mathematical Modeling of Lactate Metabolism with Applications to Sports," *Automation and Remote Control*, vol. 74, pp. 1004–1019, jun 2013.

[54] A. Erdogan, C. Cetin, H. Goksu, R. Guner, and M. L. Baydar, "Non-invasive detection of the anaerobic threshold by a neural network model of the heart ratework rate relationship," *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*, vol. 223, pp. 109–115, sep 2009.

[55] S. C. Huang, R. Casaburi, M. F. Liao, K. C. Liu, Y. J. Chen, T. C. Fu, and H. R. Su, "Noninvasive prediction of Blood Lactate through a machine learning-based approach," *Scientific Reports*, vol. 9, no. 1, pp. 1–7, 2019.

[56] A. Arriandiaga, E. Portillo, J. A. Sánchez, I. Cabanes, and A. Zubizarreta, "Downsizing training data with weighted FCM for predicting the evolution of specific grinding energy with RNNs," *Applied Soft Computing Journal*, vol. 61, pp. 211–221, 2017.

[57] A. Ng, *Machine Learning Yearning*. 2018.

[58] Y. Bar-yam, "When Systems Engineering Fails — Toward Complex Systems Engineering," *Exchange*, vol. 16, pp. 1990—-1993, 2003.

[59] S. N. Staff, "From AI to protein folding: Our Breakthrough runners-up," 2016.

[60] W. Rawat, "Deep Convolutional Neural Networks for Image Classification : A Comprehensive Review," *Neural Computation*, vol. 2449, pp. 2352–2449, 2017.

[61] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, may 2015.

[62] H. A. Simon, "Rational Choice and the Structure of the Environment," *Psychological Review*, vol. 63, no. 2, pp. 129–138, 1956.

[63] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H. T. Lin, *Learning from Data: A Short Course*. AMLBook.com, 2012.

[64] G. C. Cawley and N. L. C. Talbot, "On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation," *Journal of Machine Learning Research*, vol. 11, p. 20792107, 2010.

[65] R. C. R. Davison, K. A. van Someren, and A. M. Jones, "Physiological monitoring of the olympic athlete," *Journal of Sports Sciences*, vol. 27, no. 13, pp. 1433–1442, 2009.

[66] T. Reilly, T. Morris, and G. Whyte, "The specificity of training prescription and physiological assessment: A review," *Journal of Sports Sciences*, vol. 27, no. 6, pp. 575–589, 2009.

[67] H. a. M. Daanen, R. P. Lamberts, V. L. Kallen, A. Jin, and N. L. U. Van Meeteren, "A systematic review on heart-rate recovery to monitor changes in training status in athletes," *International Journal of Sports Physiology and Performance*, vol. 7, no. 3, pp. 251–260, 2012.

[68] K. Meyer, "Ventilatory and lactate threshold determinations in healthy normals and cardiac patients: Methodological problems," *European Journal of Applied Physiology and Occupational Physiology*, vol. 72, no. 5-6, pp. 387–393, 1996.

[69] R. K. Tanner, K. L. Fuller, and M. L. R. Ross, "Evaluation of three portable blood lactate analysers: Lactate Pro, Lactate Scout and Lactate Plus," *European Journal of Applied Physiology*, vol. 109, pp. 551–559, jun 2010.

[70] I. Standard, "Iso 5725-5 Accuracy (trueness and precision) of measurement methods and results," 1998.

[71] A. Zimek, E. Schubert, and H.-P. Kriegel, "A Survey on Unsupervised Outlier Detection in High-Dimensional Numerical Data," *Science And Technology*, vol. 4, no. 5, pp. 497–511, 2010.

[72] Https://www.encyclopediaofmath.org/index.php/Law_of_large_numbers, "Law of large numbers."

[73] Https://www.encyclopediaofmath.org/index.php/Central_limit_theorem, "Central Limit Theorem."

[74] E. W. Steyerberg, M. J. Eijkemans, F. E. Harrell, and J. D. F. Habbema, "Prognostic modeling with logistic regression analysis: In search of a sensible strategy in small data sets," *Medical Decision Making*, vol. 21, no. 1, pp. 45–56, 2001.

[75] Z. Gong, P. Zhong, and W. Hu, "Diversity in Machine Learning," *IEEE Access*, vol. 7, pp. 64323–64350, 2019.

[76] J. Gareth, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. Springer, 2013.

[77] J. Heaton, "An Empirical Analysis of Feature Engineering for Predictive Modeling," in *SoutheastCon 2016*, pp. 0–5, 2016.

[78] V. Kumar, "Feature Selection: A literature Review," *The Smart Computing Review*, vol. 4, no. 3, 2014.

[79] J. F. Moxnes and Ø. Sandbakk, "The kinetics of lactate production and removal during whole-body exercise," 2012.

[80] H. Asgari, X. Chen, M. Morini, M. Pinelli, R. Sainudiin, P. R. Spina, and M. Venturini, "NARX models for simulation of the start-up operation of a single-shaft gas turbine," *Applied Thermal Engineering*, vol. 93, pp. 368–376, 2016.

[81] Http://lactatus.com/, "Lactatus web page."

[82] U. Etxegarai, A. Insunza, J. Larruskain, J. Santos-Concejero, S. M. Gil, E. Portillo, and J. Irazusta, "Prediction of performance by heart rate-derived parameters in recreational runners," *Journal of Sports Sciences*, vol. 00, no. 00, pp. 1–9, 2018.

[83] W. K. Allen, D. R. Seals, B. F. Hurley, A. A. Ehsani, and J. M. Hagberg, "Lactate threshold and distance-running performance in young and older endurance athletes," *Journal of Applied Physiology*, vol. 58, no. 4, pp. 1281–1284, 1985.

[84] J. A. Heuberger, P. Gal, F. E. Stuurman, W. A. De Muinck Keizer, Y. M. Miranda, and A. F. Cohen, "Repeatability and predictive value of lactate threshold concepts in endurance sports," *PLoS ONE*, vol. 13, p. e0206846, nov 2018.

[85] S. Makridakis, M. Hibon, and C. Moser, "Accuracy of Forecasting: An Empirical Investigation," *Journal of the Royal Statistical Society. Series A (General)*, vol. 142, no. 2, p. 97, 1979.

[86] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," *PLoS ONE*, vol. 13, no. 3, pp. 1–26, 2018.

[87] D. J. C. MacKay, "A Practical Bayesian Framework for Backpropagation Networks," *Neural Computation*, vol. 4, no. 3, pp. 448–472, 1992.

[88] A. Arriandiaga, E. Portillo, J. a. Sánchez, I. Cabanes, and I. Pombo, "A new approach for dynamic modelling of energy consumption in the grinding process using recurrent neural networks," *Neural Computing and Applications*, 2015.

[89] Https://es.mathworks.com/help/nnet/ref/layrecnet.html, "Matlab NN- Layer recurrent neural network architecture (accesed 26.04.2017)."

[90] M. T. Hagan and M. B. Menhaj, "Training Feedforward Networks with the Marquardt Algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[91] D. Mackay, "Probable networks and plausible predictions a review of practical Bayesian methods for supervised neural networks," *Network: Computation in Neural Systems*, vol. 6, pp. 469–505, 1995.

[92] D. Nguyen and B. Widrow, “Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights,” *IJCNN Int. Joint Conf. Neural Networks*, vol. 13, p. C21, 1990.

[93] M. R. Chernick and R. A. LaBudde, *An introduction to bootstrap methods with applications to R*. Wiley, 2011.

[94] T. D. Noakes, K. H. Myburgh, and R. Schall, “Peak treadmill running velocity during the VO2 max test predicts running performance.,” *Journal of sports sciences*, vol. 8, pp. 35–45, mar 1990.

[95] U. Etxegarai, E. Portillo, J. Irazusta, A. Arriandiaga, and I. Cabanes, “Estimation of lactate threshold with machine learning techniques in recreational runners,” *Applied Soft Computing*, vol. 63, pp. 181–196, 2018.

[96] Http://lactatus.com/guia/#/, “Lactatus software.”

[97] H. Tanaka, K. D. Monahan, and D. R. Seals, “Age-predicted maximal heart rate revisited,” *Journal of the American College of Cardiology*, vol. 37, pp. 153–156, 2001.

[98] M. Buchheit, M. B. Simpson, H. Al Haddad, P. C. Bourdon, and A. Mendez-Villanueva, “Monitoring changes in physical performance with heart rate measures in young soccer players,” *European Journal of Applied Physiology*, vol. 112, pp. 711–723, 2012.

[99] A. Legaz-Arrese, H. Kinfu, D. Munguía-Izquierdo, L. E. Carranza-Garcia, and F. J. Calderón, “Basic physiological measures determine fitness and are associated with running performance in elite young male and female Ethiopian runners,” *Journal of Sports Medicine and Physical Fitness*, vol. 49, pp. 358–363, dec 2009.

[100] V. Vesterinen, A. Nummela, S. Äyrämö, T. Laine, E. Hynynen, J. Mikkola, and K. Häkkinen, “Monitoring training adaptation with a submaximal running test under field conditions,” *International Journal of Sports Physiology and Performance*, vol. 11, pp. 393–399, 2016.

[101] A. Lucía, J. Hoyos, M. Pérez, and J. L. Chicharro, “Heart rate and performance parameters in elite cyclists: a longitudinal study.,” *Medicine and science in sports and exercise*, vol. 32, pp. 1777–1782, oct 2000.

[102] R. Cardiorrespiratoria, “Propuesta metodológica para el entrenamiento de la resistencia cardiorrespiratoria.,” *Journal of Sport and Health Research*, vol. 4, no. 2, pp. 119–136, 2012.

[103] U. Etxegarai, E. Portillo, J. Irazusta, L. Koefoed, and N. Kasabov, “A heuristic approach for lactate threshold estimation for training decision-making: An accessible and easy to use solution for recreational runners,” *European Journal of Operational Research*, 2019.

[104] I. Laña, E. Villar-Rodriguez, U. Etxegarai, I. Oregi, and J. Del Ser, “A Question of Trust: Statistical Characterization of Long-Term Traffic Estimations for their Improved Actionability,” *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 1922–1928, 2019.

[105] C. M. Fairman, R. L. Lafountain, A. R. Lucas, and B. C. Focht, “Monitoring resistance exercise intensity using ratings of perceived exertion in previously untrained patients with prostate cancer undergoing androgen deprivation therapy,” *Journal of Strength and Conditioning Research*, vol. 32, no. 5, pp. 1360–1365, 2018.

[106] A. Aliverti, B. Kayser, A. L. O. Mauro, M. Quaranta, P. Pompilio, R. L. Dellacà, J. Ora, L. Biasco, L. Cavalleri, L. Pomidori, A. Cogo, R. Pellegrino, and G. Miserocchi, “Respiratory and leg muscles perceived exertion during exercise at altitude,” *Respiratory Physiology and Neurobiology*, vol. 177, pp. 162–168, jul 2011.

[107] E. Borg, G. Borg, K. Larsson, M. Letzter, and B. M. Sundblad, “An index for breathlessness and leg fatigue,” *Scandinavian Journal of Medicine and Science in Sports*, vol. 20, pp. 644–650, aug 2010.

Appendix A

Experimental tests documentation

HOJA DE INFORMACIÓN

TÍTULO DEL ESTUDIO**: LACTATUS "Sistema avanzado de apoyo al entrenamiento de resistencia basado en un Sensor Virtual de Lactato, obtenida a base de técnicas de Inteligencia Artificial."**
INVESTIGADORA PRINCIPAL: Susana Gil, Departamento de Fisiología Facultad de Medicina y Enfermería, UPV/EHU
Investigadora Principal: Susana Gil Departamento: Fisioterapia. Departamento de Fisiología Centro: Facultad De Medicina y Enfermería. UPV/EHU

Nos dirigimos a usted para informarle sobre un estudio de investigación en el que se le invita a participar. El estudio se realizará de acuerdo con la LEY 14/2007, de 3 de julio, de Investigación biomédica, cumpliendo con todos los criterios éticos, habiendo sido evaluado de modo positivo por el Comité de Ética para la investigación en Seres Humanos de la UPV/EHU.
Nuestra intención es tan solo que usted reciba la información correcta y suficiente para que pueda evaluar y juzgar si quiere o no participar en este estudio. Para ello lea esta hoja informativa con atención y nosotros le aclararemos las dudas que le puedan surgir después de la explicación. Además, puede consultar con las personas que considere oportuno.

PARTICIPACIÓN VOLUNTARIA

Debe saber que su participación en este estudio es voluntaria y que puede decidir no participar o cambiar su decisión y retirar el consentimiento en cualquier momento del estudio, sin que por ello se deriven consecuencias negativas para usted, ni se produzcan represalias directas o indirectas por su decisión.

DESCRIPCIÓN GENERAL DEL ESTUDIO

El lactato sanguíneo es un parámetro muy utilizado para estimar la forma física y el nivel de entrenamiento en los deportistas. Actualmente la determinación del lactato sanguíneo debe realizarse por medio de una punción y posterior medición del mismo.
El objetivo del presente proyecto es desarrollar, mediante técnicas de Inteligencia Artificial, un sistema que permita estimar la curva de lactato de un individuo a partir de la medida de sus pulsaciones y otras variables y parámetros del individuo sin necesidad de tomar muestras de sangre ni de utilizar dispositivos adicionales sobre el cuerpo. Para ello se realizarán numerosas pruebas de esfuerzo a triatletas y corredores en nuestro laboratorio con el objeto de obtener parámetros que puedan ser utilizados en el citado sistema de Inteligencia Artificial.

¿En qué consiste su participación?

- En contestar a un cuestionario sobre los datos personales: fecha de nacimiento, género, contestar un cuestionario sobre el ejercicio físico y los entrenamientos que realiza habitualmente, así como información acerca de las mejores marcas.
- En permitir que le realicen diversas mediciones: Talla y peso.
- Realizar una prueba incremental máxima en un tapiz rodante comenzando a 9km/h e incrementando 1.5km/h cada 4 minutos, hasta 13,5 km/h; posteriormente la velocidad incrementará en 1km/h. Habrá una recuperación entre los escalones de 1 minuto. La prueba continuará hasta que usted lo desee o cuando no sea capaz de mantener la velocidad necesaria. La prueba será realizada por una persona licenciada en medicina con experiencia en este tipo de pruebas. Los participantes estarán asegurados por medio de un arnés durante la prueba en el tapiz rodante. En dicha prueba se medirán:
    - mediante una pequeña punción en la oreja durante el minuto de recuperación se extraerá una gota de sangre para medir el lactato
    - Medición de la frecuencia cardiaca por medio de un pulsómetro durante la prueba
    - El peso corporal previo y posterior a la prueba de esfuerzo
    - En cada descanso se le preguntará por el esfuerzo percibido que lo deberá indicar en una escala de 0 a 10.
- Participar en la carrera de 10km que el Club Deportivo Donostiarrak organiza en Donostia el 19/03/2017 a las 10:00 AM.

Las participantes deberán vestir pantalón corto y camiseta, así como las zapatillas habituales del entrenamiento. Las mediciones tendrán una duración aproximada de una hora, y se realizarán en la Facultad de Medicina y Odontología, Campus de Leioa, UPV/EHU.

BENEFICIOS Y RIESGOS DERIVADOS DE SU PARTICIPACIÓN EN EL ESTUDIO

Beneficios:
- A cada sujeto se le proporcionará un informe con los resultados individuales obtenidos en su prueba.

Riesgos:
- Las pruebas físicas pueden tener el riesgo que tiene cualquier entrenamiento intenso o competición. En casos muy raros, pueden presentarse desmayos e incluso problemas cardiacos. Por ello, se solicitará la aptitud para su deporte previo a la realización de la prueba. Además se hará un control continuo durante las pruebas por parte de un médico especializado en pruebas de esfuerzo. En el caso de presentarse cualquier signo adverso la prueba de esfuerzo será detenida inmediatamente.

- Las tomas de lactado serán realizadas por el médico deportivo, intentando hacer la menor invasión posible, es decir, un solo pinchazo en uno de los dos lóbulos de la oreja. En algunas ocasiones, se deberá proceder a hacer un segundo pinchazo debido a la vasoconstricción de los capilares de la oreja, si se diese el caso, se intentará realizar este segundo sobre el primero ya realizado. Puede aparecer un pequeño hematoma en los días posteriores que desaparece espontáneamente.

CONTRAINDICACIONES

Los participantes deberán haber superado un reconocimiento médico-deportivo durante el último año. Será necesario aportar el justificante del mismo.

CONFIDENCIALIDAD

El tratamiento, la comunicación y la cesión de los datos de carácter personal de todos los participantes se ajustará a lo dispuesto en la Ley Orgánica 15/1999, de 13 de diciembre de protección de datos de carácter personal. De acuerdo a lo que establece la legislación mencionada, usted puede ejercer los derechos de acceso, modificación, oposición y cancelación de datos, para lo cual deberá dirigirse a su investigador de referencia.

Los datos recogidos para el estudio estarán identificados mediante un código y solo la investigadora principal del estudio podrá relacionar dichos datos con usted y con sus datos personales. Por lo tanto, su identidad no será revelada a persona alguna salvo excepciones, en caso de urgencia médica o requerimiento legal. El fichero de datos de los participantes ha sido dado de alta en un fichero del tipo de “Investigación de nivel alto” de la UPV/EHU con el nombre INA-LACTATUS, código: 2080310015-INA0118.

Podrá consultar en cualquier momento los datos que ha facilitado o solicitar que rectifique o cancele mis datos o simplemente que no los utilicen para algún fin concreto de esta investigación. La manera de hacerlo es dirigiéndose al Responsable de Seguridad LOPD de la UPV/EHU, Rectorado, Barrio Sarriena, s/n, 48940-Leioa-Bizkaia.

Los datos obtenidos serán tratados en ordenadores de la UPV/EHU previa disociación de los datos personales, y el acceso a su información personal quedará restringido únicamente a la investigadora principal del proyecto cuando lo precise para comprobar los datos y procedimientos del estudio, pero siempre manteniendo la confidencialidad de los mismos de acuerdo a la legislación vigente. Una vez finalizado el estudio, cuya duración se prevé de un año, los datos personales serán guardados durante 5 años.

COMPENSACIÓN ECONÓMICA

Su participación en el estudio no le supondrá ningún gasto, ni compensación económica alguna. Los costes de la inscripción a la carrera de 10km que el Club Deportivo Donostiarrak organiza en Donostia el 19/03/2017 a las 10:00 AM serán financiados por este estudio. A su vez, se facilitará el transporte a dicha prueba poniendo a disposición de los atletas un autobús que les llevará y traerá de vuelta.

OTRA INFORMACIÓN RELEVANTE

Cualquier nueva información referente al estudio que se descubra durante su participación y que pueda afectar a su disposición a participar en el mismo, le será comunicada por su investigadora de referencia (Dra. Susana Gil Orozko) lo antes posible y personalmente.

Si usted decide retirar el consentimiento para participar en este estudio, ningún dato nuevo será añadido a la base de datos y puede exigir la destrucción de todas las muestras identificables previamente retenidas para evitar la realización de nuevos análisis.

También debe saber que puede ser excluido del estudio si los investigadores del estudio lo consideran oportuno, ya sea por motivos de seguridad, por cualquier acontecimiento adverso que se produzca o porque consideren que no está cumpliendo con los procedimientos establecidos. En cualquiera de los casos, usted recibirá una explicación adecuada del motivo que ha ocasionado su retirada del estudio.

Al firmar la hoja de consentimiento adjunta, se compromete a cumplir con los procedimientos del estudio que se le han expuesto.

**En caso de necesitar más información o tener alguna duda póngase en contacto con la investigadora responsable Susana Gil, tel. 94 601 2958, e-mail: Susana.gil@ehu.eus

Vicerrectorado de Investigación	CUEID/CEISH

CONSENTIMIENTO

TÍTULO DEL ESTUDIO: **LACTATUS "Sistema avanzado de apoyo al entrenamiento de resistencia basado en un Sensor Virtual de Lactato, obtenida a base de técnicas de Inteligencia Artificial."**	
INVESTIGADORA PRINCIPAL: SUSANA MARIA GIL OROZKO	
Nombre: SUSANA MARIA GIL OROZKO Departamento: FISIOTERAPIA. DEPARTAMENTO DE FISIOLOGIA Centro: FACULTAD DE MEDICINA Y ODONTOLOGIA. UPV/EHU	
Tf: +34 94 601 2859	E-mail: susana.gil@ehu.eus

Yo, D/Dña……….... ...., mayor de edad, y con D.N.I. ....,

DECLARO QUE:

- He leído la hoja de información que se me ha entregado.
- He podido hacer preguntas sobre el estudio.
- He hablado con: Susana Gil Orozko /Ainhoa Insunza
- He recibido suficiente información sobre el estudio.

Resumen del estudio: El lactato sanguíneo es un parámetro muy utilizado para estimar la forma física y el nivel de entrenamiento en los deportistas. Actualmente la determinación del lactato sanguíneo debe realizarse por medio de una punción y posterior medición del mismo. Además, requiere de personal cualificado y de equipamiento específico.

El objetivo del presente proyecto es desarrollar mediante técnicas de inteligencia Artificial un sistema que permita estimar la curva de lactato de un individuo a partir de la medida de sus pulsaciones y otras variables y parámetros del individuo sin necesidad de tomar muestras de sangre ni de utilizar dispositivos adicionales sobre el cuerpo.

Las intervenciones que se me van a realizar son:

o Un cuestionario con preguntas sobre mis datos personales (genero, fecha de nacimiento) y el historial deportivo

o Una encuesta sobre la actividad física y entrenamientos realizados.

o Una antropometría: talla y peso

o Una prueba incremental máxima en un tapiz rodante comenzando a 9km/h e incrementando 1.5km/h cada 4 minutos, hasta 13,5 km/h; posteriormente la velocidad incrementará en 1km/h. Habrá una recuperación entre los escalones de 1 minuto, en el cual se tomará una gota de sangre para medir el lactato por punción en la oreja. Durante la prueba llevaré un arnés de sujeción y estaré conectado a un electrocardiograma.

o Además, se medirán continuamente las pulsaciones por medio de un pulsómetro

o Peso corporal previo y posterior a la prueba

o Indicaré mi percepción del esfuerzo realizado en una escala de 1 a 10 o Participaré en la carrera de 10km que el Club Deportivo Donostiarrak organiza en Donostia el 19/03/2017 a las 10:00 AM.

El equipo investigador cumplirá estrictamente la legislación en materia de protección de datos, en concreto los preceptos de la Ley Orgánica 15/1999, de 13 de diciembre de protección de datos de carácter personal y el Real Decreto 1029/2007 sobre medidas de seguridad.

Los datos personales que nos ha facilitado para este proyecto de investigación serán tratados con absoluta confidencialidad de acuerdo con la Ley de Protección de Datos. Se incluirán en un fichero de la UPV/EHU, código: 2080310015-INA0118, y sólo se utilizarán para los fines del proyecto.

Podré consultar en cualquier momento los datos que he facilitado o solicitar que rectifique o cancele mis datos o simplemente que no los utilicen para algún fin concreto de esta investigación. La manera de hacerlo es dirigiéndome al Responsable de Seguridad LOPD de la UPV/EHU, Rectorado, Barrio Sarriena, s/n, 48940-Leioa-Bizkaia.

Los datos obtenidos serán tratados en ordenadores de la UPV/EHU previa disociación de los datos personales, y el acceso a su información personal quedará restringido únicamente a la investigadora principal del proyecto cuando lo precise para comprobar los datos y procedimientos del estudio, pero siempre manteniendo la confidencialidad de los mismos de acuerdo a la legislación vigente. Una vez finalizado el estudio, cuya duración se prevé de un año, los datos personales serán guardados durante 5 años.

La participación en el estudio no me supondrá ningún gasto.

- Comprendo que la participación en el estudio es voluntaria.
- Comprendo que es posible retirarse del estudio:
    1. En cualquier momento
    2. Sin tener que dar explicaciones.
    3. Sin que esto suponga represalias de ningún tipo.
       Para ello, me podré poner en contacto con la investigadora principal del estudio.
- Participo libremente en el estudio y doy mi consentimiento para el acceso y utilización de sus datos en las condiciones detalladas en la hoja de información.

Y para que así conste firmo el presente documento en ...., a

.... de 2017

Nombre:	Nombre:
Firma de la participante:	Firma del investigador/a:
DNI:	DNI:

Appendix B

Data Collection Protocol

Several devices are used for data collection: a HR monitor, a lactate measurement device for blood lactate measurement, a precision stadiometer for height, a balance (Seca, Bonn, Germany) for weight and a bio-impedance meter for body composition measures.

HR is monitored by a HR monitor (Garmin 910XT, George Town, Caiman Islands) and lactate concentration by a portable lactate analyzer (Lactate Pro, Arkray, KDK Corporation, Kyoto, Japan.) which has been validated as an effective analyzer for lactate measurements [69]. Additionally, a 0-10 Borg scale [51; 105] (see figure B.1) is used to determine the Borg feature of the athlete at the end of each stage.

Rating	**Descriptor**
0	Rest
1	Very, Very Easy
2	Easy
3	Moderate
4	Somewhat Hard
5	Hard
6	-
7	Very Hard
8	-
9	-
10	Maximal

Figure B.1: 0-10 Borg scale: Collection of rate of perceived exertion data

The data acquisition protocol is supervised and executed by a physician and a two-member supporting staff.

Preparation, calibration and start-up:
The following list describes the detailed steps prior to starting the data acquisition required at the beginning of every testing day.

- Start-up temperature and humidity sensors
- Start-up the HR monitor

  - Power on the device
  - Make a mock test to ensure that the monitor is correcly working

- Calibrate the Lactate measurement device
- Prepare the ITSP (see chapter 3.2) in the treadmill
- Check if all the material is prepared
  - Lactate measurement device
  - HR monitor + HR band
  - Reactive strips
  - Lancets
  - Measuring tape
  - Bio-impedance meter patches + meter
  - Fungibles: paper roll, latex globes, alcohol, cotton...
- Check temperature and humidity sensors
- Start-up of treadmill mock test

Confirmation of pre-requisite compliance:

Once the initial preparations are made, the next step is to ensure that the athletes fulfill all the test, health and legal requisites described in section 3.2. Failing to fulfil these requisites means that the ITSP is not performed.

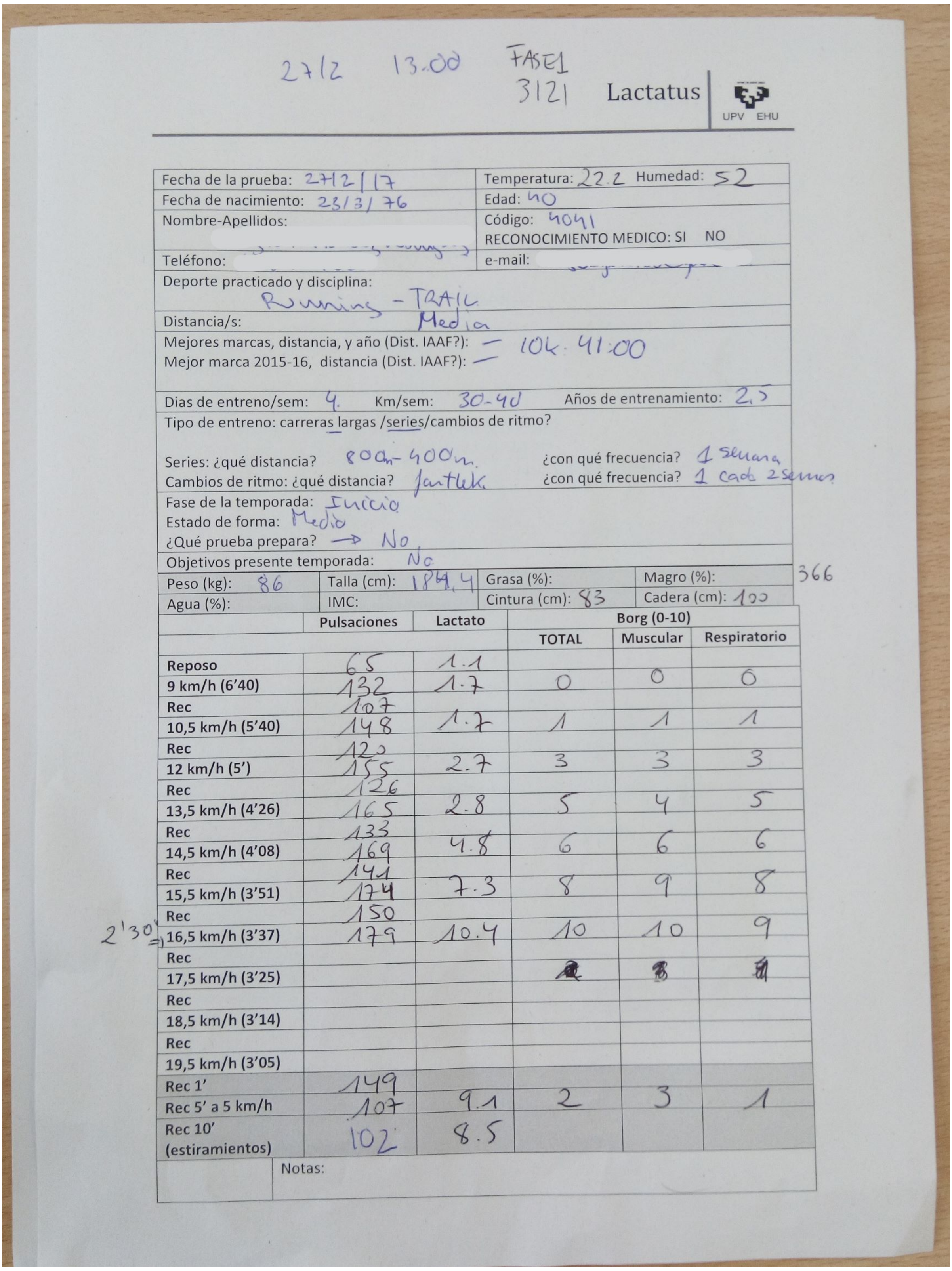
Formatting results:

Regarding the format of the database, every experiment is collected by the experimenters in parallel in two different formats, on paper and electronically in an (a priori formatted) excel file. This redundancy allows to minimize notation errors and the posterior check and correction of inconsistencies or incoherence. Images B.2 and B.3 show the paper and electronic formats respectively.

Static feature collection:

Then, the feature collection starts from the static ones following the steps described bellow:

- Questionnaire to collect historical data (training years, type of training...) -
- Anthropomorphic measurements: height, weight, waist and hip.
- Bio-impedance: Connect the device and make the measurement
- Place the HR band (with the athlete on the litter)
- Rest 5 minutes
- Measure resting HR
- Measure resting lactate



27/2 13.00 FASE1 3121

Lactatus

UPV EHU

Fecha de la prueba: 27/2/17	Temperatura: 22.2 Humedad: 52
Fecha de nacimiento: 23/3/76	Edad: 40
Nombre-Apellidos:	Código: 4041
	RECONOCIMIENTO MEDICO: SI NO
Teléfono:	e-mail:

Deporte practicado y disciplina: Running - TRAIL

Distancia/s: Media

Mejores marcas, distancia, y año (Dist. IAAF?): — 10k. 41:00

Mejor marca 2015-16, distancia (Dist. IAAF?): —

Dias de entreno/sem: 4. Km/sem: 30-40 Años de entrenamiento: 2,5

Tipo de entreno: carreras largas /series/cambios de ritmo?

Series: ¿qué distancia? 800m-400m. ¿con qué frecuencia? 1 semana

Cambios de ritmo: ¿qué distancia? fartlek. ¿con qué frecuencia? 1 cada 2 semanas

Fase de la temporada: Inicio

Estado de forma: Medio

¿Qué prueba prepara? → No

Objetivos presente temporada: No

Peso (kg): 86	Talla (cm): 184,4	Grasa (%):	Magro (%): 366
Agua (%):	IMC:	Cintura (cm): 83	Cadera (cm): 100

	Pulsaciones	Lactato	Borg (0-10) TOTAL	Muscular	Respiratorio
Reposo	65	1.1			
9 km/h (6'40)	132	1.7	0	0	0
Rec	107				
10,5 km/h (5'40)	148	1.7	1	1	1
Rec	120				
12 km/h (5')	155	2.7	3	3	3
Rec	126				
13,5 km/h (4'26)	165	2.8	5	4	5
Rec	133				
14,5 km/h (4'08)	169	4.8	6	6	6
Rec	141				
15,5 km/h (3'51)	174	7.3	8	9	8
Rec	150				
2'30" ⇒ 16,5 km/h (3'37)	179	10.4	10	10	9
Rec					
17,5 km/h (3'25)					
Rec					
18,5 km/h (3'14)					
Rec					
19,5 km/h (3'05)					
Rec 1'	149				
Rec 5' a 5 km/h	107	9.1	2	3	1
Rec 10' (estiramientos)	102	8.5			

Notas:

Figure B.2: Results collected in paper format

ID. Paciente	4041
Apellido:	
Nombre:	
Sexo:	M
F. Nacimiento	23/03/1976
Edad	40
Altura:	184,4
Peso:	86
IMC	25,3
Estado Forma	NA
Temperatura	22,2
Humedad	52
Observaciones	
BIA_ID	366

Medidas		C. Athlete	
Cintura cm	83	Discipline	Running
Cadera cm	100	Distance	long
% grasa	10,6	PB (IAAF M)	246
Grasa kg	9,1	PB (IAAF F)	719
Magro kg	76,9	Days	4
Peso Magro Sec	19,9	Series	yes
Agua %	66,3	Years	2,5
Agua lt	57	Protocol	1,5-1
Calculo met bas	2320	ADQ	Garmin
		T	1
		10k (time)	0:41:08
		10k (IAAF M)	246
		10k (IAAF F)	NA

Escalón	Speed	TrackRecID"	Time	Heart Rate	Lactato mmol/l	BORG Global	BORG Muscular	BORG Respiratorio
REC	INICIO	1"	12:28:41	65	1,1			
RUN	9_KMH	2"	12:32:42	132	1,7	0	0	0
REC	0_KMH	4"	12:33:42	107				
RUN	10,5_KMH	6"	12:37:42	148	1,7	1	1	1
REC	0_KMH	8"	12:38:42	120				
RUN	12_KMH	10"	12:42:43	155	2,7	3	3	3
REC	0_KMH	12"	12:43:42	126				
RUN	13,5_KMH	14"	12:47:42	165	2,8	5	4	5
REC	0_KMH	16"	12:48:42	133				
RUN	14,5_KMH	18"	12:52:44	169	4,8	6	6	6
REC	0_KMH	20"	12:53:44	141				
RUN	15,5_KMH	22"	12:57:43	174	7,3	8	9	8
REC	0_KMH	24"	12:58:44	150				
NF	16,5_KMH	26"	13:01:18	179	10,4	10	10	9
RECF	0_KMH	28"	13:02:19	149				
REC5	0_KMH	30"	13:07:18	107	9,1	2	3	1
REC10	0_KMH	31"	13:12:19	102	8,5			

Figure B.3: Results collected in digital format

- Sit down and dress up
- Ask if they want to hydrate (last chance prior to the test)

Time-series feature collection:

As previously mentioned, during the ITSP the time-series features are collected. Respiratory RPE [106], muscular RPE [107] and total RPE were assessed using the 10-point Borg scale [51]. HR was measured just at the end of each stage and at the end of the 1 minute recovery. Lactate concentration was measured at the end of each stage. More precisely, after starting the ITSP, at the end of each stage the following steps are performed:

- Check and annotate the HR in the result documents
- The physician measure blood lactate and say it out loud the support staff to annotate it in the results documents
- Ask RPE values and annotate in the result documents
- Check that the protocol is being correctly followed
- Annotate any incidence (double measurement required, dizziness...)
- After reaching exhaustion, push the button that initiates start the final recovery period
- 5 minute resting period walking on the treadmill at 5 kilometers/hour
- Check and annotate the HR value in the result documents

- The physician measure blood lactate and say it out loud the support staff to annotate it in the results documents
- 5 minute resting period seated
- Check and annotate the HR value in the result documents
- The physician measure blood lactate and say it out loud the support staff to annotate it in the results documents

Protocol for equipment failures:
During the tests, several equipment failures may occur. Here we define a protocol to properly handle and respond to different failure scenarios:

- Failure in the first or second step (9 kilometer/hour or 10.5 kilometer/hour) - Re-start the test
- Failure in higher steps - Discard the test, and date the athlete for another day
- HR monitor failure - Use the spare HR band

Digitalization of results:
Once the data collection is finished, as shown in figure B.4, results of the ITSP are gathered in a single excel file combined with the data of the HR monitor.

ID. Paciente	4008
Apellido:	
Nombre:	
Sexo:	M
F. Nacimiento	03/08/1970
Edad	46
Altura:	174
Peso:	64
IMC	21,1
Estado Forma	medium
Temperatura	22
Humedad	64
Observaciones	
BIA_ID	331

Medidas		C. Athlete	
Cintura cm	79	**Discipline**	Running Tri
Cadera cm	90	**Distance**	long
% grasa	15,8	**PB (IAAF M)**	405
Grasa kg	10,1	**PB (IAAF F)**	863
Magro kg	53,9	**Days**	6
Peso Magro Seco	12,4	**Series**	yes
Agua %	64,8	**Years**	6
Agua lt	41,5	**Protocol**	1,5-1
Calculo met basal	1609	**ADQ**	Garmin
		T	1
		10k (time)	0:37:54
		10k (IAAF M)	405
		10k (IAAF F)	NA

Escalón	Speed	TrackRecID"	Time	Heart Rate	Lactato mmol	BORG Muscul	BORG Respira	BORG Global
REC	**INICIO**	1"	10:10:50	45	1			
		1"	2017-02-07T10:10:51Z	54				
		1"	2017-02-07T10:10:52Z	53				
		1"	2017-02-07T10:10:53Z	54				
		1"	2017-02-07T10:10:54Z	54				
		1"	2017-02-07T10:10:55Z	55				
		1"	2017-02-07T10:10:56Z	58				
		1"	2017-02-07T10:10:57Z	60				
		1"	2017-02-07T10:10:58Z	63				
		1"	2017-02-07T10:10:59Z	67				
		1"	2017-02-07T10:11:00Z	71				
		1"	2017-02-07T10:11:01Z	73				
		1"	2017-02-07T10:11:02Z	75				
		1"	2017-02-07T10:11:03Z	78				
		1"	2017-02-07T10:11:04Z	80				
		1"	2017-02-07T10:11:05Z	83				
		1"	2017-02-07T10:11:06Z	85				

Figure B.4: Digitalized results with heart rate monitor data

Appendix C

Lactatus Report

Test Lactatus

Fecha de la prueba : **09/09/2018**

Hola [redacted] , recuerda que **todas las pruebas realizadas en un entorno controlado como el laboratorio o el gimnasio hay que ponerlas a prueba en una sesión de entreno al aire libre** y mejor si lo haces siguiendo las **recomendaciones que te detallamos en el punto 3.3.**

1.Datos del Deportista

Nombre: [redacted] Fecha Nacimiento: [redacted] Edad: [redacted]

Peso Actual: **77,00** Altura: **178 cm**

2.Perfil Deportivo

Deporte: **Triatlón** Tipo Prueba: **Media Larga Distancia**

Horas Entreno Semanal: **10 hrs** Días Entreno Semanal: **6 días** Años Entrenando: **4 años**

Fase Temporada: **Específico** Estado de Forma: **Bueno**

Series Velocidad: **Si** Frecuencia Semanal: **1,00 días**

Cambios de Ritmo: **No**

FC Max. Teorica: **177 ppm** (según formula Tanaka et al. (2001)) FC Reposo Prueba: **50 ppm**

FC Max. Test Lactatus: **174 ppm (-2% FC Max. Teórica)** FC Reposo Perfil: **50 ppm**

D. Max Test Lactatus: **156 ppm**

web: www.lactatus.com
mail: info@lactatus.com

3.Datos de la prueba

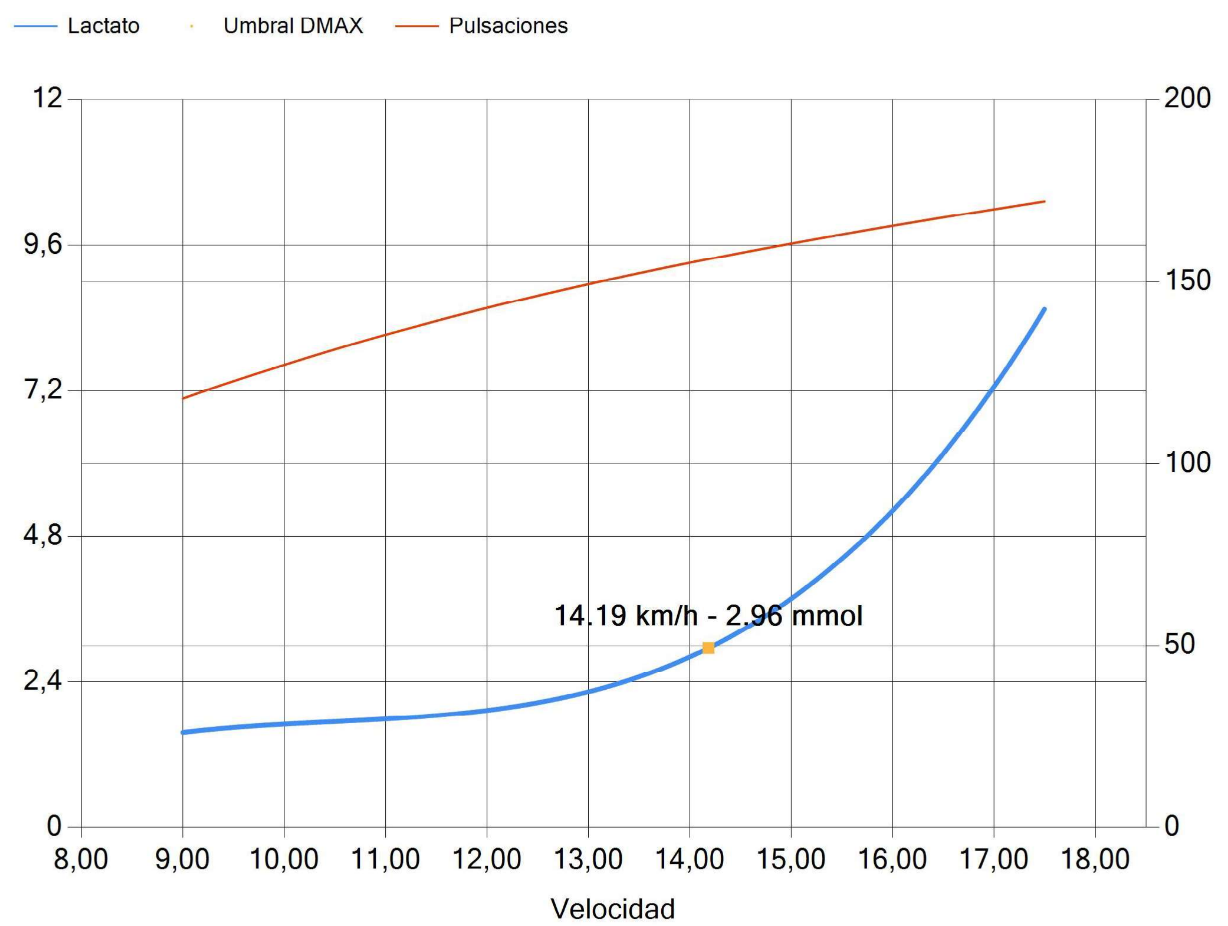
3.1.Escalones

Escalones	Pulsaciones (ppm/min)	Lactato (mmol)
Reposo	50	
9 kmh (6:40 mins/km)	118	1,58
Recuperación	84	
10.5 kmh (5:42 mins/km)	131	1,75
Recuperación	86	
12 kmh (5:00 mins/km)	143	1,94
Recuperación	96	
13.5 kmh (4:26 mins/km)	153	2,56
Recuperación	106	
14.5 kmh (4:08 mins/km)	157	3,11
Recuperación	112	
15.5 kmh (3:52 mins/km)	163	4,49
Recuperación	119	
16.5 kmh (3:38 mins/km)	168	6,38
Recuperación	131	
17.5 kmh (3:25 mins/km)	172	8,54
Recuperación	136	
18.5 kmh (3:14 mins/km) No finalizado solo corrió 87,00 s.	174	8,54
Recuperación	99	
Rec 5' a 5km/h	91	
Rec 10' parado	91	

3.2.Gráfica Umbral Lactatus

Esta grafica **nos da los valores concretos de pulsaciones, lactato, velocidad y ritmo por km** a partir de los cuales tu cuerpo se acerca a un punto de esfuerzo que si lo superas, puede suponer una caída en el rendimiento a medio o largo plazo dentro de una prueba o entreno concreto.

En el apartado de conclusiones presentamos algunas recomendaciones sobre cómo aplicar los resultados de este informe y ajustar los valores de los umbrales obtenidos, teniendo en cuenta que tu estado de forma actual que es **MedioAlto**, las **60** horas semanales que entrenas y que llevas haciendo este tipo de entrenamiento los últimos **4** años.Los **valores** aportados por la gráfica **Umbral Lactatus** que debes tener en cuenta para adaptar los entrenamientos a tu estado de forma actual y así maximizar la eficacia de los mismos son estos:



Valores DMAX

Umbral_LACTATUS	Ritmo_LACTATUS	Velocidad	Lactato
156 ppm	4:13 mins/km	14,19 km/h	2,96 mmol

3.3.Zonas de entrenamiento:

Éstas son las zonas de entrenamiento basadas en los resultados obtenidos en la prueba y **utilizando como Umbral Lactatus el valor de 156 ppm**

ZONA	Pulsaciones	Ritmo (mins/km)	Velocidad
Z1: Recuperación (inferior a 84%)	*Inferior a 131 ppm*	*5:43 mins/km - 5:12 mins/km*	*10,49km/h -11,51km/h*
Z2: Aeróbico (84%-89%)	*131 ppm - 139 ppm*	*5:43 mins/km - 5:12 mins/km*	*10,49km/h -11,51km/h*
Z3: Tempo - Resistencia (89%-94%)	*139 ppm - 147 ppm*	*5:12 mins/km - 4:44 mins/km*	*11,51km/h -12,66km/h*
Z4: Sub-Umbral (94%-99%)	*147 ppm - 154 ppm*	*4:44 mins/km - 4:20 mins/km*	*12,66km/h -13,8km/h*
Z5: Umbral (99%-103%)	*154 ppm - 161 ppm*	*4:20 mins/km - 3:58 mins/km*	*13,8km/h -15,12km/h*
Z6: Supra-Umbral (103%-106%)	*161 ppm - 165 ppm*	*3:58 mins/km - 3:46 mins/km*	*15,12km/h -15,93km/h*
Z7: Anaeróbica Máximo (mayor a 106%)	*Mayor a 165 ppm*	*3:58 mins/km - 3:46 mins/km*	*15,12km/h -15,93km/h*

Recuerda siempre que los **valores obtenidos en un test como este deben ser puestos a prueba y ajustados haciendo entrenos controlados en un entorno real como la pista de atletismo una salida en asfalto pro un terreno homogéneo preferiblemente llano.**

Lo ideal para poner a prueba estos valores, es que **pongas a prueba tu zona de Resistencia Aeróbica Máxima** con el siguiente test:

-**Calienta bien 10-15 minutos para subir el pulso de Z1 a Z3** de lo reflejado en la tabla, **en tu caso trata de subir hasta 147 ppm (Z3 alta) acabando con uno par de Sprints de 15-20 metros.**

-**Corre entre 40 y 60 minutos** a un ritmo constante **en Z4,** lo más cerca que puedas del límite superior de Z4, **en tu caso 154 ppm.**

-Al acabar deberías estar en uno de estos 2 casos:

-**Que aguantes bien,** en ese caso puedes **repetir el test aumentando el tiempo (10-15minutos), manteniendo pulsaciones en Z4 alta, en tu caso 154 ppm, o pasar a Z5 media, en tu caso 161 ppm y mantener el tiempo de la salida.**

-**Que no lo aguantes** y que tengas que **repetir el test a pulsaciones en Z4 media-Baja, en tu caso 147 ppm**

Repite el test hasta que tengas claro cuál es tu límite aeróbico máximo. Si estás en forma y entrenas con frecuencia no tendrás problemas en **trabajar sobre ese límite e incluso algo por encima para sacar el máximo provecho a la gran cantidad de entrenos que se hacen en Z4-Z5 a lo largo de una temporada.**