

G. Seleridis

EVALUATION OF THE CURRENT TEST METHODS OF WATER SENSITIVITY AND PERMANENT DEFORMATION

Evaluation Of The Current Test Methods Of Water Sensitivity And Permanent Deformation

by

Georgios Seleridis

in partial fulfilment of the requirements for the degree of

Master of Science

in Structural Engineering

at the Delft University of Technology,

to be defended publicly on Wednesday September 6, 2017 at 1:00 PM.

Student number:	4519213	
Supervisor:	Prof. dr. ir. S. Erkens	
Thesis committee:	Dr. K. Anupam,	TU Delft
	Ing. W. Verwaal,	TU Delft
	Ing. D. Van Vliet,	TNO
	Ir. L. Houben,	TU Delft

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.



Abstract

The adaption of the functional characterization of asphalt concrete mixtures by the Dutch authorities, led to a big necessity in acquiring a fundamental understanding on its properties. In order to achieve this, the long-term project *NL-Lab* was initiated by *InfraQuest* in 2012. With the advantage of its longevity and the multiple parties contributing with material, testing and analysing, the knowledge aimed for, is gradually accumulating with the time passing. After 5 years of running, a considerable number of data have been collected and analysed, leading to many observations, that in turn raised questions and doubts. These doubts are the subject of this research, focusing on the functional characteristics of water sensitivity and permanent deformation.

A common problem for both properties regards the production and compaction of type-testing specimens. The practice so far suggests both stages to take place in laboratory conditions. However, there were indications that the performance of the same asphalt mixtures taken from the field instead of the lab, differed. Testing and comparing in terms of water sensitivity and permanent deformation, these indications were indeed repeated. Lab and field specimens did not match in performance in the majority of the tests, suggesting that differences between the two conditions are projected in differences in performance.

The second common problem also evolves around the type-testing, this time regarding the test procedure as a whole. It is a highly time and money consuming process that has to be repeated in a trial-and-error manner, until a satisfying mixture in terms of performance is found. The aim of this research was to limit this process as much as possible by replacing it with regression relations. For the water sensitivity test, the relations obtained were of a fairly good predicting quality, and could be eventually used in the preliminary mixture design. The permanent deformation relations however, were at low levels of predictability, due to the test's high sensitivity in being reproduced and the inconsistent test setups used by the various contractors taking part in the project.

Focusing only on water sensitivity, the conditioning protocol described in the current standard, is considered as not very capable of capturing the property it aims for, and that it is conservative in the damage it causes to the mixture. Utilizing two conditioning protocols, MIST and Freeze-Thaw protocol, the standard method indeed was found to be the least damaging of all. With the MIST being the most destructive, at extreme levels in most cases, and the Freeze-Thaw following it, each method's pros and cons were found.

The last aspect tested referred to the friction reduction method used during the permanent deformation test (triaxial cyclic compression). The old and new standard versions both prescribe two different materials, while the requirements for rutting remained unchanged. Testing using these two materials, Teflon and Latex, plus one new, it was found that the differences in the triaxial test output were major. Consequently, this mismatch means that inconsistent comparisons take place, leading to conclusions on rutting performance not representative of the actual situation.

Preface

This thesis is written as the final exam for my Master's degree in Civil Engineering at the Technical University of Delft. Within the section of Pavement Engineering, I conducted this research in collaboration with Rijkswaterstaat and TNO. The task of this research included many problems, related to the materials, the testing equipment and the communication of all the people involved. But in the end they were faced successfully, and this is what made this journey even more interesting.

Of course, despite it being an individual's research, this could not have happened without the contribution of professor Sandra Erkens. She made it possible by trusting me with the duties of an intern at Rijkswaterstaat, and making me part of the NL-Lab team for the last year. Her guidance pushed me to consider aspects of problems that I would not think of at first.

Secondly, I want to thank Mr. Dave van Vliet for the very interesting and constructive cooperation we had during this year, and for the willingness he always showed in giving answers to all my doubts.

My next thanks go to Mr. Wim Verwaal, also a member of my assessment committee. It was very easy-going, and at the same time, effective, working with him and the CT-scanner, and I appreciate the time he spent with me on it. I also want to thank Dr. Kumar Anupam for being member of my assessment committee and helping me with his contribution at the weekly meetings we had.

I would like to thank the PhD candidate at Aristotle University of Thessaloniki, and close friend, Polyzois Bountzis, for dedicating a lot of his precious time to help me in the statistical analyses of this research, and taking them many steps further from what I would have managed alone.

For the material tested, I want to thank Rijkswaterstaat for making it possible financially, and the contractor Dura Vermeer for providing the asphalt specimens on time, and cooperating for every question I had.

For the laboratory work, I want to express my gratitude to Marco Poot and Michele van Aggelen. Without their expertise in the testing equipment and the adjustments and repairs we needed to make, this research would not have finished on time.

Finally, my biggest thanks go to my parents, Triantafyllos and Pepi, for literally making this Master's degree possible.

*G. Seleridis
Delft, September 2017*

CONTENTS

1 Introduction	13
1.1 Position of this Thesis	13
1.2 NL-Lab	13
1.2.1 Initiation	13
1.2.2 Research Questions	15
1.2.3 Overall Approach	15
1.2.4 Functional Characteristics	17
1.2.5 Materials	17
1.2.6 Recipe Based and Functional Characterization of Asphalt Mixtures	18
1.2.7 Relevance and Importance of the Project for the Stakeholders	19
1.2.8 My Internship on NL-LAB	20
1.2.8.1 Tasks	20
1.2.8.2 Outcome	21
1.2.8.3 Recommendations	22
2 This Thesis.....	24
2.1 Problem description	24
2.1.1 Lab vs Field specimens	24
2.1.2 Costly Type-Test realization and mixture's composition relation to performance ..	26
2.1.3 Moisture conditioning protocol's inability to capture the property	26
2.1.4 Undefined friction reduction methods in the triaxial test	26
2.2 Research objectives	29
2.3 Hypotheses	30
2.4 Methodology	30
2.4.1 General	30
2.4.2 Comparing lab to field data	31
2.4.3 Reliably predict functional characteristics based on composition data	31
2.4.4 Moisture conditioning protocols	32
2.4.5 Friction reduction methods	32
2.5 Testing plan	33
2.6 Thesis outline	34
3 Literature Review	36
3.1 Moisture Sensitivity	36
3.1.1 Definition	36
3.1.2 Long-term mechanisms	37
3.1.2.1 Adhesion	37
3.1.2.2 Cohesion	38

3.1.3	Short-term mechanisms.....	38
3.1.4	Influencing factors	39
3.2	Permanent Deformation.....	40
3.2.1	Definition	40
3.2.2	Types of permanent deformation	40
3.2.3	Development stages	42
3.2.4	Influencing factors	42
3.3	Analytical Tools	45
3.3.1	Regression Analysis.....	45
3.3.2	Box-plots.....	50
4	Laboratory Stage	52
4.1	Water Sensitivity	52
4.1.1	Introduction.....	52
4.1.2	Specimens	53
4.1.2.1	Organization.....	53
4.1.2.2	Production and Storage	53
4.1.2.3	Mixture composition	54
4.1.2.4	Compaction	57
4.1.2.5	Volumetrics.....	57
4.1.3	Indirect Tensile Test.....	60
4.1.4	Standard conditioning method	61
4.1.4.1	Procedure	61
4.1.5	Frost damage conditioning protocols	63
4.1.5.1	Different test temperature.....	67
4.1.6	MIST conditioning protocol.....	67
4.1.7	Overview of protocols.....	68
4.2	Permanent Deformation.....	70
4.2.1	Specimens	70
4.2.1.1	Organization.....	70
4.2.1.2	Volumetrics.....	70
4.2.2	Triaxial Cyclic Compression Test.....	73
4.2.2.1	Norm description.....	73
4.2.2.2	Friction reduction material application	76
4.2.2.3	Temperature conditioning	79
5	Moisture Sensitivity Analysis	81
5.1	Test Results	81
5.1.1	Visual Inspection.....	81

5.1.2	CT-Scans	84
5.1.3	Volumetric changes.....	85
5.1.4	Indirect Tension Test	86
5.2	Conditioning methods comparison.....	90
5.2.1	Analysis of Variance (ANOVA).....	90
5.2.2	Comparison	93
5.3	Comparison of lab to field determined properties	96
5.3.1	Tests output of this thesis.....	96
5.3.1.1	ITT	96
5.3.1.2	CT-Scans	98
5.3.2	Tests output from entire NL-Lab database.....	100
5.3.2.1	Analysis of Variance (ANOVA)	100
5.3.2.2	Lab to Field comparison.....	104
5.3.3	Relation to composition parameters.....	106
5.3.3.1	Statistical Relation	106
5.4	Performance prediction	112
5.4.1	Introduction.....	112
5.4.2	Data included	113
5.4.3	Analysis and checks	113
5.4.4	Predictors combinations.....	121
5.4.5	ITSR prediction.....	124
5.4.6	Risks.....	124
5.4.7	Example of application	125
6	Permanent Deformation Analysis	126
6.1	Test results.....	126
6.1.1	Creep curves.....	126
6.1.1.1	Stage 1.....	128
6.1.1.2	Stage 2.....	130
6.1.2	Linear and logarithmic curve fit.....	132
6.1.3	Cases with difficulty in slope definition and cases of failure	134
6.2	Comparison of lab to field determined properties	136
6.2.1	Tests output from this thesis	136
6.2.1.1	ANOVA	136
6.2.1.2	Comparison	139
6.2.2	Tests output from entire NL-Lab database.....	140
6.2.2.1	ANOVA	140
6.2.2.2	Comparison	142
6.2.3	Relation to composition parameters.....	143

6.2.3.1	Statistical relation	143
6.2.3.2	Comparison	147
6.3	Performance prediction	149
6.3.1	Introduction	149
6.3.2	Data included	149
6.3.3	Analysis and checks	150
6.3.4	Predictors combinations	154
6.3.4.1	Strain slope (f_c)	154
6.3.4.2	Permanent strain (ϵ_{1000})	155
6.4	Friction reduction methods	157
7	Conclusions & Recommendations.....	161
7.1	Conclusions	161
7.1.1	Moisture Sensitivity	161
7.1.1.1	Conditioning methods	161
7.1.1.2	Comparison of lab to field determined properties	162
7.1.1.3	Performance prediction	162
7.1.2	Permanent deformation	163
7.1.2.1	Comparison of lab to field determined properties	163
7.1.2.2	Performance prediction	163
7.1.2.3	Friction reduction methods.....	164
7.1.2.4	General conclusions	164
7.2	Recommendations	165
7.2.1	Moisture sensitivity	165
7.2.2	Permanent deformation	165
8	Bibliography.....	166

1 INTRODUCTION

1.1 Position of this Thesis

The programme that this thesis contributes to is called “NL-Lab” and stands for National-Living Lab. This programme started in 2012 and is conducted by InfraQuest, a competence centre made up of three parties; Rijkswaterstaat, TU Delft and TNO. Prior to this thesis, an internship in NL-Lab’s Program Team was followed, and served as an introduction and familiarization period, to eventually this research.

Firstly, the information describing the idea behind this project, and the relevant details needed to proceed with the research, are presented in Chapter 1. Also, an overview of the internship’s specific tasks and outcomes is included. The detailed information referring exclusively to this thesis are then presented in a separate chapter, Chapter 2.

1.2 NL-Lab

1.2.1 Initiation

The design and construction of an asphalt pavement is a complex process. Starting from the very early preliminary choice of its components and the mixture’s design, all the way to the final step of field compaction, there are numerous factors that determine the turn of the process, each one in its own way and extend. The final product has to satisfy certain requirements in order to serve its role through its lifetime and consider this process as successful.

In general terms, these factors can be divided into four main categories: technical, human, natural, and legislative. Technical related factors refer to all the equipment and machinery related aspects. Similarly, human related to the decisions, choices and handling by the people involved in the process, and natural to the effect of weather and environmental conditions. Finally, legislative factors are the ones related to the standardization of the process by the governing authorities. Each of these categories affects the entire process at different phases, with an importance that varies depending on the time-frame in which they appear.

The challenge of defining the relation of these factors with the final product has been the subject of numerous researches throughout the years. These researches have led to significant knowledge acquisition on this topic, which in turn has helped in optimizing the process of asphalt pavement construction. Also, it helped in shaping the process to match each geographic region's, and each time period's special characteristics. For example, from variations in the mixture composition based on the precipitation or noise reduction requirements of a country, to the policies imposed by the ever-developing human activities that have raised awareness on material shortage and environmental sustainability.

Despite the high level of optimization achieved, there is always the need to be one step ahead of changes. This need is legitimate and desired, being the driving force behind every development and advance taking place. It derives from changes in all four factors of the construction process. Introduction of new technologies, changes in human behaviour related to education and training, changes in the weather conditions and climate. In addition, competition in the field of constructions, at national and international level, and the aforementioned policies that define the framework in which construction activities take place. Together they result in a continuous need of researching deeper into the process.

A direct outcome of this was the initiation of the NL-Lab program. In 2008, the harmonized CEN standards for Asphalt Concrete were introduced in Europe. The way to characterize several types of Asphalt mixtures, Reclaimed Asphalt and the requirements for testing the mixes, and ensuring production quality are described in a series of standards. For Asphalt Concrete (AC) (NEN-EN 13108-1) the standard offers the choice between a classical, recipe based characterisation, or a functional characterisation, based on more mechanical-type properties in combination with some limited composition requirements.

Between these two methods of characterization, the Netherlands adopted the functional characterisation for AC mixtures, with the aim of developing a more fundamental and in-depth understanding of asphalt concrete response, providing in this way the opportunity to develop even better performing mixtures. However, the current understanding at that point was far from complete, despite the fact that the experiences since 2008 showed that, this approach allows for a better, more fundamental understanding of Asphalt Concrete.

During the period of 2008 to 2012, a number of developments took place, aiming at evaluating the current functional tests and establishing the relationship between these tests and the performance in the pavement. In specific, these developments included:

- Stiffnesses were reportedly higher than those known from past tests.
- Repeatability and reproducibility of the tests showed some issues, related either to experience, or also possibly, to the long time between the Type Tests.

- The effect of higher percentages (60-70%) of reclaimed asphalt led to surprising observations, since it appeared to improve all functional requirements, without the typical interrelation where an increase in stiffness corresponds with a decrease in fatigue resistance
- For some, especially low temperature mixtures, laboratory production proved to be difficult. This raised the question of how well actual field conditions are represented by lab conditions, which has a direct impact on the reliability of the performance predictions.

These experiences led to the initiation of this program using the Dutch road network as a living laboratory [1]. Its aim is:

1. To get an up to date reference frame based on commonly used mixtures, as well as a frame work for the evaluation and possibly improvement of the functional tests and the requirements based on them.
2. Assess the effects of mixing and compaction on functional properties.
3. Establish the predictive quality of (the current) lab determined functional properties for field performance [1].

1.2.2 Research Questions

The research questions that were derived from the experiences and developments that initiated the program are:

1. How (well) do the functional characteristics relate to field performance?
2. Is testing on laboratory mixed and compacted the correct choice?
3. Are the current tests able to distinguish “good” from “bad” mixtures?
4. How do the composition and bitumen parameters relate to the mixture’s performance?
5. How accurate and reliable the prediction of a mixture’s performance can be made, on the basis of these parameters?

1.2.3 Overall Approach

As mentioned, the Dutch road network is used as a living laboratory to answer the research questions. Although the Netherlands is a small country, the density of its road network (6th densest of the world with 331km of road per 100km² of land area), provide ample opportunity for field testing. However, field tests alone will not provide answers to the research questions. Several contractors have contributed in the project, firstly by providing the materials needed, and secondly by carrying out the tests on these materials, altogether forming the project’s wide database. The way the materials are

produced and tested differs, with each differentiation aiming at investigating a certain aspect in the testing process. In particular, the program combines lab testing with field monitoring as follows:

1. Assess the effect of mixing and compaction on the lab determined properties

This step is addressed by making specimens in three different ways:

- Phase 1 (F1): Lab Mixed – Lab Compacted according to test 62, RAW 2015¹
- Phase 2 (F2): Plant Mixed – Lab Compacted according to test 62, RAW 2015¹
- Phase 3 (F3): Plant Mixed – Field Compacted (specimens taken from the pavement soon after construction)

This step gives insight in the effects of mixing and compacting as well as providing a first indication of the relation between the predictive quality of lab mixed and compacted for field properties.

2. Follow the changes of lab determined properties over time

Directly after construction, specimens are taken for immediate testing (F3), as well as for testing after 2 (F3c) and 6 years (F3d). In addition, bitumen properties are studied 6 and 12 months after construction (F3a and F3b respectively). The cores are stored under controlled conditions. This way the effect of traffic is excluded and the changes in properties can solely be related to the changing material characteristics, related to ageing. An overview of all phases of production is seen in Table 1.1.

3. Monitor the pavement performance over time

This is straight forward for wearing courses, whereas for binder and base courses it is more complicated. For those locations, the monitoring is more indirect, based on the performance of the pavement structure as a whole.

4. Predict the functional properties

By making use of the data already recorded for the previous steps, the predictive quality of lab determined functional properties for field performance is established. Such predictive relations are

¹ Gyrotory compaction for cores (resistance to rutting and moisture sensitivity) and plate compaction for beams (stiffness and resistance to fatigue)

crucial in the preliminary design of a mix, to get a good first approximation and avoid the trial and error process of reaching this design.

Table 1.1 Overview of specimens phases

Phase	Production	Compaction	Time	Tests
F1	Lab	Lab	After construction	Asphalt & Bitumen (fresh and extracted)
F2	Plant	Lab	After construction	Asphalt & Bitumen (from silo and extracted)
F3	Field	Field	After construction	Asphalt & Bitumen
F3a	Field	Field	After 6 months	Bitumen
F3b	Field	Field	After 12 months	Bitumen
F3c	Field	Field	After 2 years	Asphalt & Bitumen
F3d	Field	Field	After 6 years	Asphalt & Bitumen

1.2.4 Functional Characteristics

The determined properties that are included in the program and are the criteria for all the comparisons made, are described by four functional characteristics. These characteristics are used by the Dutch standards for asphalt mixtures, *RAW2015*, as the functional requirements in an AC mixture's design, and include:

- (a) Resistance to Fatigue (NEN-EN 12697-24:2012) determined by a four-point bending test in continuous, full sinusoidal strain control at 20°C and 30Hz, aimed at determining the strain at which the material can take 1×10^6 load repetitions,
- (b) Stiffness (NEN-EN 12697-26:2012) determined by a four-point bending test in continuous, full sinusoidal strain control at 20°C and 8Hz,
- (c) Water Sensitivity (NEN-EN 12697-23:2003, NEN-EN 12697-12:2008) determined by the ITS ratio of the conditioned and unconditioned ITT of the material at 15°C, and
- (d) Resistance to Permanent Deformation (NEN-EN 12697-25:2016) determined by a Cyclic triaxial compression test, with temperature and loading conditions dependent on the position of the material in the pavement, aimed at determining the minimum slope of the permanent deformation versus load repetition curve.

1.2.5 Materials

The project is organized in terms of works. Until the end of 2016, 4 works had been carried out, each referring to a different pavement construction project. As a result, for each work a batch of specimens, provided by a different contractor is available. Each batch has different mixture design characteristics (Table 1.2), to assess the effect of their variation on the performance of Asphalt Concrete. All

mixtures have in common that they are defined based on the specifications set for base/binder courses, and all of them are categorized as asphalt concrete mixtures (AC). The specimens from work 1 were distributed over two different labs, to assess the effect of different specimen production (lab mixing and compaction) and test equipment. In work 2, the same lab used different compaction methods for the four point bending specimens (stiffness and fatigue) The passing material percentages in the final mix composition might slightly differ from the design values due to measuring accuracy limitations in the construction.

Table 1.2 Mixtures' design specifications

Construction Project	A4		N345		A28, HRL 157.700-156.100km		Bennenbroekerweg te Hoofddorp	
Contractor	OOMS/MNO		KWS		Van Der Lee		Boskalis	
Mix Identification	P1		P2		P3		P4	
Mixture Type (EN13108-1)	AC 22 Base, 50% PR		AC 22 Base 35/50, 60% PR		AC 22 Base/Binder 60% PR		AC 22 Base 40/60 60% PR	
Mixture Code	251		167163/267163		27774		A252	
Date Type Test Report	23-11-2011		09-09-2011		11-2013		21-12-2011	
Report Number Type Test	K FEC 2.0 APRR Platen		035-11		FEC 2.0_fase A		11806364 A	
Constituent Materials	% "IN" 100% mass							
Stone	Norwegian Granite 8/16	7,2	Bestone 8/11	14,92	Scottish Granite 16/22	10,8	Scottish Granite 16/22	8
	Norwegian Granite 16/22	9,6	Bestone 16/22	8,93		Scottish Granite 8/16	13,8	
	ECO-gravel	10,0						
Sand	ECO-sand	20,3	Course sand	13,17	River sand	20	Washed sand	12
Filler	Baghouse dust	1,0	Own Dust	1,22	Wigras 40k	2,6		
Reclaimed Asphalt	Crushed DAC 0/20	25,0	Crushed DAC 0/20	57,36	Milled AC 0/16	32,5	Frees 0/20	32,5
	Milled PA	25,0			Milled PA 0/16	32,5	Gebroken frees 0/20	32,5
Bitumen	70/100	1,9	70/100	1,76	160/220	1,6	70/100	1,6
	From RAC	2,4	From RAC	2,6	From RAC	2,4	From RAC	3,1
Composition	% Through Sieve							
	C22.4	100,0		100,0		99,0		97,0
	C16	94,9		91,0		87,0		90,0
	C11.2	80,5		84,0		80,0		80,0
	C8	64,9		71,0		60,0		65,0
	C5.6	55,0		58,0		52,0		55,0
	2 mm	39,9		47,0		43,0		44,0
	63 µm	5,6		6,9		8,0		6,6
	Filler	5,6		6,6		8,0		6,6
	Bitumen (in 100% mass)	4,3		4,3		4,5		4,3

1.2.6 Recipe Based and Functional Characterization of Asphalt Mixtures

The main reason that triggered the initiation of NL-LAB was the transition from the traditional recipe based characterization to the functional characterization. For this reason, a clarification of the facts and processes that differentiate the two methods is necessary at this point.

A characteristic of a recipe based (empirical) method is that it relies on practical experience rather than theories. This makes an empirical method descriptive rather than explaining. An empirical law can describe a phenomenon without providing an understanding, although the empirical law itself could be considered a sort of “understanding”; yet, this differs from an understanding in terms of fundamental principles, which have more general predictive value. An empirical law is predictive merely in its own reference system.

A system of contractual requirements and technical specifications works satisfactorily as long as it is operated within its framework of standardised technology. With any new development the road authority asks if current requirements are applicable, and if not, to develop new requirements. This question unfolds the recipe-based system’s restriction. The restriction lies in its empirical character and causes the limited applicability of the existing requirements and specifications to newly developed products, and the long time needed to evaluate the performance of new products, and consequently, a long time to develop new requirements. The time needed to develop new requirements, let alone the time needed to develop the knowledge to be able to develop a more fundamental approach, causes the implementation of innovative techniques and materials to stay at a low pace, until a system is developed which permits development of more generally applicable requirements.

An empirical methodology requires renewal of empirical reference data, based on practical experience. To gain practical experience with a new pavement design, or a new type of asphalt mixture, requires monitoring of the nominal service-life, to gather reference data, and to verify the performance (cost-effectiveness with respect to standard pavement designs, respectively asphalt mixtures). This leads to a delay of innovation that is no longer acceptable in the competitive field of constructions.

This situation improves when requirements are more generally applicable, not just to standardised technology, but to new technology as well. In a functional or performance related approach concerned with the prediction of pavement behaviour, and the evaluation of the cost-effectiveness and risk of failure, the material composition is irrelevant; relevant are only the properties needed to predict or judge the cost-effectiveness and the risk of failure. The link between the design characteristics and the functional behaviour is crucial in this sense [2].

1.2.7 Relevance and Importance of the Project for the Stakeholders

As far as Rijkswaterstaat and other agencies are concerned, NL-LAB has more than one target to fulfil, bringing numerous benefits for the organization. Specifically:

- Conclude on whether the lab tests can be reliably used as an indication of the actual field performance and adjust the requirements and QA/QC procedure accordingly.

- Modify the regulations, test procedures and test set-ups due to the redefined distinction between “good” and “bad” mixtures.
- Endorse competition in the research field regarding the introduction of innovations in asphalt mixtures, by abolishing the restrictions necessitated by the recipe based design.

Not only Rijkswaterstaat, but also the contractors themselves have a number of benefits to attain from the project:

- Save time and money by avoiding the time-consuming and expensive lab tests in the mix design process. The prediction of asphalt performance is done by using the relations provided by NL-Lab instead of carrying out the type-tests. This will give them a good indication of the properties in the preliminary design phase of the mixture and will eliminate the trial-and-error process in defining a mixture that satisfies the requirements.
- Freedom for the introduction of innovations and new materials. Restrictions resulting from recipe-based mixtures no longer apply, benefiting in extension the final user of the product, in this case the public.
- A new wide space of economic benefit potential for the contractors themselves.
- The potential to eventually go to functional verification of mixes in projects.

1.2.8 My Internship on NL-LAB

1.2.8.1 Tasks

For the fulfilment of the requirements of the Masters in Structural Engineering at TU Delft, I followed an internship related to Pavement Engineering. The internship took place in Rijkswaterstaat (Ministry of Infrastructure and the Environment) and lasted for 3 months, starting in September 2016. As mentioned, the duties assigned were targeted in contributing to the NL-LAB’s team actions.

Just before the start of the internship, contractors had carried out the type-tests up to Work 4 for phases 1-2-3a-3b, meaning that a large amount of data was available. The NL-Lab program team at that point had completed the processing and analysis of the results up to Work 1, not including the permanent deformation tests. After discussions with the parties involved, the emphasis of the internship was decided to be on two of the four functional properties included in NL-LAB, Permanent Deformation and Water Sensitivity, taking into consideration all 4 works available. The analysis was divided in two main targets for each property, covering two of the project’s aims:

- 1) Derivation of a performance prediction formula via a regression analysis
- 2) Comparison of the lab and field determined properties (F1 vs F3)

Due to the limited time available for the internship, the first part of the analysis relating to the performance prediction, was based on the corresponding analysis carried out in the Mechanistic – Empirical Pavement Design Guide (MEPDG) formed by the National Cooperative Highway Research Program (NCHRP) in the USA, taking the relations suggested as a starting point. After an extensive literature review on the two properties studied, and a period of familiarization with the tools used to derive the prediction formulas, the analysis was carried out. Several observations were made leading to significant conclusions, that in turn triggered topics to be further researched and eventually led to this thesis' scope of study.

1.2.8.2 Outcome

The analysis carried out and the conclusions reached from it, are described in the Technical Report submitted to the Internship Office of the Civil Engineering Faculty of TU Delft [3]. Looking into the prediction possibilities in the property of water sensitivity, the results were found to be fairly acceptable. An indicating factor of this, the R-squared value of the formula, showed values of $R^2=0,71$, which in combination with the equality scatter plot gave the image of a quite good prediction quality.

In the case of permanent deformation, the outcome of the analysis was different from the water sensitivity case. The prediction accuracies were not at satisfactory levels, lying around $R^2=0,50$, and the prediction could not safely be considered as accurate enough. Even though the experience in the analysis at this point was not very high because it was studied at the early stages of the internship, the observation was still that the problem did not lie on the analysis itself, but in the nature of the property and the dataset. It is a known fact that rutting is a very complex phenomenon with a lot of parameters playing a role in its occurrence. Trying to include as many of them as possible in a predictive relation, expectedly leads to difficulties. Another factor that affected the results, was the possible inappropriateness of the MEPDG as a basis for the prediction. The formulas suggested were referring to totally different test principles and conditions. The transformations done in order to correspond them, probably led to some loss in accuracy. For this reason, a safe conclusion could not be drawn and further research following different approaches was suggested in order to achieve this.

Lab specimens clearly showed both lower strength values and lower strength ratio values comparing to field specimens. Their performance in terms of water damage was thus considered as worse. One possible reason behind this inconsistency is the higher densities achieved in the field. The general trend shows a tendency for over-compacted field specimens. The result of this is less voids, thus less penetrating water in the mixture's body. The outcome of these deviations works on the safety side. Lab determination of water damage resistance is more conservative and is like having a safety factor applied. Coming to answer the fourth of the research questions stated at the beginning of the project, "How well do the functional characteristics relate to field performance?", the answer is positive with

respect to water damage. However, when we come to a more global consideration, this conservative (i.e. safe side) estimation of moisture damage resistance, which is the result of the mismatch in densities, becomes optimistic (i.e. risk) regarding rutting resistance, as it will be seen in the next paragraph.

Finally, looking at the comparison in the case of permanent deformation, the general image was that both in terms of ϵ_{1000} and in terms of f_c there is a significant difference between lab and field. Lab specimens seemed to perform better in rutting than the field specimens, both unaged and aged. An oddity was observed when comparing aged and unaged filed specimen with respect to ϵ_{1000} . They showed similar values, with a small tendency of aged values to be above the unaged. This was against the expectations stemming from the fact that ageing hardens bitumen, and in consequence leads to better rutting performance. In this case, the relation of lab to field properties was characterized as optimistic. Contrary to water damage, it does not lie on the safety side and should not be used as a reliable indicator of a mixture's performance. In addition, a direct relation between the two criteria (f_c and ϵ_{1000}) was not observed, meaning that even having a good correlation for one of them, it does not necessarily lead to deducting information about the other [3].

1.2.8.3 Recommendations

The completion of the internship's analysis with the results obtained led to several recommendations on some aspects that should more extensively be studied [3]. Also, some new parameters added in the project that would enhance its quality and covering of subjects were suggested. In specific, it was suggested to broaden the dataset by looking into deviations from the standard testing. This is explained below.

Regarding the water sensitivity determination, so far one method of conditioning the specimens was followed, the one described in NEN-EN 12697-12. Sometimes the information obtained from the test are insufficiently distinctive, meaning that they are not able to distinguish explicitly different mixtures behaviour. In general, besides the low ratios when the test was first used, all mixtures seem to pass this test. For this reason, different conditioning methods was suggested to also be investigated and conclude on whether the result is more representative and informative. In particular, two methods were proposed:

- (1) The moisture conditioning protocol followed in the frost damage method developed by TNO based on NEN 2872, and
- (2) the MIST method (Moisture Induced Sensitivity Tester) that has been proven to be able to distinguish among mixtures with different moisture damage characteristics.

In this way, a broader and more detailed data set will be obtained regarding the moisture susceptibility of the mixtures tested. The addition of extra data can be used to further support or reject the

conclusions drawn in the comparison between lab and field determined properties, and also possibly enhance the quality of the prediction relations.

So far, the triaxial tests for the determination of rutting resistance were all carried out at the same temperature and same loading conditions. Temperature is one of the most crucial factors affecting the performance of a mixture, the state of stress is another one. In this way, with no variability in the test conditions, differences in the performance cannot be linked to the temperature or stress state effect. It is suggested for this reason to follow a more diverse test temperature selection, that still lie within the test standards, and also assess the effect of different loading conditions, by using not only the binder/base layer test conditions, but also those for wearing courses.

As it was mentioned in the internship's conclusions, the use of the MEPDG relations as a basis was possibly a drawback for the final prediction quality. It would be more understandable and easily processed to start from point zero. Setting an initial regression equation directly for f_c and ϵ_{1000} , would certainly enhance the quality. This directness will in turn make the analysis and its final product friendlier to the researcher and the user.

The regression analysis regarding permanent deformation can be continued in more depth. The possible parameter combinations were not studied to their maximum extent because of the aforementioned lack of experience in the early stages. For this reason, more combinations can be tested in order to possibly enhance the quality or even come to the same conclusion, that the problem indeed lies in the dataset. A more complete analysis can be made, with a direct connection to the standards.

An additional recommendation refers to the friction reduction methods used in the triaxial cyclic compression test. Specifically, there are indications of a variety of methods being used by various contractors. This variety is the result of the standard's guidelines not explicitly defining the method to be followed, leaving space for interpretations. In combination, with the known importance of the material choice, a research on the effect of different materials' behaviour was suggested.

Finally, alternative computational tools for the data process are recommended. For the entire internship's analysis, Microsoft Excel was the software used, whether it was for the comparison of lab to field using boxplots, or for the performance prediction using the Regression function. Due to certain limitations imposed by the software, it was suggested to use more statistically dedicated packages e.g. SPSS or MATLAB, to take advantage of the wider capabilities offered [3].

2 THIS THESIS

The initiation and forming of this thesis was the outcome of the internship that preceded and contributed to NL-Lab. During that time, the ideas around that topic were evolving, to eventually, along with the conclusions of the internship, lead to the formation of this thesis' scope. However, this process of defining the targets and methodology was not restricted to the period before its start, but was an on-going process throughout its duration, that required adaptation based on the observations made. The problems presented are in general terms independent of each other. The common ground lies in their relation to the current norms in effect. The description of these problems, along with the methodology followed, are presented in this chapter, as a continuation of the work carried out previously. Some of the problems are an extension of NL-Lab, and some are entirely new additions that will work supplementary.

2.1 Problem description

2.1.1 Lab vs Field specimens

In general practice, during the design phase of a mixture there are certain property requirements for a mixture to be considered suitable and appropriate for use. These requirements depend on the pavement layer that the mixture is intended to be used in and on the traffic on the road. In order to evaluate these properties, Type Tests consisting of a number of tests to characterize the mixture, take place on specimens constructed with a certain composition. These tests include the determination of density, void content, stiffness, fatigue, rutting and moisture resistance.

Theoretically, the Type Test procedure is repeated until all requirements are met. This means that if the test outputs fulfil the requirements, the design process moves on to the next phase. If they are not fulfilled, another mixture composition is designed and tested until all the requirements are met. However, this process requires a highly elaborate and time consuming work, making it unattractive and not very realistically feasible. Also, since at the moment there is no Dutch Mix Design Method, Type Testing in the Netherlands is only done for the purposes of mix assessment.

The problem in this phase lies in the production of the specimens. The aforementioned type tests done to assess the mixture's resistance in certain distress types, are carried out on specimens that are produced in laboratory conditions. There are numerous studies on this matter that suggest that there is a difference in a mixture's performance that is related in the way in which it is produced and compacted. A comparative summary of the factors that affect the process is seen in Table 2.1. It should be noted that this table, coming from a US source, will reflect aspects and conditions that differ from what is the case in the Netherlands, but it can still generally relate to the fundamental aspects. A task included in this thesis will be to recognize these factors, in the framework of NL-Lab.

Table 2.1 Laboratory vs Field conditions [4]

Laboratory Conditions	Field Conditions
Binder	
<p>Aging is simulated using the TFO, RTFO or PAV. All of these methods are only rough simulations of actual asphalt binder aging.</p> <p>After mixing, the loose mix is generally aged to allow for asphalt binder absorption and an increase in viscosity.</p>	<p>Aging is much more complex – especially after construction when it is highly dependent upon construction quality and the environment.</p> <p>After mixing the loose mix can be immediately transported to the construction site or can be placed in storage silos for up to a week.</p>
Aggregates	
<p>Gradation is carefully measured and controlled</p> <p>Aggregate used is completely dry</p> <p>Fines are retained during the mixing process</p> <p>Oven heating of the aggregate usually results in uniform heating of the coarse and fine aggregate</p> <p>If RAP is used, it is heated to the same uniform temperature as the virgin aggregate.</p>	<p>During the manufacturing process, aggregate gradation will change slightly as it passes through the cold feed bins, aggregate dryer and drum mixer/pugmill</p> <p>Even after drying, aggregates typically contain between 0.1 – 0.5 % moisture by weight</p> <p>Some fines are collected in the mix plant baghouse</p> <p>In a drum plant there is often a distinct temperature difference between the coarse and fine aggregate</p> <p>If RAP is used its degree of heating may be different than the virgin aggregate.</p>
Mixing Process	
<p>The mixing process occurs on essentially unaged asphalt binder.</p>	<p>The mixing process can substantially age the asphalt binder. A mixing time of 45 seconds can increase asphalt binder viscosity by up to 4 times</p>
Compaction	
<p>Compaction uses a laboratory device and a small cylindrical or rectangular slab of HMA. This combination attempts to simulate the practical orientation achieved by field compaction with rollers</p> <p>Compaction is relatively quick (<5 minutes) and thus occurs at an almost constant temperature</p> <p>Compaction occurs against a solid foundation</p>	<p>Particle orientation and compactive effort can vary widely depending upon roller variables and the environment (e.g., temperature, wind speed).</p> <p>Compaction can take a significant amount of time (30 minutes or more in some cases) and thus occurs over a wide range of mix temperatures</p> <p>Foundation rigidity will affect compaction. Compaction can occur against a range of foundations: some can be quite stiff (like old pavement) while some can be quite soft (like a clay subgrade).</p>

It is obvious that a considerable number of factors differ between the two methods. In addition, early studies during the internship confirmed this mismatch in the mixture's performance. This leads to the formulation of the first problem studied in this thesis:

Is testing lab produced and compacted specimens in order to evaluate a mixture's performance representative of the actual performance when the mixture will finally be laid in the field?

Are we able to safely characterize a mixture's behavior only by testing a lab produced specimen?

2.1.2 Costly Type-Test realization and mixture's composition relation to performance

Similarly to the process described in the previous paragraph, this problem is related to the preliminary choice of the mixture's materials and composition, and the type-testing that takes place to assess whether it meets the requirements or not. This is a trial and error process that may require a series of many tests to eventually design a mixture that fulfils all the standards. The number of tests depends on the designer's experience in choosing the mix parameters, and in the complexity of the requirements. This procedure is highly time- and money-consuming, both for the contractor and the governing authorities.

2.1.3 Moisture conditioning protocol's inability to capture the property

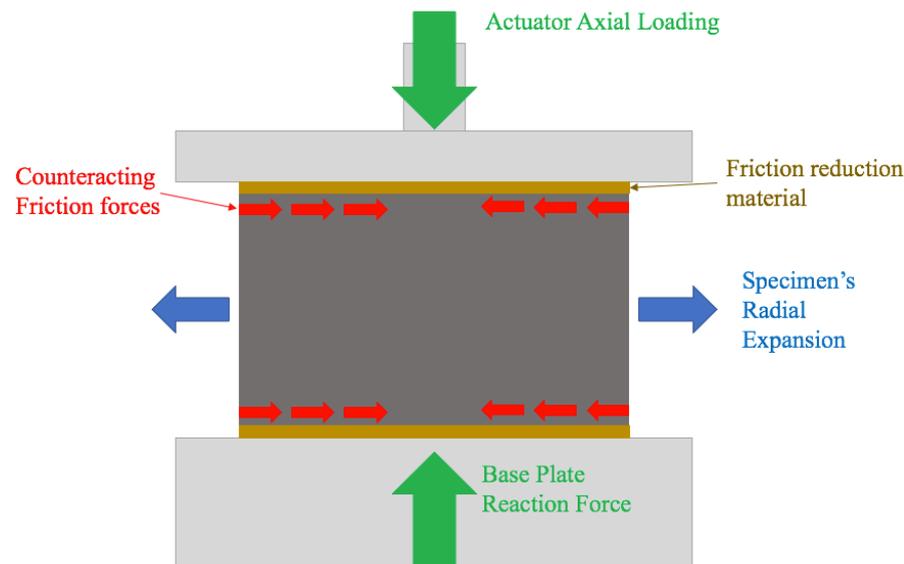
As it will be analyzed in detail in Chapter 3.1, moisture damage in asphalt pavements is distinguished in long- and short-term processes, based on the time frame over which each process occurs. Moisture diffusion refers to long-term, whereas pumping action resulting from wheel passes takes place in the short-term.

The standards currently utilized to define the moisture susceptibility of a mixture describe a moisture conditioning protocol in NEN-EN 12697-12:2008, where the specimens are firstly subjected into a vacuum water bath to saturate, and then into a 70-hour water bath at 40°C. However, there are several drawbacks that have been reported about this protocol, including poor correlation with field performance and micro-cracks induced during the vacuum suction process. Taking into account the nature of the conditioning, being a plain 3-day water bath, it is safe to say that it fails to capture the time frame over which moisture infiltration occurs and does not take into account the aforementioned short-term damages resulting from pumping action [5].

2.1.4 Undefined friction reduction methods in the triaxial test

The determination of the resistance in permanent deformation is done according to NEN-EN 12697-25. This standard utilizes the Triaxial Cyclic Compression Test on a cylindrical asphalt concrete specimen to determine its creep curve.

During the axial loading of the specimen, between the top and bottom interfaces of the loading plate and the specimen, high friction is generated. This is because of the radial expansion of the specimen, which is the result of its axial compression in combination with its Poisson's ratio. Naturally, this phenomenon influences the specimen's deformation and its test result, and this is the reason the standard refers to the reduction of friction using various methods.



Picture 2.1 Schematic representation of Friction development during a Triaxial Cyclic Compression Test

The original abstract from the current standard (EN12697-25:2016) that describes these methods is:

“8.4.3 To minimize the friction between the upper and lower loading platens and the test specimen, the end faces of the test specimen shall be smooth and plain. Brush the hand over the test specimens' surface. If it feels even without blemishes, it shall be considered adequate, otherwise it shall be polished or ground.

A friction reducing system shall be applied. The friction-reducing system shall consist of a circular disk cut out of a PTFE-sheet (e.g. Teflon or equivalent). The PTFE-sheet shall have a thickness of 0,5 mm and a shore hardness between D50 and D60. The diameter of the disk shall comply with the diameter of the loading platen.

Other systems to reduce the friction between loading platen and test specimen surface may be applied if proven that the alternative systems reduces the friction in a similar extend as prescribed system and doesn't influence the axial strain measurement.

NOTE The amount of friction between the loading platens and the test specimen is known to have a large impact on the results.”

At this point. it is interesting comparing this abstract, with the one from the withdrawn version of the standard, which was valid until the previous year (EN12697-25:2005):

“5.4.6 To minimize the friction between the loading platens and the test specimen, a membrane-lubricant-membrane-system shall be used between the loading platens and the specimen.

NOTE 1 The membrane may e.g. consist of a disk cut out of typical geotechnical latex rubber membranes, e.g. ELE P/N EL-25-7621 or WFI P/N 11091 or equivalent, of the same diameter as the specimen. A small amount of silicon grease should be applied between both membranes.

NOTE 2 The amount of friction between the loading platens and the specimen is known to have a large impact on the results.

NOTE 3 Instead of putting the test specimen in direct contact to the loading platens, the specimen may be glued by its extremities on steel plates

It is obvious that the two different versions suggest two different materials for the friction reduction; the current suggests a Teflon sheet, while the withdrawn Latex membranes with silicon grease. The explicitness in the way in which they suggest them is worth comparing.

First of all, the 2005 version suggests the material in a Note paragraph, and not in the main text. Notes have an advisory role and it is not compulsory to be followed. Also, the vocabulary and context used in the note (‘may’, ‘e.g.’, ‘or equivalent’) do not strictly imply to follow this suggestion, but leave it open to the preference of the tester.

Looking at the updated version of 2016, there is a noticeable improvement in the way the friction reduction material is prescribed. The material definition is done on the main text, and while possibility for alternatives methods is still given, it is required to be proven that their function will be similar. However, the use of words ‘shall’, ‘may’ and ‘e.g.’, still gives the impression of not explicitly standardizing the material.

The issue that arises by this change, which also happened in the recent past, is that there might still be labs that prefer to follow the withdrawn version with their ‘own’ interpretation of the definition. Taking into account the indications that different methods lead to different results, their choice is a subjective task of the tester and the fact that certain methods lead to more “attractive” and favorable numbers (i.e. lower deformation (rates)), can lead to systematic differences between labs. On top of that, considering the fact that despite the change in the standard, the requirements remained the same, the need to assess the interference of each friction reduction method in the testing process and explicitly standardize the one with least influence, is obvious.

2.2 Research objectives

The identification and forming of the problems described, aims in getting an understanding of the situation, and based on this, translate the problems into objectives, whose solution will eventually be sought for. Even though the 4 different problems might seem independent of each other, they are all related to the current standard that applies in Europe, and specifically in the Netherlands. This forms the general objective of this research, which is to assess the standard, from various aspects, identify any inaccuracies and eventually suggest adjustments or adaptations of alternative procedures.

In addition, the research on each of the problems stated, is not going to be independent. The steps followed lead to information relative to more than one problem, meaning that there are interrelated aspects between them. The research objectives that were formed according to these problems, and all of them collectively aim to the standard's assessment are:

1. Determine whether producing and testing specimen in the lab can be utilized to determine if a mixture will satisfy the requirements for application in the field. Is it an optimistic or a pessimistic approximation?
2. Identify the composition and bitumen parameters that are both physically and statistically related to the mixture's performance. In this way, a deeper understanding on the factors that determine performance will be obtained, and the attention will be driven on them.
3. Produce relations that predict the functional characteristics (specifically: ITS(R) and rutting resistance) based on composition and constituent material properties for the preliminary design of a mixture's composition. Additionally, broaden the input's predicting parameters included in the relations, by taking into account more elaborate test conditions.
4. Give an outlook of each moisture conditioning protocol's ability to capture the moisture susceptibility property.
5. Determine if and to what extent the friction reduction method in the cyclic triaxial compression test affects the result, and advice on the best system to use, the one that intervenes the least in the loading process.

2.3 Hypotheses

The research questions and objectives have been used to define the following hypotheses:

1. Lab mixed and compacted specimens can be used to determine a mixture's field properties. This means that there is no significant difference between F1 and F3 results.
2. It is possible to give an estimate of the functional properties (moisture sensitivity and rutting resistance) based on regression relations, using mix composition and constituent material properties. This means the predicted and measured values for the property are the same with 90% reliability.
3. The standard moisture conditioning protocol is sufficiently able to expose a mixture's moisture susceptibility. This means that the difference in strength reduction between the different protocols is not significant.
4. The various friction reduction methods prescribed in the norms, lead to consistent results with small or no interventions in the testing procedure. This means that mixture is expected to behave the same, regardless the friction reduction method chosen, and will result in similar strain rate (f_c) and permanent deformation levels (ϵ_{1000})

2.4 Methodology

2.4.1 General

The methodology followed to cover the research objectives, builds on the main procedures adopted by NL-Lab. Having an already vast database of 450 specimens tested since 2012, divided in 4 Works, the project was extended to an additional fifth work, in the framework of which this thesis took place. In this fifth project, the contractor, as for the other four projects (or works), produced specimens in the lab that were mixed and compacted there (F1), specimens that were mixed in an asphalt plant and compacted in the lab (F2), and specimens taken from the field compacted asphalt mix (F3). Besides this set of specimens that was tested by the contractor, an additional batch of cores for each of the three phases was produced and delivered to TU Delft for additional testing for this thesis. An overview of the specimens for Work 5 is given in Table 2.3 and Table 2.4, in Paragraph 2.5. The contractor results are also used in this project, along with the results from other contractors in the previous 4 works. Based on this data, each of the four research questions is answered in a different manner.

2.4.2 Comparing lab to field data

First of all, in order to establish the comparison between lab produced and field produced specimens, the specimens used for the testing were constructed, as the target suggests, in two different ways; half produced in lab conditions, and half in actual field conditions. After their testing in water sensitivity and permanent deformation, the comparison is facilitated through a statistical analysis and presented in boxplots and bar charts, distinguished by phase (F1 respectively, F3). In this way, the relative comparison between the values is visualized and conclusions are easily drawn. In addition, the study of the CT-scans is used to try to visualize the differences in void distribution within the specimens'. Besides the comparison of the test results, the specimens' volumetric properties are also compared through boxplots, again distinguished by phase. In this way, the explaining factors behind possible matches or mismatches between the two phases will be recognized and traced back to potential differences in the construction phase.

2.4.3 Reliably predict functional characteristics based on composition data

Regarding the second research objective, the principle is, through the big database of tests and properties, to predict the outcome of the tests. The main statistical tool used in such researches, in all kinds of scientific disciplines, is the Regression Analysis. In this way, using as an input the specimen's properties (volumetric, bitumen, production), the relation obtained by the regression analysis will give as an output the predicted mixture's performance in a certain distress type. The crucial point during the analysis, is choosing the parameters of the mixture that should be included in the relation. This is achieved by various statistical analyses that take place before the regression analysis, and indicate the importance, significance and the correlation of the parameter with the property that is aimed to be predicted.

This exact methodology is also followed in NL-Lab. What differentiates this thesis is the addition of two extra parameters in the database, and the investigation of their significance in the final prediction. These parameters are related to the test setup of the permanent deformation triaxial test and are the test temperature and the maximum stress applied on the specimen during the cyclic loading (deviator + confining pressure). The differentiation of these parameters based on layer are seen in (Table 2.2). So far, until Work 4, all the tests took place at the same temperature (40°C) and maximum stress (450 kPa), according to the standards specification for a base/binder layer (RAW 2015). However, it is widely known that both temperature and stress state greatly affect the specimen's behavior, and for this reason it was decided to perform the tests using all the combinations of these parameters in the Standard.

Table 2.2 Stress conditions and specifications for a top layer and other layers

	Top Layer	Other Layers
Compaction Method	Gyrator	Gyrator
Sample diameter	100 mm	100 mm
Sample height ($D \leq 16\text{mm}$)	60 mm	60 mm
Sample height ($D > 16\text{mm}$)	80 mm	80 mm
Conditioning Temperature	15°C	15°C
Test Temperature	50°C	40°C
Axial signal shape	Haversine	Haversine
Axial stress pulse duration	0.4 s	0.4 s
Axial stress rest-time	0.6 s	0.6 s
Axial stress amplitude (σ_v)	0.30 MPa	0.20 MPa
Confinement (σ_c)	0.15 MPa	0.05 MPa
Peakload ($\sigma_{1,\text{max}}$)	0.75 MPa	0.45 MPa
Max number of repetitions	10000	10000

The additional temperature is 50°C and the maximum stress 750 kPa, which are the parameters that are usually used for a surface layer. The friction reduction method used in this stage did not follow the standard's prescription, which in the current version of 2016 is Teflon sheets. Instead, plastic foil with conventional soap was used, for the reason that there were indications about a doubtful function of Teflon sheets and the better performance of this plastic foil. This comparison is elaborately answered by the fourth research question that follows this paragraph. An overview of the combinations is presented in paragraph 2.5.

2.4.4 Moisture conditioning protocols

As far as the moisture sensitivity conditioning methods are concerned, three different protocols were used. The standard method which is used by the EN standard, also in NL-Lab, and two additional methods that are not (yet) part of the standard; the MIST protocol (4.1.6) and the Freeze Thaw protocol (4.1.5). Both protocols will be described in more detail in Chapter 4.1. By looking at the way and extent each method reduces the specimens initial indirect tensile strength and the extent to which they “magnify” the mixture's sensitivity to water damage, a conclusion will be drawn on which protocol would be more appropriate to follow.

2.4.5 Friction reduction methods

The final research objective is dealt with using the Phase 2 (F2) specimens (plant produced - lab compacted) for triaxial testing. Testing them in cyclic triaxial compression test according to the standard (EN 12697-25, method B), using various methods of friction reduction and a reference set of tests without any friction reduction, will give a fair indication of how each method intervenes in the capturing of the specimen's performance. In the end, taking also into account the ease of access of

each product in the market and the ease of installation, a first recommendation will be made. The test setup in this case will follow the standard's guidelines for a binder/base course, meaning a temperature of 40°C and maximum stress of 450 kPa. The various friction reduction systems used are described in paragraph 4.2.2.2.

2.5 Testing plan

In total, 94 tests were planned in order to sufficiently cover the questions. 50 of those referred to the permanent deformation part, and the remaining 44 to the moisture sensitivity. Additionally, 4 extra permanent deformation tests were carried out as a backup, for reasons that are explained in chapter 4, increasing the total number of tests to 98. The testing period started in the middle of February 2017, and finished on the 19th of May 2017. The conditioning of the moisture damage protocols was taking place in parallel. The analytic overview of the number of specimens divided by the property investigated is seen in the tables below.

Table 2.3 Overview of the Permanent Deformation testing stage

Phase	Protocol	Temperature °C	Maximum Stress kPa	Friction Reduction Method	Number of tests	Test Date
F1	I	40	450	Plastic + Soap	4	8 th March – 1 st April 2017
	II	40	750		4	
	III	50	450		4	
	IV	50	750		4	
F3	I	40	450	Plastic + Soap	4	8 th March – 1 st April 2017
	II	40	750		4	
	III	50	450		4	
	IV	50	750		4	
F2	I	40	450	No reduction	4	9 th – 19 th May 2017
				Plastic + Soap	4	
				Teflon	4	
				Latex	6	
Total					50	

Table 2.4 Overview of the Moisture Sensitivity testing stage

Conditioning Method	Condition	Phase	Number of tests	Test Date
Standard	Dry	F1	3	13 th February 2017
		F3	3	
	Wet	F1	3	
		F3	3	
MIST	Wet + MIST	F1	4	15 th February 2017
		F3	4	
Freeze Thaw	Dry	F1	3	10 th May 2017
		F3	3	
	Thermal loading	F1	3	
		F3	3	
	Wet	F1	3	
		F3	3	
	Wet + Thermal Loading	F1	3	
		F3	3	
Total			44	

2.6 Thesis outline

This thesis consists of 6 chapters that intend to give all the vital information needed and finally answer the research questions. The introductory chapter describes in detail the foundation of this thesis and the work that preceded; InfraQuest's program NL-Lab and my internship that created the basis and set the framework for the formation of this thesis. It is information necessary for the in-depth understanding of the work that follows.

The second chapter describes the specific problems that constitute the subject of the thesis. Problems already defined by NL-Lab, and additional ones that came up during the internship period. The methodology and testing plan that were designed to answer these questions is also presented in this chapter.

Chapter 3 covers all theoretical background needed to carry out this research. It includes a literature review regarding the nature of the two distress types studied, permanent deformation and moisture sensitivity. Also, the statistical tools used to reach the targets are described in detail.

The laboratory procedures followed, are described in Chapter 4. The steps in every conditioning protocol and testing, as they were defined in the standards, or in the independent manuals, are given in detail, along with the mixture's and specimens' specifications. Also, important observations or alterations during the testing phase, are noted to help in the result interpretation and data analysis.

Chapters 5 and 6 deal with the functional characteristics of Moisture Sensitivity and Permanent Deformation respectively. Firstly presenting the results obtained from the tests, followed by the

necessary statistical analyses, data processing and their interpretation. All the questions set at the beginning of the thesis are answered in this chapter using the appropriate graphs and tables.

Chapter 7 consists of conclusions deriving from this research, and the recommended additions and adjustments for future research.

Finally, chapter 8 contains the bibliography studied during the course of this thesis.

3 LITERATURE REVIEW

The two major properties of asphalt concrete studied in this graduation project are Permanent Deformation and Water Sensitivity. The background information needed to go deeper into these phenomena and conduct the analysis to reach the targets set, is presented in this chapter. In addition, a description of the methodology followed in the data processing is presented.

3.1 Moisture Sensitivity

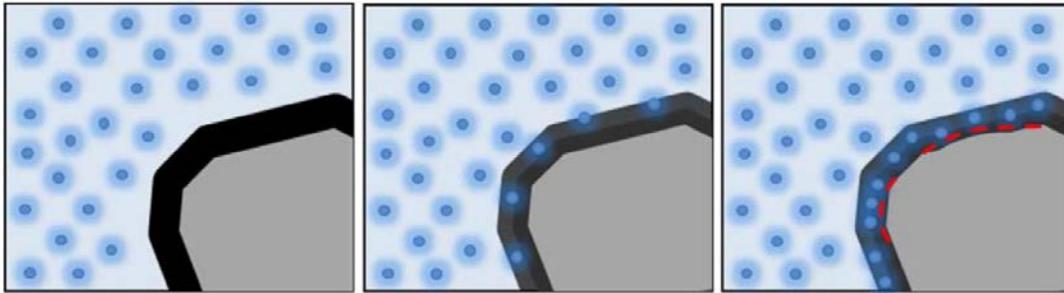
3.1.1 Definition

During their service life, asphalt pavements undergo a combination of traffic and environmental loadings that lead to a deterioration in their overall performance, leading eventually in various forms of damage like rutting, ravelling and cracking. Moisture damage has proven to be a major contributor to accelerated deterioration in asphalt pavements [5]. Moisture damage is mostly manifested through three mechanisms, with the first two being the most important:

- 1) Loss of cohesion through a gross softening of the bitumen or weakening of asphalt concrete mixtures.
- 2) Loss of adhesion between the aggregate and the bitumen, also known as stripping.
- 3) Degradation or fracture of individual aggregate particles when subjected to freezing.

It is a generally agreed fact that moisture has a disrupting effect in the integrity of the structure of bituminous mixtures, through these three mechanisms.

Depending on the traffic loading characteristics and climate conditions, the asphalt mixture type as well as the type of asphalt binder in the quality of the aggregate-binder bond, failure due to moisture damage can be either of cohesive or adhesive nature. Both mechanisms eventually show their results on the long-term. Reduction of the strength and stiffness of a mixture is often the result of cohesion reduction. A pavement with reduced strength loses its ability to support traffic-induced stresses and strains. Also, loss of bond between aggregate and bitumen, leads to a reduction in pavement support. Both mechanisms result in weaker pavement layers which are susceptible to deformations under traffic loading, and in the case of stripping, loss of material and deterioration of the mixture.



Picture 3.1 Damage of the binder and the binder-aggregate interface due to moisture diffusion [5]

3.1.2 Long-term mechanisms

3.1.2.1 Adhesion

An understanding of the factors that cause the loss of adhesion requires the knowledge of the mechanisms through which it occurs in a mixture. There are 4 main ways of asphalt binder – aggregate adhesion:

Mechanical

Irregularities and pores in the surface of the aggregate allow the asphalt binder to enter and create an interlock with its hardening. In case moisture is present on the aggregate, it can interfere with the binder's penetration in the aggregate and deteriorate the mechanical interlock. This increases the susceptibility to stripping.

Chemical

Chemical adhesion is caused through the asphalt binder's and aggregate's surface chemical reaction. Generally, aggregates with acidic surfaces react weaker with asphalt binders, potentially resulting to other moisture damage factors, if not sufficiently strong.

Adhesion tension

“Wetting line” is the edge of the drop, as a drop spreads over a surface. The tension between the asphalt binder and aggregate along the wetting line is in general lower than the tension between water and aggregate. For this reason, if all three are in contact, asphalt binder will be displaced by water, resulting in poor wetting of the aggregate surface by the asphalt binder. This is a cause of stripping.

Molecular orientation

When in contact with aggregate, asphalt molecules tend to orient themselves in relation to the ions on the aggregate surface essentially creating a weak attraction between the asphalt binder and aggregate surface. If water molecules, which are dipolar, are more polar than asphalt binder molecules, they may

preferentially satisfy the energy demands of the aggregate surface. The resulting weak asphalt binder-aggregate bond can result in stripping [6].

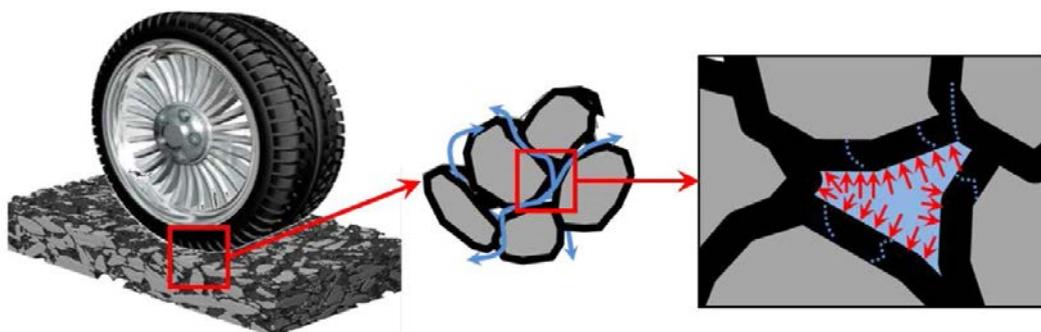
It is most likely that two or more mechanisms occur simultaneously in a mixture to cause loss of adhesion, and all of them may occur to some extent in any asphalt- aggregate system. In the case of this thesis the mechanisms that are expected to occur are mainly the mechanical and chemical. The standard conditioning method, with just the water bath, results in chemical changes only. There is still the potential for mechanical damage during the vacuum water bath at the initial stage, however this a side-effect and is not desirable by the protocol. In the MIST and Freeze Thaw protocols, the expected mechanisms are both chemical (by the water-bitumen interaction during the water bath), and mechanical (by the pressure loading or ice expansion).

3.1.2.2 Cohesion

Under the assumption that adhesion between aggregate and asphalt is adequate, cohesive forces will develop in the asphalt film or matrix. Factors such as viscosity of the asphalt-filler system can influence the cohesion values. Water can affect cohesion through the intrusion into the asphalt binder film or through saturation and expansion of the void system (swelling) [7]. Also, as moisture infiltrates into the asphalt mixture, the physio-chemical properties of the asphalt binder can change, thus reducing its cohesive strength.

3.1.3 Short-term mechanisms

Pumping action is a short-term damage mechanism that can act accumulatively and possibly accelerate the long-term damage mechanisms. The interconnection of asphaltic mixture pores allows the water to move through the mixture's mass. High water pressure fields within the pores that are filled with water can be caused by the dynamic traffic loads. These pore pressures can cause the binder film to crack, facilitating in this way the penetration of moisture to the asphalt binder-aggregate interface. On top of that, the intense pore pressures can cause desorption of the already weak binder, referred to as erosion [5].



Picture 3.2 Pore pressure development due to pumping action [5]

3.1.4 Influencing factors

Moisture susceptibility is a phenomenon which depends upon the mechanisms described above, hence its complexity. Their interaction makes it difficult to predict with certainty the importance of a particular characteristic as a factor in determining moisture susceptibility. A general rule suggests that *“moisture susceptibility is increased by any factor that increases moisture content in the HMA, decreases the adhesion of asphalt binder to the aggregate surface or physically scours the asphalt binder”* [6]. The factors described below have an influence on moisture susceptibility, but none of them are fully definitive for predicting it. They refer to the mixture design and construction characteristics, but not climatic or traffic conditions.

Asphalt binder characteristics: Viscosity is an important property of bitumen because it may be an indicator of higher asphaltenes concentrations, which can create higher adhesion tension and molecular orientation adhesion. For this reason, lower viscosities, and consequently lower asphaltenes concentrations, are in general more susceptible to stripping. Other components in asphalt binders such as sulfoxides, carboxylic acids, phenols and nitrogen bases can also potentially lead to stripping [6].

Aggregate characteristics: Hydrophilic aggregates (attract water) are more prone to strip than hydrophobic aggregates (repulse water). The key properties that determine this characteristic are the surface chemistry (acidic aggregate surfaces are more susceptible to stripping), porosity and pore size; high porosity leads to high absorption and more asphalt binder has to be used to achieve the desired effective binder content. If this is not considered, not sufficient binder will be available for the creation of the film around aggregate particles, resulting in faster aging and stripping [6].

Air voids: When air voids exceed about 8% of the volume, they will possibly become interconnected and allow water to penetrate with ease. In the case where the water stays in these voids, moisture damage will happen by the water-bitumen interaction. In the case of non-interconnected voids, the ‘trapped’ water in the voids will cause moisture damage through pore pressure or ice expansion. For this reason, mix design has to adjust binder content and aggregate gradation, to achieve the desirable void content. Construction stage also defines this factor. Inadequate compaction will result in lower density levels, meaning that more voids than the designed remain in the mixture’s structure. Poor compaction can be caused either by not well executed compaction plan, or by cool weather condition during the construction [6].

3.2 Permanent Deformation

3.2.1 Definition

Permanent deformation in a pavement layer is a very common phenomenon and causes the development of ruts along the wheel path at the surface. Hence, when talking about permanent deformation, we talk in terms of rutting.

Rutting is defined as a longitudinal surface depression occurring in the wheel paths of roadways. It is often followed in later stages by an upheaval along the sides of the rut [8]. Rutting accumulates incrementally with small permanent deformations from each load application (i.e. each wheel pass) over the pavement's service life and is by definition a load-related pavement distress. It is a high temperature phenomenon, i.e. most often occurs during the summer when high pavement temperatures occur. Its importance in the pavement performance lies in the fact that it can lead to functional failure and potential danger from hydroplaning [9].



Picture 3.3 Rutting due to weak subgrade (left), and rutting due to weak asphalt layer [10] (right)

3.2.2 Types of permanent deformation

There are three types of rutting that are distinguished by the cause and the layer in which they appear.

a) One-dimensional densification or vertical compression

A depression near the centre of the wheel path without an accompanying hump on either side of the depression is caused due to material densification. This densification is generally caused by excessive air voids or inadequate compaction after the placement of the asphalt layer. In this way the material is

allowed to further compact when it is subjected to traffic load. This type of rutting usually results in a low to moderately severe levels of rutting [11].

b) Lateral flow or plastic movement

This type is caused by the localized shear failure by overstressing the pavement with high tire pressure. A depression near the centre of the wheel path with humps on either side of the depression is caused by lateral flow. It occurs in mixtures with inadequate shear strength or an insufficient amount of total voids in the asphalt layer. In such cases lateral flow occurs because the low voids allow the asphalt to act as a lubricant rather than a binder. It is higher at higher temperatures, and less on highways with higher speeds due to the visco-elastic behaviour of asphalt. This type of rutting usually results in moderate to highly severe levels of rutting and is most difficult to predict [11].

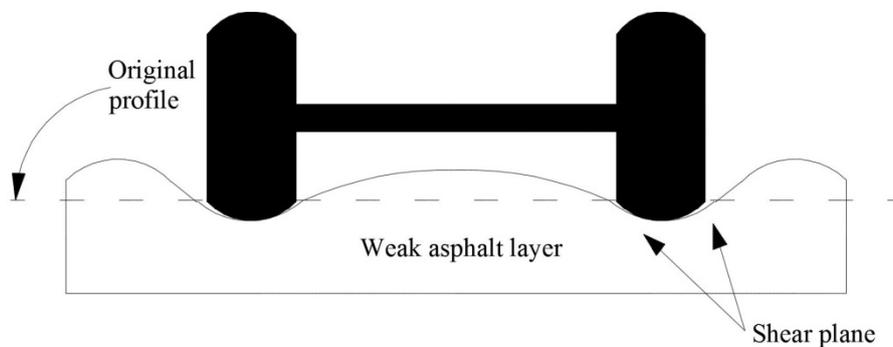


Figure 3.1 Rutting caused by weak asphalt layer [12]

c) Mechanical deformation

This third type of rutting is related to the unbound materials below the asphalt surface and their consolidation, densification, and/or lateral movement. It is a result of subsidence in the base, subbase or subgrade and is usually accompanied by a longitudinal cracking pattern at the pavement's surface, in the case of very stiff mixtures. These longitudinal cracks generally occur in the centre and along the outside edges of the ruts [11].

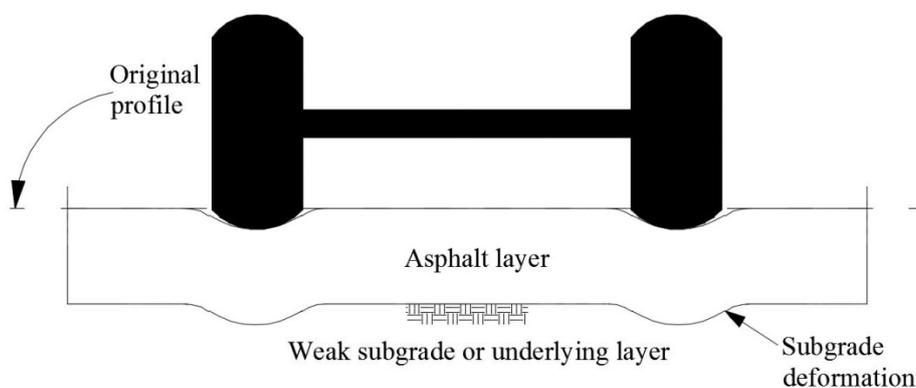


Figure 3.2 Rutting caused by weak subgrade [12]

The interest of this research is limited to the first two types of permanent deformation, vertical compression and lateral flow, since the effect of a weak subgrade cannot be looked into by the tests carried out.

3.2.3 Development stages

Rutting in asphalt layers develops in three stages (Figure 3.3):

- *Primary (initial) stage* is related to the deformation caused by traffic compaction (densification, volume reduction) at the early stages of the pavement's service life (usually within the first year).
- *Secondary (middle) stage* is considered to be representative of the pavement's deformation behaviour for the greater part of its lifetime. Rutting rate is constant and is caused by horizontal and vertical traffic loads resulting in shear stresses in asphalt.
- *Tertiary (last) stage* is characterized by accelerated rutting and excessively rapid plastic deformations [13].

The most common practice is rehabilitating the pavement prior to reaching the tertiary stage, since at that point rutting has already reached the regulation's threshold or another distress triggers the need for maintenance. For this reason, rutting modelling omits the last stage and is restricted to the secondary stage [9].

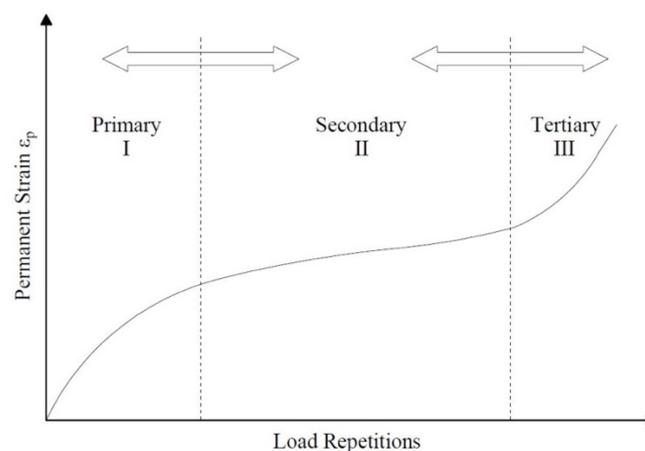


Figure 3.3 Rutting stages

3.2.4 Influencing factors

The permanent deformation of asphaltic mixtures is a complex phenomenon where the contribution of various components like the properties of the aggregates, bitumen, contact of aggregates and bitumen, etc. make up the overall performance. These properties are not constant but they are changing through time to the end of the pavement's service life, i.e. till the failure due to excessive permanent

deformation or cracking. An overview of the various factors affecting the permanent deformation as well as effects of their changes are given in Table 3.1.

Table 3.1 Factors affecting rutting of asphalt mixtures [14]

Factor		Change in Factor	Effect of Change in Factor on rutting Resistance
Aggregate	Surface texture	Smooth to rough	Increase
	Gradation	Gap to continuous	Increase ¹⁾
	Shape	Rounded to angular	Increase
	Size	Increase in maximum size	Increase ²⁾
Binder	Stiffness ³⁾	Increase	Increase
Mixture	Binder content	Increase	Decrease
	Air void content ⁴⁾	Increase	Decrease
	Voids in the mineral aggregate ⁵⁾	Increase	Decrease
	Method of compaction	- ⁶⁾	- ⁶⁾
Test or field conditions	Temperature	Increase	Decrease
	State of stress/strain	Increase in tire contact pressure	Decrease
	Load repetitions	Increase	Decrease
	Water	Dry to Wet	Decrease if mixture is water sensitive

¹⁾In the Netherlands this is not necessary the case, because in the Porous Asphalt (Gap graded), rutting resistance is higher than more continuous mixtures

²⁾Assuming constant layer thickness.

³⁾Refers to stiffness at temperature at which rutting propensity is being determined. Modifiers may be utilized to increase stiffness at critical temperatures, thereby reducing rutting potential.

⁴⁾When air void content is less than about 3%, the rutting potential of mixture increases.

⁵⁾It is argued that very low (i.e. less than 10%) voids in the mineral aggregate should be avoided.

⁶⁾The method of compaction, whether laboratory or field, may influence the structure of the system and therefore the propensity for rutting

Apart from the general overview presented in the previous section, the bitumen properties and their relation to rutting were investigated in more detail. In particular, the bitumen stiffness expressed as G^* and the zero shear viscosity (ZSV) expressed by η_0 were considered as potential parameters to be included in the rutting prediction.

- Bitumen Stiffness

Stiffer binders at high service temperatures have less rutting. In general, the stiffer the asphalt binder, the stiffer the mixture and the more resistant to permanent deformation [15] (high G^* produces mixtures less susceptible to rutting [16]), which was also what was initially expected when considering the addition of this parameter. In the case of NL-LAB it was decided to work with the bitumen's stiffness at a low frequency level, passes because it's a good indication of the behaviour at

the low frequency conditions. Hence the values at 0.1 rad/s (0.016 Hz) were used. The master curves where these values were taken from were constructed by DSR tests at 40°C performed by TNO.

- Zero Shear Viscosity

Looking into various studies regarding ZSV, the main conclusion was that ZSV (η_0) is a suitable indicator to evaluate the contribution of the bituminous binder to the rutting resistance of the asphalt pavement layers [17]. In particular according to a research [18], a good correlation between rutting rate and η_0 was found for all the binders tested including unmodified and polymer-modified bitumen. The advantage of ZSV comparing to G^* is that there is an apparent inability of G^* to capture the contribution to rutting resistance afforded by polymer modification [19]. In the case of pure binders, the correlation between this indicator and results from rutting tests on asphalt mixes is good. For Polymer modified binders on the other hand, it generally underestimates the resistance to rutting [20].

Literature regarding rutting and binder properties suggests that in order to characterize the rheological behaviour of a thermoplastic material in a certain temperature range, at least two properties should be estimated:

- Consistency at a certain temperature (e.g. penetration Pen_{25} at 25°C) or in a certain rheological state (e.g. softening point $T_{R\&B}$ or T_{800}) and
- Temperature susceptibility (PI), or in the case of NL-LAB, $\log A$, which is interrelated with PI.

$$PI = \frac{20 \cdot (1 - 25 \cdot A)}{1 + 50 \cdot A} \quad [-] \quad \text{Eq. (3.1)}$$

where,

- PI is the penetration index of bitumen (-);
 A is the temperature susceptibility of bitumen (-);

$$A = \frac{\log Pen_{25} - \log 800}{25 - T_{R\&B}} \quad [-] \quad \text{Eq. (3.2)}$$

where,

- Pen_{25} is the penetration value of bitumen at 25°C expressed in 0.1 mm (Pen units);
 $T_{R\&B}$ is the softening point of bitumen determined by the ring and ball test, expressed in temperature degrees (°C);

However, when using these parameters in a regression to predict a mixture's rutting behaviour, all three properties (Pen_{25} , $T_{R\&B}$ and A) shall not be included at the same time due to their interrelation. Thus only two of them might be included.

3.3 Analytical Tools

3.3.1 Regression Analysis

The second research objective of this thesis intends to provide relations that “simulate” Type Testing by predicting the Type Test results (functional properties), based on composition parameters. The process of generating these relations from a database is called Regression Analysis. It is a powerful tool used in researches of every scientific field, from psychology and medicine, to marketing and engineering. The statistical package that was used for this purpose is SPSS Statistics.

There are various forms of regression analysis. The most common and simple is the simple linear regression. It assesses the probability that a linear relationship exists between two continuous variables (infinite number of possible values; opposite of a discrete variable which can only take on a certain number of values) to predict the value of a dependent variable based on the value of an independent variable. More specifically, it will let the researcher:

- (a) determine whether the linear regression between these two variables is statistically significant;
- (b) determine how much of the variation in the dependent variable is explained by the independent variable;
- (c) understand the direction and magnitude of any relationship;
- (d) predict values of the dependent variables based on different values of the independent variable.

In this thesis, the type of regression used is multiple linear regression. A multiple regression is used to predict a continuous dependent variable (like Permanent deformation, Moisture sensitivity) based on multiple independent variables (Volumetric properties, bitumen properties, etc. of the mixture). As such, it extends the simple linear regression, which is used when there is only one continuous independent variable. Multiple regression also allows to determine the overall fit (the amount of variance in the data explained by the model) of the model and the relative contribution of each of the predictors (or independent variables) to the total variance explained.

In order to run a multiple regression, there are eight assumptions that need to be considered. The first two assumptions relate to the choice of study design and the measurements done in this study, whilst

the other six assumptions relate to how the data fits the multiple regression model. These assumptions are [21]:

Assumption 1: The dependent variable is measured at the continuous level (i.e., the interval or ratio level). Examples of continuous variables include height (measured in centimeters), temperature (measured in °C), weight (measured in kg). Examples of discrete variables (also called categorical or nominal) include condition (e.g., two categories: dry and wet), production phases (e.g., three categories: F1, F2 and F3).

Assumption 2: There are two or more independent variables that are measured either at the continuous or at the categorical level (examples of such variables are listed in Assumption 1). An independent variable with only two categories is known as a dichotomous variable whereas an independent variable with three or more categories is referred to as a polytomous variable.

Assumption 3: Independence of observations (i.e., independence of residuals). The assumption of independence of observations in a multiple regression is designed to test for 1st-order autocorrelation, which means that adjacent observations are correlated (i.e., not independent). This is largely a study design issue because the observations in a multiple regression must not be related. In SPSS Statistics, independence of observations can be checked using the Durbin-Watson statistic.

Assumption 4: There needs to be a linear relationship between (a) the dependent variable and each of the independent variables, and (b) the dependent variable and the independent variables collectively. The assumption of linearity in a multiple regression needs to be tested in two parts. Firstly, establish if a linear relationship exists between the dependent and independent variables collectively, which can be achieved by plotting a scatterplot of the studentized residuals against the (unstandardized) predicted values. Secondly, establish if a linear relationship exists between the dependent variable and each of the independent variables, which can be achieved using partial regression plots between each independent variable and the dependent variable.

Assumption 5: The data needs to show homoscedasticity of residuals (equal error variances). The assumption of homoscedasticity is that the residuals are equal for all values of the predicted dependent variable (i.e., the variances along the line of best fit remain similar as you move along the line). To check for heteroscedasticity, we can use the plot created to check linearity in the previous section, namely plotting the studentized residuals against the unstandardized predicted values, which were produced as part of the multiple regression procedure on the previous page, as explained above.

Assumption 6: The data must not show multicollinearity. Multicollinearity occurs when you have two or more independent variables that are highly correlated with each other. This leads to problems with understanding which independent variable contributes to the variance explained in the dependent variable, as well as technical issues in calculating a multiple regression model. Multicollinearity is checked through an inspection of correlation coefficients and Tolerance/VIF values.

Assumption 7: There should be no significant outliers, high leverage points or highly influential points. Outliers, leverage and influential points are different terms used to represent observations in a data set that are in some way unusual when we wish to perform a multiple regression analysis. These different classifications of unusual points reflect the different impact they have on the regression line. An observation can be classified as more than one type of unusual point. However, all these points can have a very negative effect on the regression equation that is used to predict the value of the dependent variable based on the independent variables. Outliers are detected using casewise diagnostics and studentized deleted residuals; and influential points using a measure of influence known as Cook's Distance.

Assumption 8: Check that the residuals (errors) are approximately normally distributed. In order to be able to run inferential statistics (i.e., determine statistical significance), the errors in prediction – the residuals – need to be normally distributed. Two common methods used to check for the assumption of normality of the residuals are: (a) a histogram with superimposed normal curve and a P-P Plot; or (b) a Normal Q-Q Plot of the studentized residuals [21].

In order to check these assumptions, it is first needed to run the standard multiple regression procedure. This is because many of these assumptions are checked by inspection of the residuals, which can only be calculated once a regression line has been fitted/generated. Since it is not uncommon for a dataset to violate (i.e., fail) one or more of these assumptions, there are various steps that can be followed after that. This could include (a) making corrections to the data so that it no longer violates the assumptions, (b) using an alternative statistical test, or (c) proceeding with the analysis even when the data violates certain assumptions.

The next step is to interpret the regression results and evaluate the quality of the fit. If the desirable levels are reached, that particular relation can be used for the prediction. If not, another model with different independent variables is chosen and checked again. There are a number of statistics that can be used to determine whether the multiple regression model is a good fit for the data. These are: (a) the multiple correlation coefficient, (b) the percentage (or proportion) of variance explained; (c) the

statistical significance of the overall model; and (d) the precision of the predictions from the regression model.

(a) Multiple correlation coefficient

The multiple correlation coefficient (which is denoted by R , and is the square root of R^2 explained below), is simply the Pearson correlation coefficient between the scores predicted by the regression model and the actual values of the dependent variable. As such, R is a measure of the strength of the linear association between these two variables and can give an indication as to the goodness of the model fit with a value that can range from -1 to 1, with higher absolute values indicating a stronger linear association. A multiple correlation coefficient of 0 (zero) indicates no linear association between the dependent variable and the independent variables and a value of 1 a perfect linear association.

(b) Total variation explained

The coefficient of determination – more commonly known as R^2 – is a measure of the proportion of variance in the dependent variable that is explained by the independent variable. More specifically, it is the proportion of variance in the dependent variable that is explained by the independent variables over and above the mean model. Having for example and R^2 equal to 0,577, means that the addition of the independent variables into a regression model explained 57.7% (i.e., $0.577 \times 100 = 57.7\%$) of the variability of our dependent variable.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad \text{Eq. (3.3)}$$

$$SS_{res} = \sum_i (y_i - f_i)^2 \quad \text{Eq. (3.4)}$$

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \quad \text{Eq. (3.5)}$$

where,

SS_{res} is the sum of squares of the residuals;

SS_{tot} is the total sum of squares (proportional to the variance of the data);

y_i is the observed value;

\bar{y} is the mean of the observed data;

f_i is the predicted value;

However, R^2 is based on the sample and is considered a positively-biased estimate of the proportion of the variance of the dependent variable accounted for by the regression model (i.e., it is larger than it should be when generalizing to a larger population). There is another measure called *adjusted* R^2 which corrects for this positive bias in order to provide a value that would be expected in the population.

$$R_{adj}^2 = 1 - \left[\frac{(1 - R^2)(n - 1)}{n - k - 1} \right] \quad \text{Eq. (3.6)}$$

where,

- R^2 is coefficient of determination;
- n is the number of points in the data sample;
- k number of independent regressors, i.e. the number of variables in the model, excluding the constant

(b) Statistical significance of the model

The statistical significance of a model is measured with the p-value. If $p < 0,05$, we have a statistically significant result. This means that the addition of the independent variables used leads to a model that:

- (i) is statistically significantly better at predicting the dependent variable than the mean model; and
- (ii) is a statistically significantly better fit to the data than the mean model.

On the other hand, if $p > 0,05$, we do not have a statistically significant result [21].

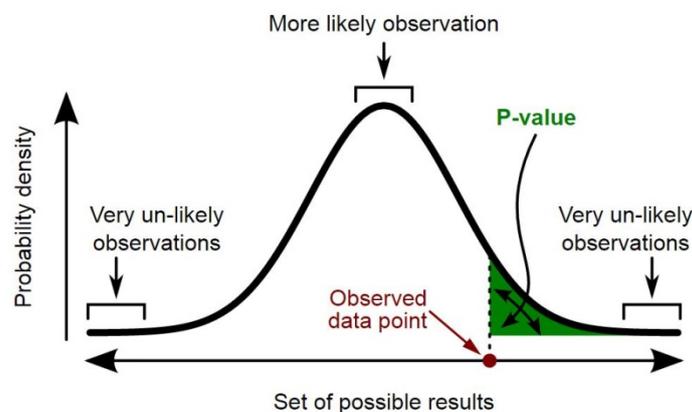


Figure 3.4 Example of a p-value computation. The vertical coordinate is the probability density of each outcome, computed under the null hypothesis. The p-value is the area under the curve past the observed data point [22].

A p-value (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true. The computation of p-value is done using statistical software, often via numeric methods [22].

3.3.2 Box-plots

The comparison of the performance of the specimens produced in the lab and the field will be facilitated through what is known in descriptive statistics as Box-Plot, or Box-and-Whiskers-Diagram. It is a convenient way of depicting groups consisting of numerical data, graphically, through their quartiles. Box plots have lines vertically extending from the boxes, that indicate the variability outside the upper and lower quartiles. Box plots are non-parametric, meaning that they display variation in samples of a statistical population, with no assumptions regarding the underlying statistical distribution.

The bottom and top of the box represent the first and third quartiles (Q1, Q3), and the band inside the box the second quartile (Q2, the median). The ends of the whiskers depict the minimum and maximum of all data. Possible outliers are plotted as individual points [23]. As far as quartiles are concerned, they are three points that divide data set into four equal groups, each group comprising a quarter of the data. The first quartile (Q1) is defined as the middle number between the smallest number and the median of the data set. The second quartile (Q2) is the median of the data. The third quartile (Q3) is the middle value between the median and the highest value of the data set. In other words:

Table 3.2 Definition of quartiles [24]

Symbol	Name	Definition
$Q1$	First Quartile	Splits off the lowest 25% of the data, from the highest 75%
$Q2$	Second Quartile, Median	Cuts dataset in half
$Q3$	Third Quartile	Splits off the highest 25% of the data, from the lowest 75%

An example of the components described is seen in a representative boxplot of the air void values generated by SPSS, in Figure 3.5.

Also, in order to demonstrate their use, two auxiliary datasets with the same mean value, but with different distribution around it, are plotted in two boxplots in Figure 3.6. In the first variable, the values are evenly and closely distributed around the mean, whereas in variable number two, the values are scattered away from the mean, which is still the same with the first variable.

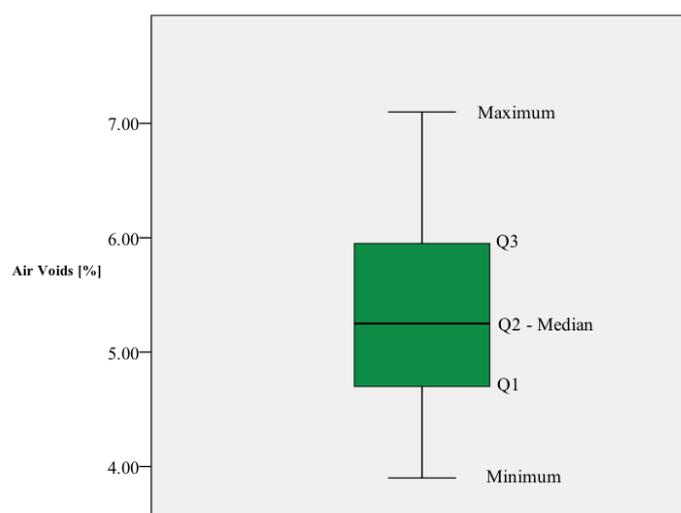


Figure 3.5 Example of boxplot's components

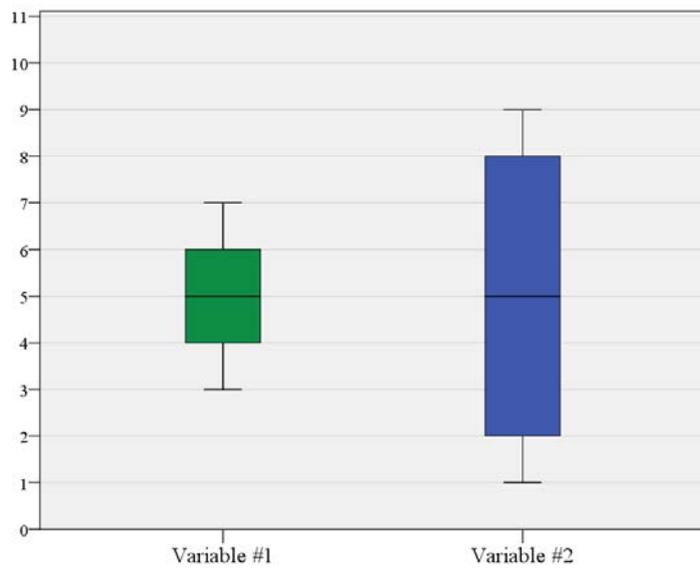


Figure 3.6 Example of two datasets with the same mean value, but different distributions

Variable #1	Variable #2
3,0	1,0
3,5	1,0
4,0	2,0
4,0	2,0
5,0	3,0
5,0	8,0
6,0	8,0
6,0	8,0
6,5	9,0
7,0	9,0
Mean	
5,0	5,0
Standard Deviation	
1,35	3,46

4 LABORATORY STAGE

The description of the standards and test procedures followed during the course of this thesis, along with important notes and observations, are presented in this chapter. Also, information regarding the specimens' and mixture specifications, how they were produced, stored and tested.

4.1 Water Sensitivity

4.1.1 Introduction

The water sensitivity of asphalt concrete, as it is suggested by the name, expresses the effect of the interaction between water and the asphaltic material, on the material properties. As it was described in the literature review in paragraph 3.1, the presence of moisture in the asphalt concrete's mass can lead to various distress types that relate to loss in stiffness and strength.

The principle of the methods that are employed to evaluate this sensitivity is that one subset of specimens is considered as the dry, unconditioned reference, where the effect of water is excluded, and one or more subsets of specimens are considered in a wet, conditioned state, where the material is subjected to damages, or in general, changes due to the presence of moisture. The way the wet subsets are conditioned differs from method to method, depending on the final use and target of that method. However, all of the methods evaluate this sensitivity on the basis of Indirect Tensile Strength Ratio (ITSR). The various conditioned subsets are tested by means of Indirect Tensile Strength after their conditioning, and each one is compared to the dry reference's strength, giving the ITSR:

$$ITSR = \frac{ITS_{conditioned}}{ITS_{unconditioned}} \times 100 \quad [\%] \quad \text{Eq. (4.1)}$$

where,

$ITSR$ is the indirect tensile strength ratio, in percent (%);

$ITS_{conditioned}$ is the average indirect tensile strength of the conditioned group, in kilopascals (MPa);

$ITS_{unconditioned}$ is the average indirect tensile strength of the unconditioned group (reference), in kilopascals (MPa);

In order to cover all the aspects of the experimental analysis and be able then to interpret and study the results, a number of information is necessary to be acquainted. This information ranges from organizational aspects to detailed practical information on the tests, and are entailed in the next paragraphs.

4.1.2 Specimens

4.1.2.1 Organization

In total, 44 samples with a target diameter of 100 mm and height 50 mm were provided by the Dutch contractor *Dura Vermeer Groep NV*, for the purposes of this thesis. These specimens were designed to be constructed according to the Type Testing specifications of Paragraph 1.2.5. However, due to a misunderstanding, the compaction method differed from the prescribed, and instead was made in the way described in Paragraph 0. The coding of the specimens was kept as it was defined by the contractor, to maintain a consistency throughout the project and help the back-tracking of any information needed.

As mentioned in the thesis' organization, the water sensitivity of the mixtures will be tested in three different ways regarding the conditioning process, each one with its specified number of specimens. An overview of the specimen's division in the different protocols, is seen on the tables in paragraph 4.1.7. Table 4.6 refers to the specimens conditioned under the Standard Method's protocol (NEN-EN 12697-12:2008 Method A). The dry specimens are the ones that were used as a reference for the unconditioned subset, both for the standard method and for the MIST. Table 4.8 refers to the specimens that underwent the MIST conditioning protocol, and Table 4.7 to the specimens conditioned according to the Frost Damage protocols. The protocol columns refers to the different protocols specified by the Frost Damage guidelines, and they are described in detail in paragraph 4.1.5.

4.1.2.2 Production and Storage

The production of the specimens, both in lab and in the field, took place on approximately the same period. Phase 1 and 2 slabs were produced in the labs of *Dura Vermeer*, whereas phase 3 slabs were extracted from the construction site of Highway A77 in the Netherlands, which was constructed in November 2016. The specific dates of production, coring and sawing for each phase are see on Table 4.1. After the production they were stored in the contractor's facilities and were incrementally

measured in their volumetric properties, before they were delivered at TU Delft, on the 26th of January. During the waiting periods in-between the tests and the conditionings, all the specimens were stored at a controlled temperature and humidity climate chamber, at the faculty's lab, to control the ageing process and keep all of them at the same ageing stage (Picture 4.1). The temperature was maintained at $13\pm 1^\circ\text{C}$ and the relative humidity at approximately $42\pm 1\%$.

Table 4.1 Key dates in specimens' production

Phase	Production	Coring	Sawing
F1	16/12/2016	11/01/2017	11/01/2017
F2	11/11/2016	17/11/2016	17/11/2016
F3	11/11/2016	03/12/2016	07/12/2016



Picture 4.1 Specimens storing conditions at TU Delft

4.1.2.3 Mixture composition

The design of the mixture's composition was done by Dura Vermeer to conform with the requirements described for each test in *RAW 2015*. For all three phases of production, the same mixture is used; an Asphalt Concrete of Nominal Aggregate Size 16 mm for a Binder/Base layer, with 60% of Reclaimed Asphalt (AC16 OL/TL 60% PR). The target density of the mixture was 2380 kg/m^3 .

According to the contractor, the mixture composition was checked three times per phase. Firstly, extracting the composition from a non-compacted mixture directly out of the mixer (uit menger), secondly after a gyratory compaction of a non-treated sample (proefdraai onbew.), and finally after a gyratory compaction on a sample treated in such a way that its dimensions correspond to the requirements for an ITSR sample, as specified by *RAW 2015* (ITSR proefstuk). Since the norms do

not give guidelines for the extraction of composition and bitumen from Triaxial samples, they were treated in the same way as the ITSR samples. The composition considered for the calculations in the analysis of this thesis, is the third, referring to the ITSR treated sample. The detailed mixture compositions, in all three phases, as they were determined in the three different ways, are seen in the following tables.

Table 4.2 Phase 1 mixture composition in % (m/m)

	From mixer	Untreated - compacted	ITSR treated
Through sieve C31,5	100	100	100
C22,4	100	100	100
C16	89	94	96
C11,2	68	74	76
C8	56	61	65
C5,6	48	52	56
2 mm	35	38	41
500 µm	26	28	30
180 µm	11	12	13
125 µm	8	8	9
63 µm	5,9	5,8	6,6
Bitumen "in"	4,0	3,9	4,1

Table 4.3 Phase 2 mixture composition in % (m/m)

	From mixer	Untreated - compacted	ITSR treated
Through sieve C31,5	100,0	100,0	100,0
C22,4	100,0	100,0	100,0
C16	94,3	90,9	93,0
C11,2	73,8	74,6	77,4
C8	65,0	64,0	64,9
C5,6	59,3	57,3	58,5
2 mm	49,7	47,2	48,5
500 µm	37,8	35,7	36,7
180 µm	12,9	12,6	12,9
125 µm	8,6	8,4	8,6
63 µm	6,0	5,8	6,2
Bitumen "in"	3,9	3,9	3,8

Table 4.4 Phase 3 mixture composition in % (m/m)

	From mixer	Untreated - compacted	ITSR treated
Through sieve C31,5	100,0	100,0	100,0
C22,4	100,0	100,0	96,7
C16	94,6	93,2	91,8
C11,2	78,6	73,4	74,6
C8	67,3	63,0	62,0
C5,6	61,0	55,4	55,7
2 mm	50,7	46,5	46,2
500 µm	38,6	35,5	35,2
180 µm	13,3	12,5	12,5
125 µm	9,1	8,5	8,5
63 µm	6,5	6,1	6,1
Bitumen "in"	4,0	3,8	3,7

It can be seen that the bitumen percentage slightly differs from phase to phase, being 4,1% in phase 1, 3,8% and 3,7% for phases 2 and 3 respectively. This percentage is the final sum of the already existing bitumen in the reclaimed asphalt and the fresh bitumen added to the mixture. The resulting aggregate gradation curves from the ITSR treated percentages are very close to each other and can be seen in

following figure (Figure 4.1). Figure 4.2 and Figure 4.2 show the gradation curves resulting from the ‘From mixer’ and ‘Untreated – compacted’ percentages. The various aggregate types used by the contractor to achieve this gradation, along with the bitumen and reclaimed asphalt type, are listed in Table 4.5, expressed in 100% of the mixture’s mass.

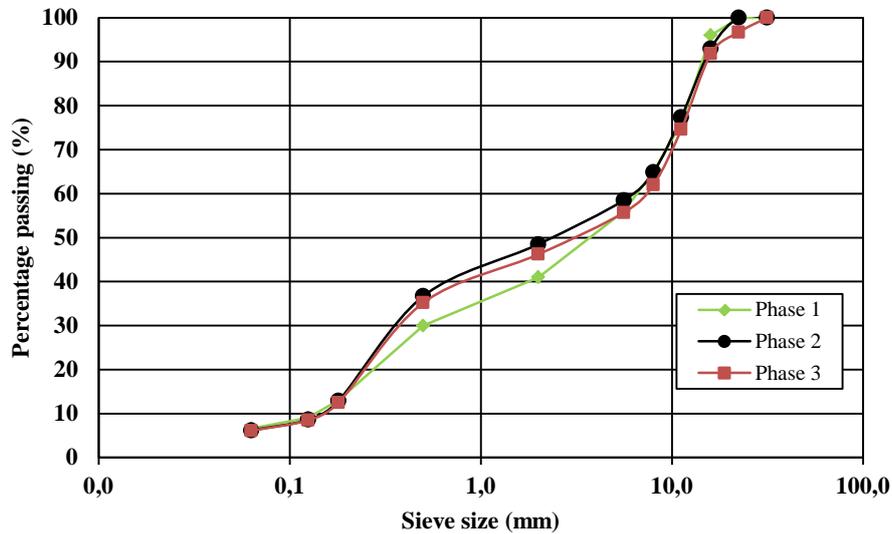


Figure 4.1 Different phases aggregate gradation curves (ITSR treated)

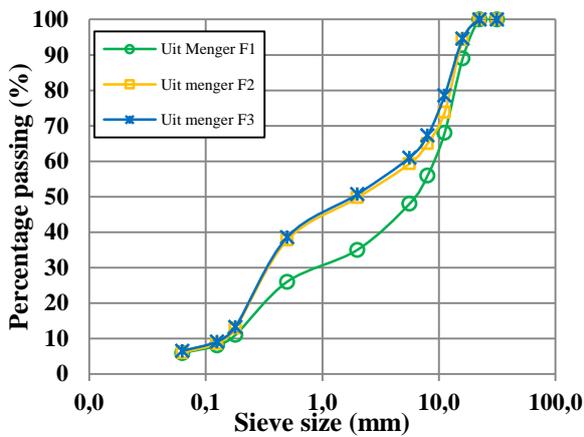


Figure 4.2 Different phases aggregate gradation curves (From mixer)

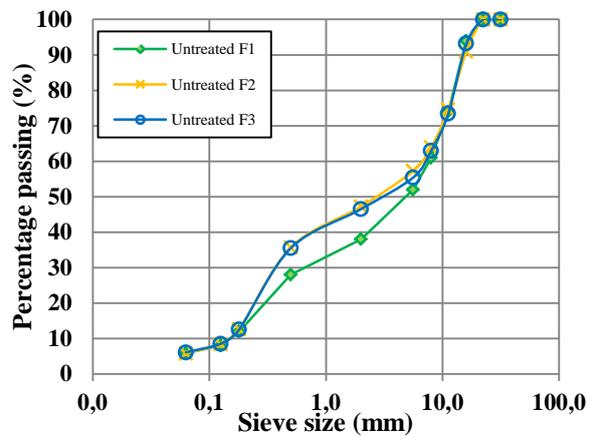


Figure 4.3 Different phases aggregate gradation curves (untreated)

Table 4.5 Mixture materials

Material	% “In”
Scottish granite	25,0
Fine natural sand	13,6
Onderlaag PR 0/20	60,0
Bitumen 160/220	1,4
Total	100,0

4.1.2.4 Compaction

For the specimens whose compaction took place in the lab, phase 1 and 2 specimens, a compaction device was used by Dura Vermeer to compact the slabs constructed for this purpose. This is different than usual in Type Testing, where cores are made using gyratory compaction. This difference arose from a misunderstanding, which is unfortunate for this specific project but has advantages in the overall NL-LAB scheme because it provides new variations. Specifically, it was a device made by *Baustoff-Prüfsysteme Wennigsen GmbH*, with a height controlled at 90 mm, as seen in the picture.



Picture 4.2 Compression device used for the slabs' compaction

4.1.2.5 Volumetrics

By visual examination, all the specimens appeared to be symmetrical with the curved side even and circular, hence none of them was excluded from the sets. The dimensions of the samples were determined by the contractor according to NEN-EN 12697-29:2002, taking for the diameter 2 perpendicular measurements at the top, middle, and bottom of the specimen, getting the average of the 6 measurements to the nearest 0,1 mm. The height was taken as the average of 4 evenly spaced measurements around the specimen's perimeter, 10 mm in from the edge, to the nearest 0,1 mm.

The bulk density (ρ_b) of each core was calculated by the contractor in accordance with NEN-EN 12697-6:2012, following both Procedure B (SSD) and D (by dimensions). For this purpose, besides the dimensions already measured, the dry, submerged, and saturated surface-dried weights were recorded for each specimen, at a certain water temperature. As advised by Annex A3 of the standard, Procedure B is suitable for measuring the bulk density of dense-graded bituminous specimens having a low water absorption level. The applicability of this procedure is related to the voids level: for continuously graded materials, such as asphalt concrete with void contents up to 5%, which corresponds to our material. Procedure D on the other hand, is said to be more suitable for void contents greater than 15%. For this reason, $\rho_{b,SSD}$ was used, and will be noted as just ρ_b in all the following calculations.

In order to calculate the void content (V_m) of the cores, the maximum density (ρ_m) was measured according to NEN-EN 12697-5:2009, following Procedure A (Volumetric procedure), again by the contractor. In this case, the maximum density was calculated for each different mixture, which means

one maximum density per phase. For each phase the maximum density was calculated as the average of 4 different measurements. The air void content of each specimen was then calculated following NEN-EN 12697-8:2003:

$$V_m = \frac{\rho_m - \rho_b}{\rho_m} \times 100 \quad [%(v/v)] \quad \text{Eq. (4.2)}$$

In the following tables (Table 4.6, Table 4.7, and Table 4.8), an overview of the average dimensions, the bulk density, the maximum density and the void content of each specimen can be seen, distinguished by conditioning protocol.

Table 4.6 Standard conditioning method specimens' volumetric properties

Core	Phase	Protocol	Height mm	Diameter mm	ρ_b kg/m ³	ρ_m kg/m ³	V_m %
40	F1	Dry	49,7	101,4	2398,1	2506,0	4,31
45	F1	Dry	49,0	101,4	2367,6	2506,0	5,52
47	F1	Dry	49,5	101,5	2392,4	2506,0	4,53
49	F1	Wet	50,1	101,5	2391,0	2506,0	4,59
52	F1	Wet	49,6	101,4	2375,9	2506,0	5,19
56	F1	Wet	49,8	102,0	2386,6	2506,0	4,76
1238	F3	Dry	49,9	101,5	2363,7	2492,0	5,15
1241	F3	Dry	49,9	101,5	2351,8	2492,0	5,62
1246	F3	Dry	50,4	101,5	2346,5	2492,0	5,84
1252	F3	Wet	49,5	101,5	2347,5	2492,0	5,80
1257	F3	Wet	50,2	101,4	2356,3	2492,0	5,45
1262	F3	Wet	49,5	101,4	2359,7	2492,0	5,31

Table 4.7 Frost damage conditioning protocol specimens' volumetric properties

Core	Phase	Protocol	Height mm	Diameter mm	ρ_b kg/m ³	ρ_m kg/m ³	V_m %
59	F1	I	49,9	101,5	2404,0	2506,0	4,07
61	F1	I	50,6	101,4	2359,7	2506,0	5,84
62	F1	I	49,4	101,6	2382,4	2506,0	4,93
63	F1	II	50,7	101,4	2389,2	2506,0	4,66
64	F1	II	50,2	101,5	2375,7	2506,0	5,20
66	F1	II	50,6	101,4	2390,0	2506,0	4,63
67	F1	III	50,0	101,4	2388,1	2506,0	4,70
68	F1	III	50,5	101,4	2384,0	2506,0	4,87
70	F1	III	50,7	101,4	2396,8	2506,0	4,36
72	F1	IV	50,7	101,3	2382,4	2506,0	4,93
75	F1	IV	49,6	101,3	2376,4	2506,0	5,17
79	F1	IV	49,9	101,4	2387,9	2506,0	4,71
1266	F3	I	49,7	101,4	2331,3	2492,0	6,45
1270	F3	I	49,5	101,5	2343,1	2492,0	5,98
1273	F3	I	50,6	101,4	2341,0	2492,0	6,06
1278	F3	II	50,1	101,5	2343,2	2492,0	5,97
1283	F3	II	49,7	101,5	2326,2	2492,0	6,65
1286	F3	II	49,7	101,5	2330,3	2492,0	6,49
1291	F3	III	50,2	101,4	2319,7	2492,0	6,92
1295	F3	III	49,8	101,5	2315,6	2492,0	7,08
1299	F3	III	49,9	101,5	2360,6	2492,0	5,27
1302	F3	IV	50,5	101,6	2367,5	2492,0	5,00
1309	F3	IV	49,9	101,5	2349,8	2492,0	5,71
1311	F3	IV	50,1	101,5	2357,1	2492,0	5,41

Table 4.8 MIST conditioning protocol specimens' volumetric properties

Core	Phase	Protocol	Height mm	Diameter mm	ρ_b kg/m ³	ρ_m kg/m ³	V_m %
34	F1	MIST	49,4	101,3	2391,7	2506,0	4,56
35	F1	MIST	49,1	101,4	2396,7	2506,0	4,36
36	F1	MIST	49,9	101,3	2391,2	2506,0	4,58
38	F1	MIST	49,8	101,5	2407,2	2506,0	3,94
1218	F3	MIST	49,2	101,5	2323,1	2492,0	6,78
1222	F3	MIST	49,7	101,6	2343,8	2492,0	5,95
1227	F3	MIST	49,8	101,5	2340,9	2492,0	6,07
1231	F3	MIST	49,7	101,5	2332,8	2492,0	6,39

Apart from these volumetric characteristics, the void content in the mineral aggregate (VMA) and the percentage of the voids in the mineral aggregate filled with binder (VFB) were calculated because of their possible use in the calculations to be done. Both quantities are described in NEN-EN 12697-8:2003 and were calculated for each core as follows:

$$VMA = V_m + B \cdot \frac{\rho_b}{\rho_B} \quad [%(v/v)] \quad \text{Eq. (4.3)}$$

where,

- VMA is the voids content in the mineral aggregate in 0,1% (v/v);
- V_m is the air voids content of the specimen in 0,1% (v/v);
- B is the binder content of the specimen (in 100% of the mixture), in 0,1% (v/v);
- ρ_b is the bulk density of the specimen, in 1 kilogram per cubic metre (kg/m³);
- ρ_B is the density of the binder, in 1 kilogram per cubic metre (kg/m³)

$$VFB = \left(\left(B \cdot \frac{\rho_b}{\rho_B} \right) / VMA \right) \times 100 \quad [%(v/v)] \quad \text{Eq. (4.4)}$$

where,

- VMB is the percentage of the voids in the mineral aggregate filled with binder in 0,1% (v/v);
- B is the binder content of the specimen (in 100% of the mixture), in 0,1% (v/v);
- ρ_b is the bulk density of the specimen, in 1 kilogram per cubic metre (kg/m³);
- ρ_B is the density of the binder, in 1 kilogram per cubic metre (kg/m³)
- VMA is the voids content in the mineral aggregate in 0,1% (v/v);

4.1.3 Indirect Tensile Test

The test method for determining the (splitting) indirect tensile strength of cylindrical specimens of bituminous mixtures is described by the European Standards in NEN-EN 12697-23:2003. Through a strain-controlled Indirect Tension Test, the maximum (calculated) tensile stress applied to a cylindrical specimen loaded diametrically until break, at the specified test temperature and speed of displacement of the compression testing machine, is defined as the Indirect Tensile Strength (ITS) (Eq. 4.5). The principle of the test is bringing the cylindrical specimen to the specified test temperature, placed in the compression testing machine between the loading strips, and loading it diametrically along the direction of the cylinder axis, with a constant speed of displacement, until it breaks. The indirect tensile strength is the maximum tensile stress calculated from the peak load applied at break, and the dimensions of the specimen. A schematic representation of the test's setup along with a photo from the calibrating phase of the equipment before the test, are seen in Figure 4.4 and Picture 4.3 respectively.



Picture 4.3 Testing head with loading strips and calibration silicon specimen

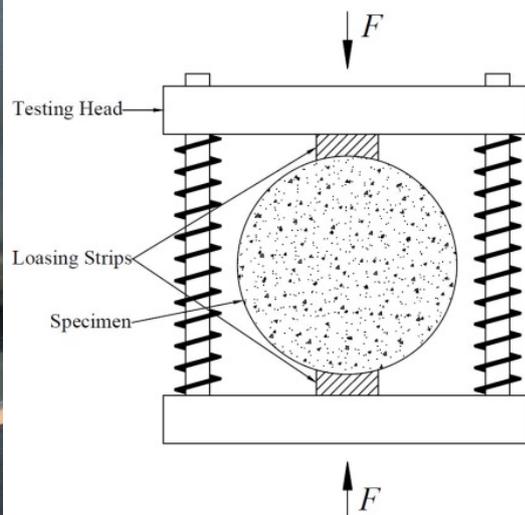


Figure 4.4. Scheme of testing head with loading strips

The test temperatures, as suggested by the European standard, shall be selected in the range between 5°C and 25°C, with a tolerance of $\pm 2^\circ\text{C}$. The Type Test Standard (NEN-EN 13108-20:2016) defines the test temperature at 15°C, hence the tests were chosen to comply with this. Prior to the test, and in order to achieve a uniform temperature among the specimens to be tested, all of them were placed in a 15°C thermostatically controlled chamber, for approximately 2 hours, regardless of their conditioning method. During this period, the ready-to-use testing equipment was also set at the test temperature, to allow the air contained in the chamber, but also the components of the equipment, both mechanical and non, to reach the test temperature uniformly.

Each specimen had to be tested within 1 minute after it was taken out from the temperature controlled chamber. To minimize the effect of experience and familiarization with the procedure, which follows the learning curve theory, specimens were tested with a distributed selection in their condition. This means, that conditioned and unconditioned, phase 1 and phase 3 specimens were tested alternatively, instead of following their numbering order. In this way, any possible inaccuracies or mistakes that were the product of low experience, were evenly distributed among all different specimen types, and not only in the first batch of specimens tested.

After the specimen was correctly aligned on the lower loading strip to ensure its diametric loading, the test was started. A diametric load was applied continuously and without a shock, at a constant speed of deformation of 50 ± 2 mm/min (or 0,85 mm/s), until the peak load P was reached. The loading continued until the break of the specimen. Since the equipment sensors measure the load applied in Volts (V), the correspondence between volts and newtons had to be made in the test's output to be able to interpret the results in engineering terms. This correspondence is $1V=5kN$ and the transformation was done after the end of all tests. After this step, the peak load P in N was recorded for each specimen and the Indirect Tensile Strength was calculated according to the norm as follows.

$$ITS = \frac{2P}{\pi DH} \quad [MPa] \quad \text{Eq. (4.5)}$$

where,

- ITS is the indirect tensile strength, expressed in Megapacals (MPa), rounded to three significant figures;
- P is the peak load, expressed in Newtons (N), rounded to three significant figures;
- D is the diameter of the specimen, expressed in millimetres (mm), to one decimal place;
- H is the height of the specimen, expressed in millimetres (mm), to one decimal place;

4.1.4 Standard conditioning method

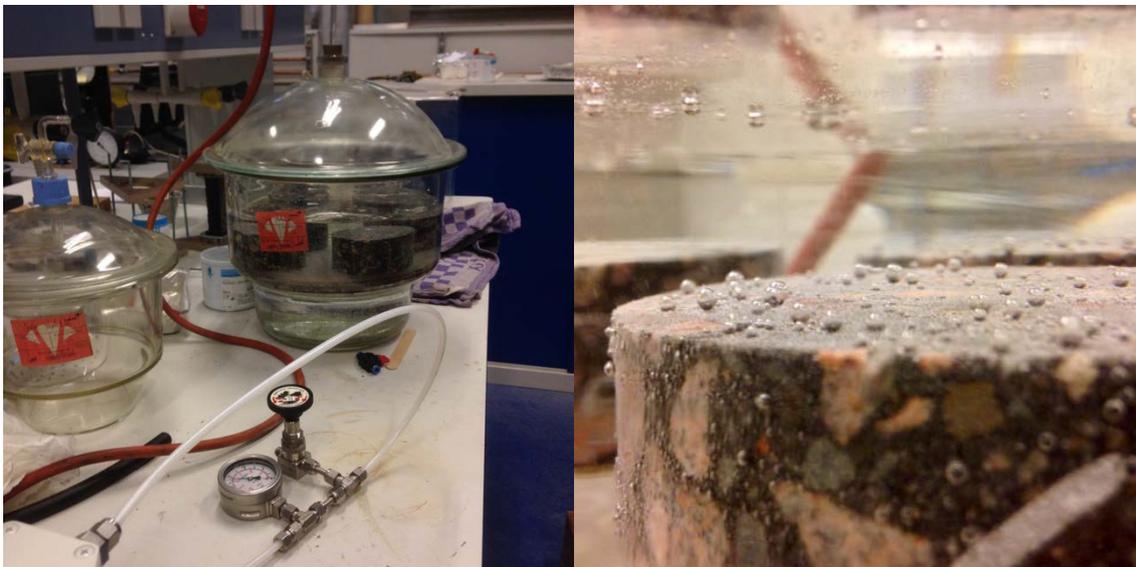
4.1.4.1 Procedure

The procedure followed in this method is standardised by NEN-EN 12697-12:2008. This norm describes three methods of conditioning, A, B and C, each one having differences in the way the specimen is treated with moisture. The Dutch standards, and as a consequence NL-Lab, adopt Method A. In this method the set of specimens is divided in two equally sized subsets (3 specimens per subset). One set is maintained dry at room temperature, while the other subset is saturated and stored in water at elevated conditioning temperature. After the conditioning, the indirect tensile strength of each of the two subsets is determined in accordance with NEN-EN 12697-23:2003, described in the

previous paragraph. The ratio of the indirect tensile strength of the water conditioned subset, compared to that of the dry subset is determined and expressed in percent.

The first step of the procedure is to force the water to enter the specimen's voids and reach a considerable level of saturation. This step is particularly important in our case, having a dense mix where water penetration in the voids is a difficult process. It is achieved by placing the wet subset of specimens on the perforated shelf of a vacuum container, filled with distilled water at $20\pm 5^{\circ}\text{C}$, to a level at least 20 mm above the upper surface of the specimens. The standard's guidelines note that the use of clear, drinkable tap water instead of distilled is allowed, provided that it has been demonstrated that the use of the local source gives the same results, as when using distilled water. It was confirmed that all the past tests in the faculty's labs were carried out using tap water, thus this option was also followed in our case.

After placing the water in the vacuum container, a vacuum of an absolute residual pressure of $6,7\pm 0,3$ kPa was gradually obtained within 10 minutes, to avoid any expansion damage of the specimens. The vacuum was then maintained for approximately 30 minutes, before the atmospheric pressure was slowly let into the vacuum container and the specimens were left submerged for another 30 minutes. A close-up of the air being forced out of the specimen's mass as well as the vacuum container connected to the manometer and pump, are seen on Picture 4.4.



Picture 4.4 Vacuum container connected to the pump (left) and air being sucked out of specimen's voids (right)

After the finish of the saturation process, the specimens were immediately placed in a water bath, to prevent any air entering the specimen's voids again. The water bath was at 40°C for a period of 70h, after which the specimens' were surface-dried and placed in a 15°C temperature chamber for 2 hours, before testing them according to paragraph 4.1.3.

4.1.5 Frost damage conditioning protocols

The protocols defined in this method were suggested in a research conducted by TNO, as part of InfraQuest, on 2013. The report's name is "Frost Damage Research, TNO 2013 R11050, IQ-2012-64".

This research was triggered by the indications that the already existing protocols in water and frost damage were not able to sufficiently distinguish between a "good" and a "bad" mixture. In addition, frost damage was noticed to appear more and more often, proving to be a problem of considerable concern for all parties involved. To determine the phenomena that leads to damage of porous asphalt during winter, the effect of water on the weakening of the mastic and the effect of frost on weakening due to expansion of water were analysed. To simulate the different situations, four protocols were proposed by InfraQuest and are utilised in this thesis. In general, what distinguishes this method from the standard method utilized so far is the addition of the thermal loading factor. It should be noted that in the original InfraQuest report five protocols are suggested in total. In this extra fifth Protocol V the specimens are subjected to moisture conditioning and thermal loading, having their macro pores saturated with water. This protocol was designed specifically for porous asphalt concrete and for this reason it was not included in this thesis. The first four protocols include a combination of water conditioning and thermal loading. The details are described below, followed by a visual overview.

Protocol I: No moisture conditioning – No thermal loading

This protocol provides the reference values, without any moisture or temperature damage

Protocol II: No moisture conditioning – Thermal loading

In this protocol, the effect of the differential thermal expansion of the various components of a sample is taken into account.

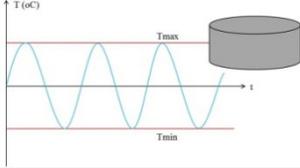
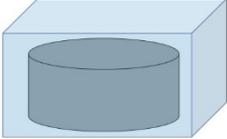
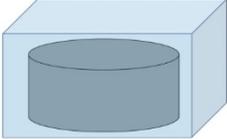
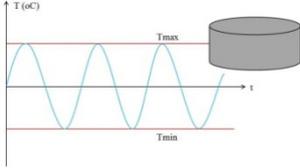
Protocol III: Moisture conditioning – No thermal loading

In this protocol, the effect of moisture that is absorbed in the mastic can be evaluated. It will show us if just the presence of moisture already leads to weakening of the bonds in the mastic or between mastic and aggregate.

Protocol IV: Moisture conditioning – Thermal loading

In this protocol, the effects of moisture in the mastic and temperature can be evaluated. It is a combination of II & III, showing if thermal loading enhances damage in a moisture conditioned mastic due to expansion of water in the mastic when it changes into ice. In this test, no water in the macro pores exists [25].

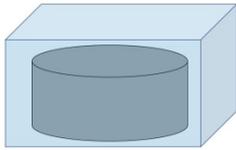
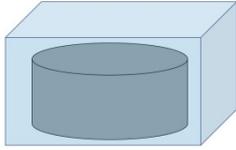
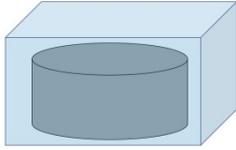
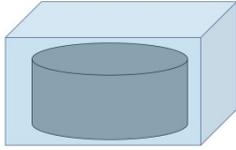
Table 4.9 Test protocol plan

Protocol	Moisture Conditioning	Thermal Loading	Evaluation Method
I	 No moisture conditioning	 No thermal loading	 ITT
II	 No moisture conditioning		 ITT
III	 Moisture conditioning at 25°C	 No thermal loading	 ITT
IV	 Moisture conditioning at 25°C		 ITT

To go into more detail in the InfraQuest report's steps on moisture conditioning, specimens in protocols I and II are not moisture conditioned. To ensure that all the specimens have the same age when tested, it was suggested that they remained in storage at 5°C during the moisture conditioning phase of the other specimens. However, due to the absence of such storing facilities available in the faculty, it was agreed that they would be stored under the same conditions as the specimens in the other methods, since ageing at such low temperatures and humidity levels is not affected much. The storing conditions were then $13\pm 1^\circ\text{C}$ and relative humidity of approximately $42\pm 1\%$.

The moisture conditionings of protocol III and IV consist of 2 steps. First the samples are saturated with water in a vacuum chamber, according to the procedure described in paragraph 4.1.4 and NEN-EN 12697-12. Afterwards the specimens are submerged in water at atmospheric pressure, at 25°C. The duration of the moisture conditioning was determined by InfraQuest by means of parametric CAPA-3D simulations, to estimate the time moisture needs to infiltrate into the mastic. Based on these, it is suggested that the samples remain submerged in water for 1000 hours (~42 days). An overview of the steps followed in moisture conditioning phase is seen on the table below.

Table 4.10 Moisture conditioning conditions

Protocol	I	 No moisture conditioning	Remain in storage at 15°C during the moisture conditioning of the other specimens			
	II	 No moisture conditioning	Remain in storage at 15°C during the moisture conditioning of the other specimens			
	III	 Vacuum saturation at 25°C	Step 1	Vacuum	6,7 (±0,3)	kPa
				t	30 (±5)	min
	IV	 Vacuum saturation at 25°C	Step 2	Water	25	°C
				t	1000	hours
	IV	 Vacuum saturation at 25°C	Step 1	Vacuum	6,7 (±0,3)	kPa
				t	30 (±5)	min
IV	 Vacuum saturation at 25°C	Step 2	Water	25	°C	
			t	1000	hours	

As far as the thermal loading is concerned, after moisturizing the specimens from protocols II and IV, they are directly subjected to thermal loading, in surface dry condition. The thermal protocol is based on the sandbox method of NEN 2872. Using again a CAPA 3D simulation the temperature propagation in an asphalt specimen was visualized and the thermal loading cycles were determined accordingly. One cycle has a total duration of 48 hours and consists of 8 steps of constant or changing temperature. In total the specimens are subjected to 24 cycles, meaning the total duration of the thermal conditioning phase lasts 1152 hours (49 days). The temperature cycles described by the temperature intervals and their corresponding durations are seen below. Also, an overview of the thermal loading distinguished by protocol is presented on Table 4.12.

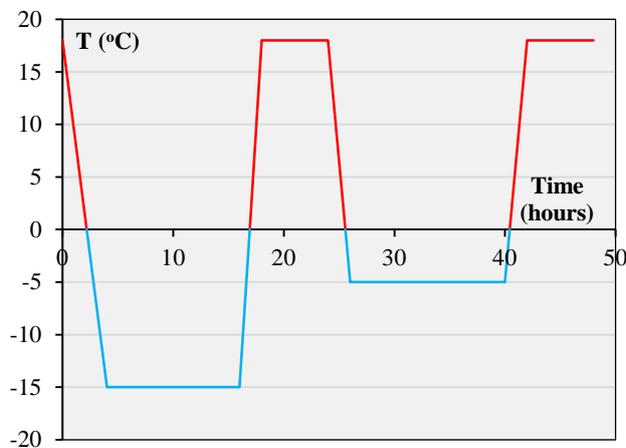
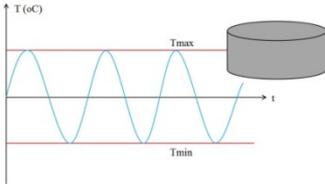
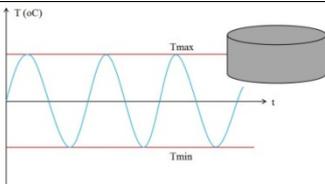


Figure 4.5 Temperature development during one thermal loading cycle of 48 hours

Table 4.11 Thermal loading cycles

Temperature (interval)	Duration (hours)
+18°C → -15°C	4
-15°C	12
-15°C → +18°C	2
+18°C	6
+18°C → -5°C	2
-5°C	14
-5°C → +18°C	2
+18°C	6

Table 4.12 Thermal loading conditions

Protocol	I	 No thermal loading	Remain in storage at 15°C during the moisture conditioning of the other specimens	
	II		# cycles	24
			Time per cycle	48 hours
			Total time	1152 hours (48 days)
	III	 No thermal loading	Remain in storage at 15°C during the moisture conditioning of the other specimens	
	IV		# cycles	24
			Time per cycle	48 hours
			Total time	1152 hours (48 days)

In common practice, the thermal loading in the actual pavement, penetrates the mass from above. To simulate this top-down thermal loading, all specimens from protocols II and IV were placed in an insulating mould during the cycles. This happened in such a way that the specimen was not restrained by the mould, to avoid any interference with the freeze thaw effect. The moulds were custom cut from a standard insulation material, with a height of 50 mm to fully cover the specimens, and an approximate thickness of 2,5 cm.



Figure 4.6 Insulation mould (left), and thermal loading chamber (right)

The moisture and thermal conditioning of all specimens finished on the 9th of May 2017.

4.1.5.1 Different test temperature

Even though the InfraQuest manual defines the testing temperature at 5°C, in this thesis the temperature of 15°C was chosen to maintain a uniform temperature along all the conditioning methods. In this way, the additional factor of temperature is eliminated and the comparison is easier. For this reason, after the end of the thermal loading, the specimens from all protocols were placed in a thermostatically controlled chamber at 15°C for 24 hours, before they were tested in Indirect Tension.

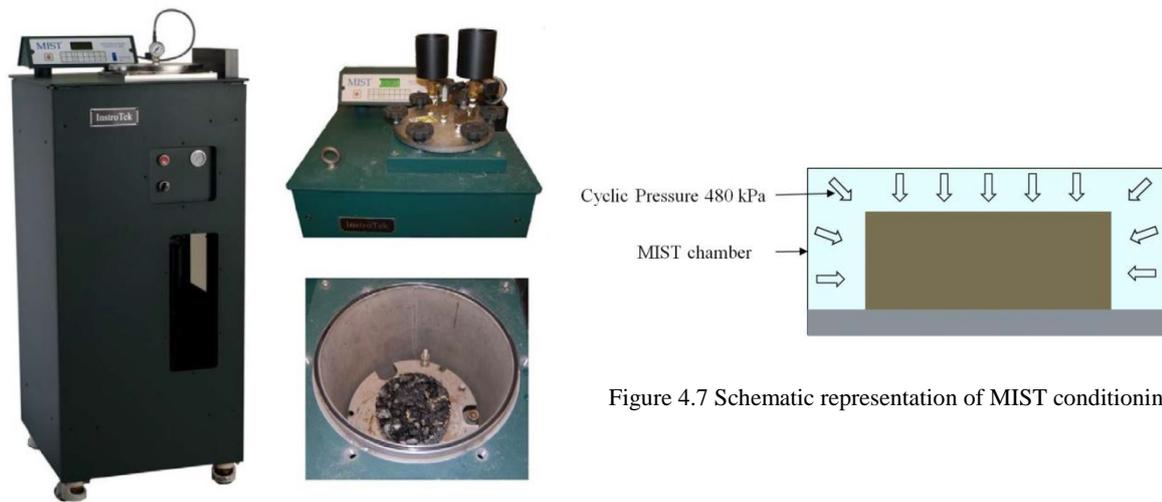
4.1.6 MIST conditioning protocol

MIST, that stands for Moisture Induced Stress Tester, is a new protocol which attempts to distinguish the contributions of long- and short-term moisture damage on asphalt mixtures. The protocol used in this thesis is based on a research carried out at TU Delft's Pavement Engineering section in 2012. The proposed protocol attempts to distinguish the contributions of long- and short-term moisture damage on asphalt mixtures, and can be used for the evaluation of their moisture damage.

The standard conditioning method of NEN-EN 12697-12 has the disadvantage of failing to capture the time frame over which moisture infiltration occurs. Furthermore it does not take into account the short-term moisture processes related to pumping action, even though it is not really present in the asphalt layers studied in this thesis (base/binder). In addition, it has been proven from field observations that stripping of open asphalt mixes is a rather localized phenomenon in trafficked areas of a pavement, which are oversaturated with water. These facts strengthened the claims that pumping action can be an important damage mechanism, and led to the development of the MIST protocol.

In order to address the differences in the time frame over which moisture diffusion and pumping action take place, this test protocol is based on the combination of two different conditioning methods: (a) bath conditioning, and (b) cyclic water pore pressure application. Water bath conditioning which facilitates the infiltration of water into the asphaltic mixture, is performed at elevated temperatures in order to accelerate the long-term degradation of the material properties. Cyclic pore pressure generation on the other hand is achieved by means of the Moisture Induced Stress Tester.

The MIST was designed as an accelerated conditioning device that aims in the evaluation of the resistance of an asphalt mixture to stripping, by simulating the high pressure which develops within an asphalt layer due to traffic loading. It is a self-contained unit developed by *InstroTek, Inc. (US)*. It includes a hydraulic pump and a piston that is designed to cyclically apply pressure inside a sample chamber. The test involves placing the specimen inside the chamber, filling the chamber with water, tightly closing the chamber lid, and starting the test choosing the preferred conditioning settings. The machine automatically heats the sample to the desired temperature and starts cycling between zero and the selected pressure [5].



Picture 4.5 MIST device [5]

Figure 4.7 Schematic representation of MIST conditioning

All the detailed steps defined by this protocol, and also followed during this thesis are:

1. Water bath conditioning at 60°C for 2 weeks.
2. Storing of specimens at approximately 13,5°C.
3. MIST conditioning one specimen at a time for 4000 cycles, at 60°C and a pressure of 70 psi (450 kPa).
4. After the MIST conditioning of each specimen, storing again at 13,5°C, until the test.
5. After the completion of all MIST tests, the specimens' dimensions and weights (Submerged, SSD and dry) were recorded. Also, a photograph of each specimen was kept for visual inspection of damage.
6. Before the ITT test, all specimens were placed at a 15°C chamber (test temperature) for 2 hours.
7. Specimens were tested in ITT, except for the two with the biggest visual damage that were kept for CT-scanning and tested 1 week later.

4.1.7 Overview of protocols

In order to have an overall image of the protocols and be able to distinguish the differences in their steps, an overview including all the steps is presented below in Table 4.13.

Table 4.13 Overview of the conditioning protocols

Standard Conditioning Protocol				
	30-min vacuum water bath at 25°C	72-hour water bath at 40°C	2 hours in thermostatically controlled chamber at 15°C	Indirect Tension Test at 15°C
MIST Conditioning Protocol				
	2-week water bath at 60°C	MIST conditioning for 4000 cycles at 60°C	2 hours in thermostatically controlled chamber at 15°C	Indirect Tension Test at 15°C
Freeze Thaw Protocol				
Protocol I	Stored at 15°C			Indirect Tension Test at 15°C
Protocol II	Stored at 15°C		49-day thermal loading (24 cycles)	Indirect Tension Test at 15°C
Protocol III	30-min vacuum water bath at 25°C	42-day water bath at 25°C	Stored at 15°C	Indirect Tension Test at 15°C
Protocol IV	30-min vacuum water bath at 25°C	42-day water bath at 25°C	49-day thermal loading (24 cycles)	Indirect Tension Test at 15°C

4.2 Permanent Deformation

4.2.1 Specimens

4.2.1.1 Organization

As in the case of water sensitivity, the specimens for the tests in permanent deformation were provided by Dura Vermeer on the 26th of January 2017. In total 54 specimens, equally divided into 3 phases, were delivered. The specimens were of 100 mm diameter and 60 mm target height. Their coding was again kept as it was defined by the contractor to maintain a consistency throughout the project and help the back-tracking of any information needed. Their production and storage, the mixture characteristics, and their compaction was identical to the specimens of moisture sensitivity (Paragraph 4.1.2), therefore they are not presented twice.

As described in the thesis methodology in Paragraph 2.4, two different stages of permanent deformation tests were considered. The first stage refers to the specimens of phase 1 (F1) and phase 3 (F3), and aims to answer the first two research questions; how well do lab and field specimens compare in terms of performance, and how accurate can the prediction of their performance be, in order to replace the costly tests. The second stage uses the specimens of phase 2 (F2) to test the different friction reduction methods. Their overview along with their volumetric properties is seen in the next paragraph.

4.2.1.2 Volumetrics

By visual examination, all the specimens appeared to be symmetrical with the curved side even and circular, hence none of them was excluded from the sets. The dimensions of the samples were determined by the contractor according to NEN-EN 12697-29:2002, taking for the diameter 2 perpendicular measurements at the top, middle, and bottom of the specimen, getting the average of the 6 measurements to the nearest 0,1 mm. The height was taken as the average of 4 evenly spaced measurements around the specimen's perimeter, 10 mm in from the edge, to the nearest 0,1 mm.

The bulk density (ρ_b) of each core was calculated by the contractor in accordance with NEN-EN 12697-6:2012, following both Procedure B (SSD) and D (by dimensions). For this purpose, besides the dimensions already measured, the dry, submerged, and saturated surface-dried weights were recorded for each specimen, at a certain water temperature. As advised by Annex A3 of the standard, procedure B is suitable for measuring the bulk density of dense-graded bituminous specimens having a low water absorption level. The applicability of this procedure is related to the voids level and the

diameter of the pores: for continuously graded materials, such as asphalt concrete with void contents up to 5%, which corresponds to our material. Procedure D on the other hand, is said to be more suitable for void contents greater than 15%. For this reason, $\rho_{b,SSD}$ was chosen as the most appropriate for the calculations, and will be noted as just ρ_b in all the following calculations.

In order to calculate the void content (V_m) of the cores, the maximum density (ρ_m) was measured according to NEN-EN 12697-5:2009, following Procedure A (Volumetric procedure), again by the contractor. In this case, the maximum density was calculated for each different mixture, which means one maximum density per phase. For each phase the maximum density was calculated as the average of 4 different measurements. The air void content of each specimen was then calculated following NEN-EN 12697-8:2003:

$$V_m = \frac{\rho_m - \rho_b}{\rho_m} \times 100 \quad [\%(v/v)] \quad \text{Eq. (4.6)}$$

In the following tables (Table 4.14, Table 4.15 and Table 4.16), an overview of the average dimensions, the bulk density, the maximum density and the void content of each specimen can be seen, distinguished by conditioning protocol.

Table 4.14 Phase 1 and Phase 3 Permanent Deformation specimens

Core Number	Phase	Protocol	Height mm	Diameter mm	ρ_b kg/m ³	ρ_m kg/m ³	V_m %
7	F1	I	60,1	99,0	2385	2506	4,84
10	F1	I	60,5	99,1	2390	2506	4,64
12	F1	I	60,6	99,0	2380	2506	5,01
21	F1	I	60,1	99,1	2370	2506	5,43
33	F1	II	60,9	101,2	2392	2506	4,54
37	F1	II	60,8	101,5	2395	2506	4,42
42	F1	II	59,8	101,3	2386	2506	4,77
44	F1	II	60,7	101,3	2385	2506	4,84
48	F1	III	60,7	101,3	2377	2506	5,14
50	F1	III	60,0	101,4	2394	2506	4,47
53	F1	III	60,3	101,3	2396	2506	4,39
55	F1	III	61,1	101,3	2406	2506	3,99
58	F1	IV	59,8	101,3	2396	2506	4,38
60	F1	IV	60,1	101,3	2396	2506	4,40
71	F1	IV	60,1	101,2	2390	2506	4,62
74	F1	IV	60,4	101,1	2400	2506	4,25
76	F1	Spare - III	59,6	101,4	2380	2506	5,05
80	F1	Spare - IV	60,3	101,4	2373	2506	5,29
1223	F3	I	61,2	101,5	2349	2492	5,72
1226	F3	I	61,1	101,5	2342	2492	6,01
1228	F3	I	60,9	101,6	2348	2492	5,80
1237	F3	I	60,3	101,4	2363	2492	5,18
1239	F3	II	59,5	101,5	2360	2492	5,31
1242	F3	II	61,0	101,6	2351	2492	5,65
1253	F3	II	59,6	101,5	2351	2492	5,67
1255	F3	II	59,3	101,4	2347	2492	5,83

1260	F3	III	60,6	101,3	2363	2492	5,19
1269	F3	III	60,9	101,4	2340	2492	6,11
1274	F3	III	60,7	101,4	2336	2492	6,26
1276	F3	III	60,0	101,4	2326	2492	6,66
1287	F3	IV	60,4	101,5	2333	2492	6,37
1290	F3	IV	60,6	101,6	2329	2492	6,54
1292	F3	IV	60,0	101,6	2320	2492	6,89
1301	F3	IV	60,2	101,5	2356	2492	5,47
1303	F3	Spare - IV	60,4	101,7	2346	2492	5,84
1306	F3	Spare - III	60,2	101,5	2360	2492	5,28

The protocol column refers to the combination of settings used during the triaxial test. Protocols I and II are at 40°C with maximum stresses of 450 kPa and 750 kPa respectively. Protocols III and IV are at 50°C with maximum stresses of 450 kPa and 750 kPa respectively. The four spare specimens were not initially assigned to a protocol, but were kept as a backup.

Table 4.15 Friction Reduction Methods Specimens overview

Core Number	Phase	Method	Height mm	Diameter mm	ρ_b kg/m ³	ρ_m kg/m ³	V_m %
1125	F2	No Reduction	59,8	101,3	2346	2490	5,79
1127	F2	No Reduction	60,3	101,4	2347	2490	5,76
1132	F2	No Reduction	61,0	101,3	2345	2490	5,82
1141	F2	No Reduction	60,8	101,4	2340	2490	6,03
1146	F2	Teflon	60,9	101,3	2344	2490	5,88
1148	F2	Teflon	60,0	101,2	2314	2490	7,05
1157	F2	Teflon	60,8	101,4	2312	2490	7,15
1162	F2	Teflon	59,9	101,3	2336	2490	6,19
1164	F2	Plastic + Soap	60,0	101,3	2315	2490	7,02
1173	F2	Plastic + Soap	59,7	101,3	2306	2490	7,40
1178	F2	Plastic + Soap	60,0	101,3	2325	2490	6,64
1180	F2	Plastic + Soap	59,2	101,3	2326	2490	6,60
1191	F2	Latex	60,0	101,1	2321	2490	6,77
1194	F2	Latex	60,0	101,3	2341	2490	5,99
1196	F2	Latex	59,0	101,3	2322	2490	6,74
1205	F2	Latex	60,2	101,4	2332	2490	6,35
1210	F2	Latex x2	60,2	101,4	2329	2490	6,45
1212	F2	Latex x3	60,1	101,4	2311	2490	7,19

Apart from these volumetric characteristics, the void content in the mineral aggregate (VMA) and the percentage of the voids in the mineral aggregate filled with binder (VFB) were also calculated in the same way as in 4.1.2.5, according to NEN-EN 12697- 8:2003.

4.2.2 Triaxial Cyclic Compression Test

4.2.2.1 Norm description

The test method followed in this thesis complies with the European Standards and is according to the guidelines specified in EN 12697-25:2016. This stress-controlled test method determines the resistance to permanent deformation of a cylindrical test specimen of a bituminous mixture by repeated load. During the test, the test's specimen change in height is measured at a specified number of loading cycles. From this, the cumulative axial strain ε_n of the specimen is determined as a function of the number of cycles. Out of the three methods described in the standard, the third method is adopted, "Method B", which refers to the determination of creep characteristics of bituminous mixtures by means of triaxial cyclic compression test (TCCT).

The minimum number of samples required for this test is three. The test is carried out by loading the specimen with a sinusoidal compressive stress in the vertical direction, and a radial confining pressure, for the simulation of the confinement of the specimen within the pavement structure. The confining pressure is held constant throughout the test to simplify the test control. In reality though, there is a sinusoidal oscillation in the confining pressure with a certain phase lag in the vertical loading due to the viscoelastic properties of asphalt concrete.

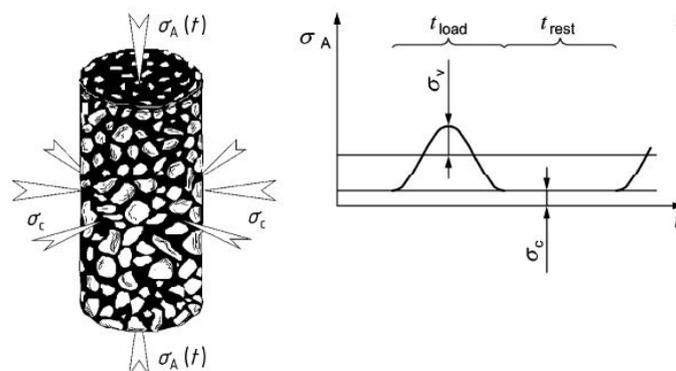


Figure 4.8 Stress conditions in triaxial test [26]

It should be noted that the actual stress conditions in the road cannot be simulated in the laboratory with simple test equipment. They depend on time (position of the wheel), the road structure, the depth in the structure, the stiffness of other layers, among other aspects. Therefore, the applied load conditions are only an approximation of the loads that occur in reality. One might suggest that application of a cyclic confining stress is to be preferred over a static confining stress, which was used here. However, given the considerations mentioned in the standard and the fact that cyclic confining stresses require advanced and expensive equipment, it is not applied for type testing.

The creep curve that is generated from the test is a display of the cumulative axial strain, expressed in %, of the test specimen as a function of the number of loading cycles. Generally the following stages can be distinguished:

Stage 1: the (initial) part of the creep curve, where the slope of the curve decreases with increasing number of loading cycles.

Stage 2: the (middle) part of the creep curve, where the slope of the curve is quasi constant and can be expressed by the creep rate f_c . The exact turning point of the creep curve lies within this stage.

Stage 3: the (last) part of the creep curve, where the slope increases with increasing number of loading cycles.

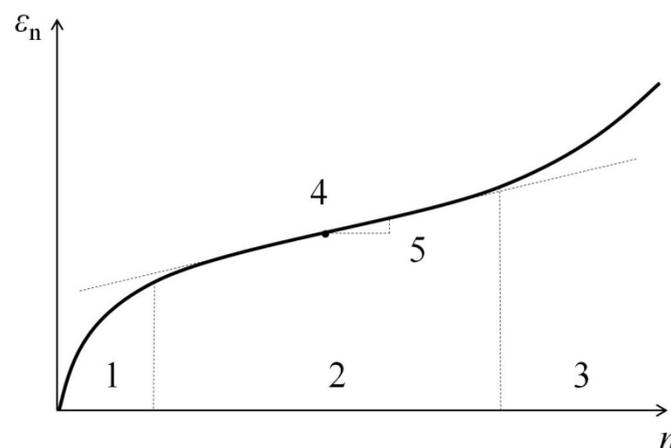


Figure 4.9 Creep curve stages

where,

- ϵ_n is the cumulative axial strain (%);
- n is the number of cycles;
- 1 is stage 1;
- 2 is stage 2;
- 3 is stage 3;
- 4 is the turning point;
- 5 is the creep rate f_c ($\mu\text{m}/\text{m}/\text{loading cycle}$)

The resistance to permanent deformation of the mixture shall be determined by interpreting the creep curve, according to one of the following methods by minimizing the squared error between curve fit and measured deformation. There are two methods of interpreting the creep curve, both of which were utilized in this research.

- Method 1: Determination of the creep rate f_c (linear fit)

If stage 2 is present the creep curve is represented on a linear scale, determining the slope B_1 from the least square linear fit of the (quasi) linear part of the creep curve (stage 2):

$$\varepsilon_n = A_1 + B_1 \cdot n \quad \text{Eq. (4.7)}$$

where,

ε_n = cumulative axial strain of the test specimen after n loading cycles, in percent (%) to the nearest 0,01%

A_1, B_1 = regression constants

The creep rate f_c in the (quasi) linear part of the creep curve in ($\mu\text{m}/\text{m}/\text{loading cycle}$) to the nearest 0,01 ($\mu\text{m}/\text{m}/\text{loading cycle}$) is:

$$f_c = B_1 \cdot 10^4 \quad [\mu\text{m} / \text{m} / \text{loading cycle}] \quad \text{Eq. (4.8)}$$

The parameter f_c is used to characterize the resistance to permanent deformation of the mixture tested. This method has the disadvantage that it is only a poor representation of the creep curve, since it only describes its secondary stage. Furthermore, the creep rate f_c depends highly on the selected interval used for curve fitting, because there is generally no part with real constant slope in the creep curve. However, it is the way in which the triaxial test results are currently reported in the Dutch Standard (RAW2015).

- Method 2: Determination of the parameters B and $\varepsilon_{1000,\text{calc}}$ (logarithmic fit)

The (quasi) linear part of the creep curve is determined from the following least square power fit:

$$\varepsilon_n = A \cdot n^B + C \quad \text{Eq. (4.9)}$$

where,

ε_n is the cumulative axial strain of the test specimen after n loading cycles, in percent (%) to the nearest 0,01%;

A is a regression constant;

B is the power least square power fit or the slope from the least square linear fit on the $\log(\varepsilon_n - C)$ versus $\log n$ -values;

C is a factor to correct deformation at the beginning of the loading;

The calculated permanent deformation after 1000 loading cycles, $\varepsilon_{1000,\text{calc}}$, in percent (%) to the nearest 0,01% is:

$$\varepsilon_{1000,\text{calc}} = A \cdot 1000^B + C \quad \text{Eq. (4.10)}$$

The parameters B and $\varepsilon_{1000,\text{calc}}$ are used to characterize the resistance to permanent deformation of the mixture.

For the purposes of this thesis, a Universal Testing Machine (UTM) was used to carry out the tests. The input parameters in the controlling software are in Table 4.16 and Table 4.17, distinguished by protocol. These parameters were accompanied by the specimen's individual number and its detailed dimensions, and were input before the start of each test.

Table 4.16 Triaxial Cyclic Compression Test – UTM Settings (Protocols I & III)

Pre-Load		Loading	
Stress (kPa)	9	Waveshape	Haversine Pulse
Load (kN)	0,071	Pulse Width (ms)	400
Hold Time (s)	120	Rest Period (ms)	6000
Confining		Contact	
Pressure (kPa)	50	Stress (kPa)	3
Hold Time (s)	10	Load (kN)	0,024
Termination		Deviator	
Axial Micro-Strain	90000	Stress (kPa)	400
Stop Test After Cycle	10000	Load (kN)	3,142

Table 4.17 Triaxial Cyclic Compression Test – UTM Settings (Protocols II & IV)

Pre-Load		Loading	
Stress (kPa)	15	Waveshape	Haversine Pulse
Load (kN)	0,118	Pulse Width (ms)	400
Hold Time (s)	120	Rest Period (ms)	6000
Confining		Contact	
Pressure (kPa)	150	Stress (kPa)	6
Hold Time (s)	10	Load (kN)	0,047
Termination		Deviator	
Axial Micro-Strain	90000	Stress (kPa)	600
Stop Test After Cycle	10000	Load (kN)	4,712

4.2.2.2 Friction reduction material application

Each specimen was mounted according to the guidelines of the norm. However, care should be taken in the friction reduction material. As described in the plan, in the first testing stage, only one method was followed; a plastic wrapping material that has a low yield strength, in combination with conventional soap (*Tricel*® *Goudzeep*). Each time a newly cut plastic sheet was used, adding approximately 2-3 grams of soap to sufficiently cover the entire surface of the base and top plate. The details of the material and its application on the equipment are seen in the following pictures.



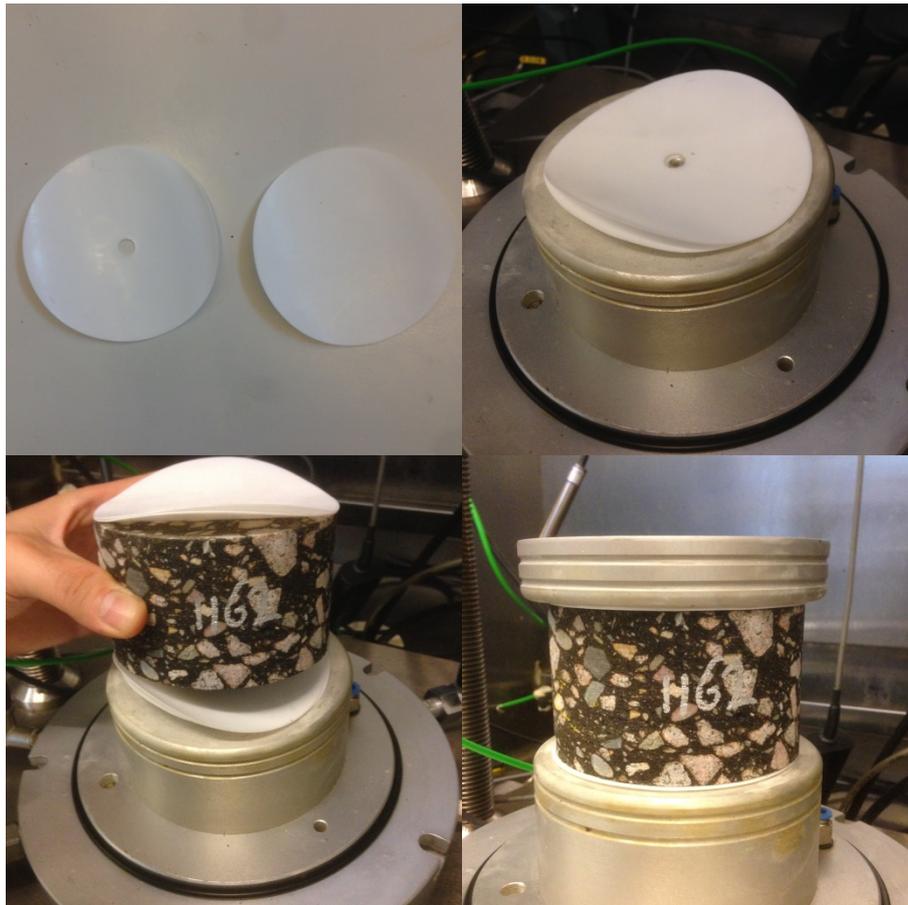
Picture 4.6 Wrapping plastic (top left), conventional soap (top middle), spreading of soap on the bottom plate (top right), placement of plastic on the soap (bottom left), specimen mounted with latex membrane (bottom middle), equipment ready for testing (bottom right)

The second testing stage in the permanent deformation studied the different friction reduction methods. In this stage, a group of specimens was tested with no material applied to reduce the friction, a group was tested with the same material as in the first stage, a group using Teflon sheets and another group with Latex membranes. Two round Teflon sheets (PTFE) of 0,5 mm thickness were placed on each interface with the loading plate.

Initially, vacuum grease was added in between the two sheets, to further reduce friction. However, after the end of the first test it was observed that the effect of grease was the opposite, causing the two layers to tightly stick together. A picture of this sticky surface after the application of vacuum grease is seen below (Picture 4.7). For this reason, the next tests were done using no material between the two sheets, since when slid by hand they seemed to behave better in this way.



Picture 4.7 Two layers of Teflon sheets stick to each other after grease application



Picture 4.8 Teflon Sheets (top left), sheets on base plate (top right), 4 sheets with specimen (bottom left), and loading plate on top fo specimen (bottom right)

Even in this case however, the two sheets were tightly attached to each other after the end of the test. This happened due to the high pressure developed at that surface during the loading. It required some effort to detach them by hands, indicating that this was also the case during the test and that friction was possibly not very effectively reduced.

The Latex membranes were of 0,3 mm thickness, having with the same material specifications with the membranes used for the vacuum isolation of the specimen (*ELE International Rubber Membrane*). In this case also, two of them were used on each interface with the loading plate, this time with an additional 1 gram of vacuum grease spread in between them (*Dow Corning ® High Vacuum Grease*).



Picture 4.9 Latex membrane and high vacuum grease (top left), spreading 1 gram of grease (top right), grease in between the two latex layers (bottom left), bottom plate (bottom right)

The Latex membranes that were re-used for a second and third time, were spread with additional grease before their use, to compensate for the grease that was forced out of the interface during the loading.

4.2.2.3 Temperature conditioning

The test temperatures varied between 40°C and 50°C, in the way described in the methodology followed. To allow the uniform specimen's temperature in its entire mass, each specimen was placed in a thermostatic chamber at the test's temperature, for approximately 3 hours before being mounted in the testing equipment.

In addition, after the equipment was ready, the test did not start before 2 hours passed, so that the fresh air that entered the chamber during the mounting, but also all the components used, reached the desired temperature. This procedure took place identically in all 54 tests done, to ensure that all the specimens were tested under the same conditions. During the test, the temperature was constantly monitored, both in the triaxial chamber where the specimen was under pressure (Skin temperature),

and in the outer chamber (Core temperature), and it was confirmed that the temperature did not deviate by more than $\pm 0,5^{\circ}\text{C}$ from the target temperature.

5 MOISTURE SENSITIVITY ANALYSIS

The results of the lab phase and analysis of the first functional characteristic studied, is the subject of this chapter. Initially, all the individual results and curves are presented, with no further comment or explanation. Each individual research question is then answered in a separate paragraph, referring to the corresponding results.

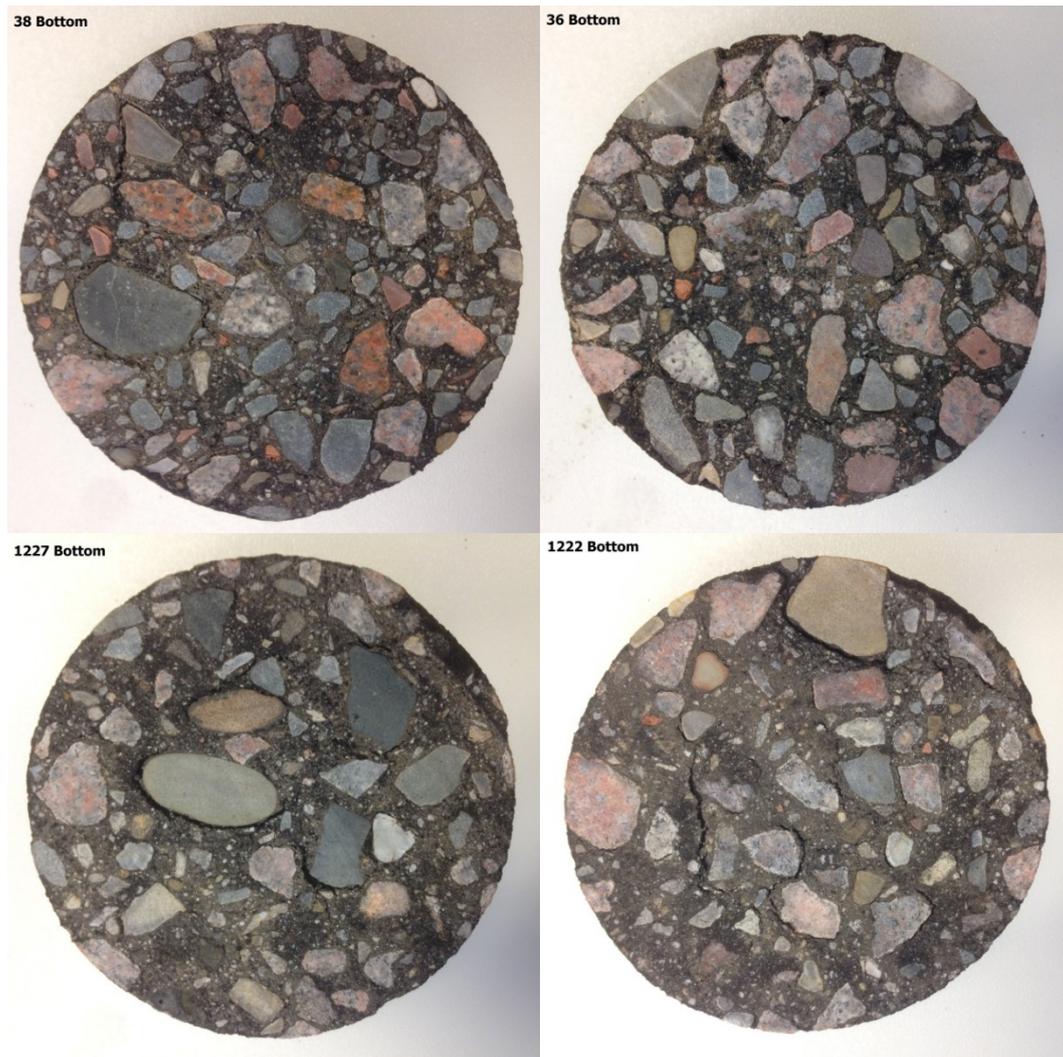
5.1 Test Results

5.1.1 Visual Inspection

During and after every conditioning and testing phase, all the specimens were carefully inspected to observe possible damages or changes in their appearance. A photographic record was kept for all the remarkable damages.

Starting from the MIST conditioned specimens, the first observation was made during their 2-week water bath at 60°C. After some days of conditioning, black stains of dissolved bitumen were noticed floating at the surface of the water. This was an early indication that a conditioning at such elevated temperatures leads to harsh changes and that bitumen undergoes accelerated deterioration. The expected natural consequence would be a highly reduced performance at the ITT.

After the water conditioning, the specimens underwent the MIST protocol with its combined water bath and pore water pressure. Taking out each specimen from the MIST chamber, an immense damage was noticed at their bottom side. Both cohesive and adhesive types of damage were clearly visible by the naked eye, entirely at the bottom side; none of them at the top. Some representative pictures of the specimens with the biggest damage noticed from Phase 1 and 3, are seen in Picture 5.1. It can clearly be seen that aggregates have been detached from the bitumen film (adhesion), and that cracks have appeared in the mortar's mass (cohesion). The biggest damage appears to be at the Phase 3 specimens (1227 and 1222), a fact which is confirmed later by the ITT results.



Picture 5.1 Lab specimens (top row) and field specimens (bottom row) directly after MIST conditioning

This extraordinary damage led to a further investigation, using CT-scans, in order to assess the internal damage of the specimens. These scans can be seen in paragraph 5.1.2.

In addition to the cracks that appeared, it was easily observed that there was a change in the shape of the specimens. Starting from perfectly symmetrical cylinders, after the MIST conditioning they ended up in a distorted shape. An expansion at the bottom side slowly turning to contraction as we move to the top side. This, in combination with the cracks, indicates the radical effect of pore water pressure increase inside the chamber. With the specimen being placed on a metal base that does not allow a pressure increase from the bottom, but only from the sides and the top, the cyclic 480 kPa pressure stresses the specimen in a compression like way. This process can be seen in the scheme below.

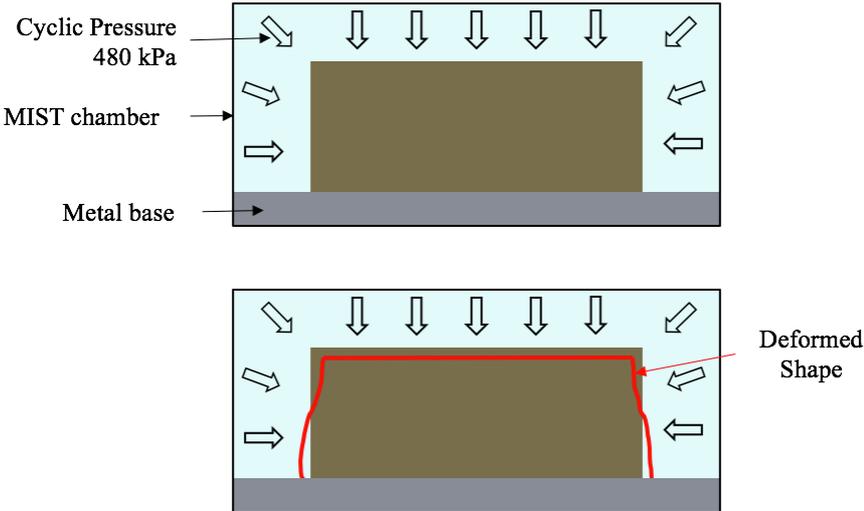
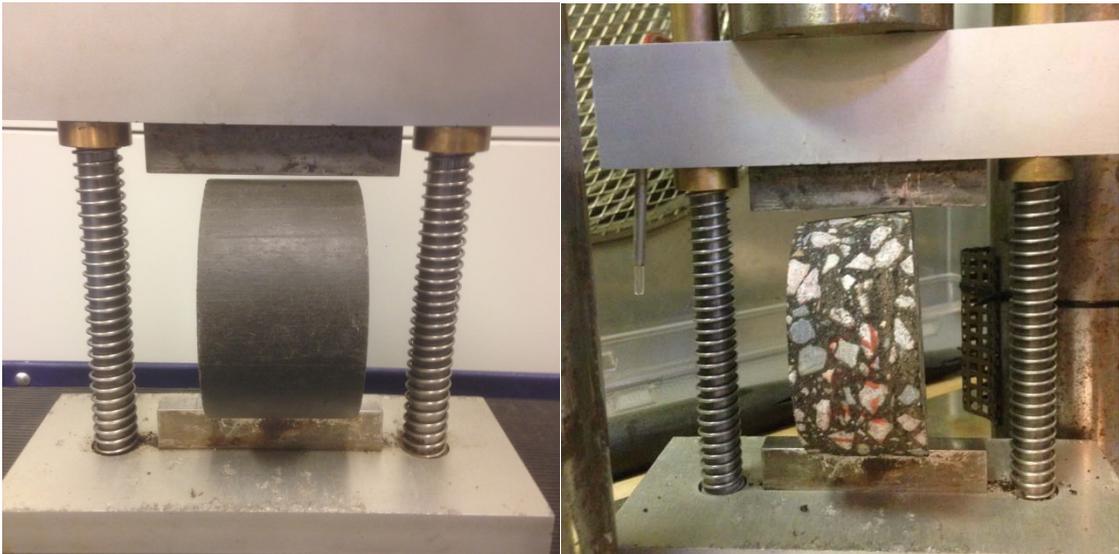


Figure 5.1 Specimen deformation during MIST conditioning

This extreme deformation led to a crucial problem during the testing process. The ITT’s setup requires the specimen to be mounted in a perfectly horizontal way, where the loading plates are in parallel. In this way the uniform diametric loading is ensured. With this deformation though, this assumption is violated and the loading is not uniformly distributed along the specimen’s height, but instead, it is concentrated at the tilted bottom side of the deformed specimen. It is obvious that this misalignment intervenes in the test’s output, since at the first moments of the loading, the stresses at the contact point are extreme, and consequently, the damage caused is not the desired. A representative picture of how the mounting should theoretically be before the loading, using a silicon calibration specimen, and how the mounting was with the deformed specimen, is seen in Picture 5.2 below.



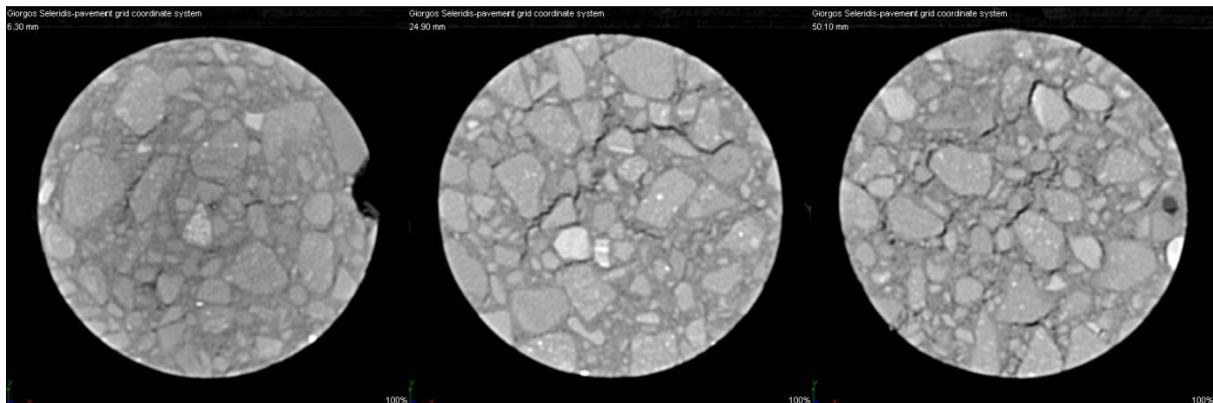
Picture 5.2 Core’s misalignment after by deformation caused from MIST conditioning

Due to the milder nature of the other two conditioning methods, no visual damage was observed. Hence, this and the next paragraph entirely refer to the observations made on the MIST conditioned specimens.

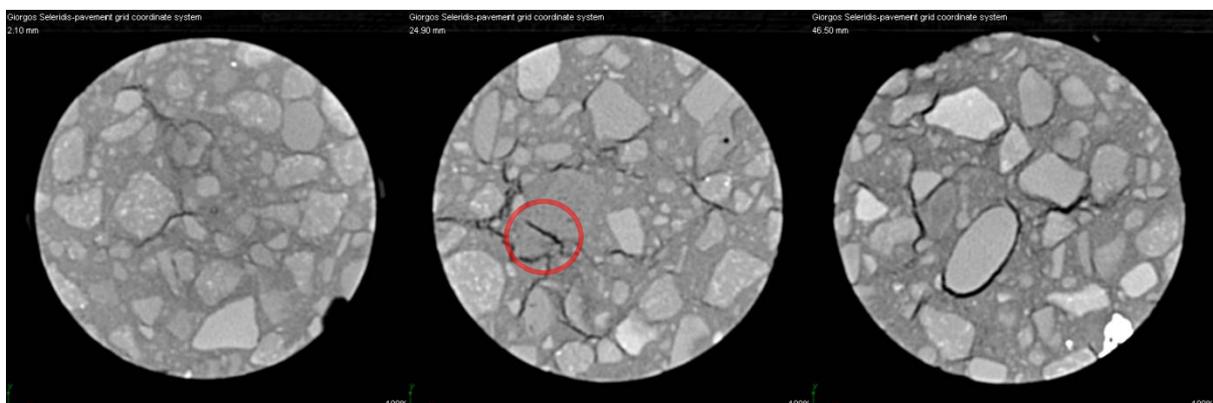
5.1.2 CT-Scans

Because of the excessive damage observed visually, two MIST conditioned specimens were CT scanned before being tested in ITT. One core was lab produced (#34) and one field produced (#1227). This was done to investigate the cracks on the surface, how they develop in the specimen's mass, and recognize types of distresses.

Looking the cores from the top, three representative 'slices' for each were chosen; one at the top of the core, one in the middle, and one at the base, where it was in contact with the MIST base plate (Picture 5.3 and Picture 5.4). Looking for the side, the 'slices' were mostly chosen to be in the center of the core, because this was the region where the biggest damages occurred and the most interesting information was found (Picture 5.5 and Picture 5.6).

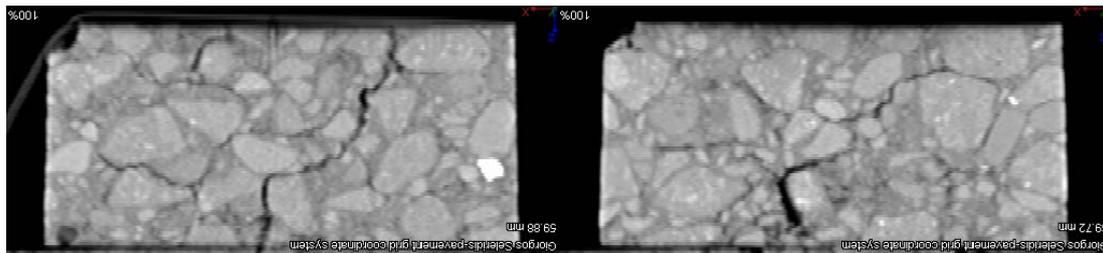


Picture 5.3 Horizontal CT-scans of Lab produced core #34 (top-left, middle-middle and bottom-right)

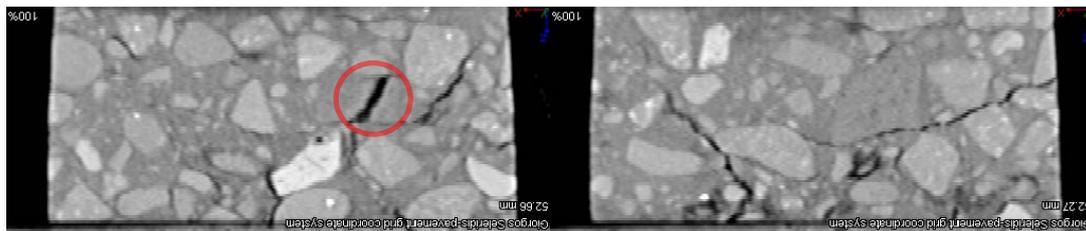


Picture 5.4 Horizontal CT-scans of Field produced core #1227 (top-left, middle-middle and bottom-right)

It is noticed that in both phases, the damages are smaller at the top and gradually increase in number and intensity as we get closer to the base. This is also visible by the slight diameter increase as we look the scans from left to right, and goes along with the deformed shape of the specimen seen in Picture 5.2. In both cases, the failures are both of cohesive and adhesive nature. In the case of the field specimen though, an aggregate breakage is also noticed (Picture 5.4, middle, red circle).



Picture 5.5 Vertical sections of Lab Core #34



Picture 5.6 Vertical sections of Field core #1227

The vertical cross sections are inverted in order to correctly depict the specimen with its base at the bottom. In both cases, the shape deformation is clearly visible, with an increasing diameter from top to bottom. Additionally, very big cracks are noticed, some of them starting from the top and being interconnected all the way to the base. This indicates radical damages and suggests a very decreased strength as it will later be seen. The aggregate failure previously noticed is also visible in this side view in Picture 5.6.

5.1.3 Volumetric changes

The radical visual damages necessitated the measurement of the specimens dimensions and weights, to quantify these damages in terms of volume change. After the MIST conditioning each specimen was stored again at $\sim 14^{\circ}\text{C}$, until their testing. Thus, due to the strict time limitation of the procedure, their complete drying could not be facilitated, leaving an amount of water in their mass. This would affect the measurement of their dry weight, since in reality they were not completely dry, and as a consequence, the accuracy of their new Bulk Density (G_{mb}) and Air Voids (V_m) as well. Therefore, the initial dry weight, measured before the conditioning, was used in the calculations. This imperfection in

the drying however, worked in favor of the accuracy in the submerged (W_{sub}) and surface dry (W_{SSD}) weights measurement. The changes in bulk density, air voids in the mass and dimensions is seen in Table 5.1 below. It should be noted that the diameter measurement refers to the bottom side of the specimens, hence the positive value of increase. As it is natural, there was an almost equivalent diameter decrease at the top side of the specimens.

Table 5.1 Volumetric changes in MIST conditioned specimens

Core Number	G_{mb} kg/m ³	G_{mb} after MIST kg/m ³	V_m %	V_m after MIST %	V_m change	Height change	Diameter change
34	2391,7	2353,4	4,56	6,09	33%	-1,0%	0,8%
35	2396,7	2364,0	4,36	6,89	58%	-1,4%	2,6%
36	2391,2	2387,2	4,58	5,79	26%	-0,8%	1,5%
38	2407,2	2384,8	3,94	6,05	53%	-2,1%	2,8%
1218	2323,1	2326,1	6,78	9,37	38%	-0,9%	1,9%
1222	2343,8	2320,2	5,95	9,55	60%	-0,6%	2,3%
1227	2340,9	2328,2	6,07	9,36	54%	-1,0%	2,5%

As it can be seen, there is a reduction in the bulk densities, and consequently an increase in their air voids. This is slightly bigger for the Phase 3 specimens, with an average air voids increase of 51%, in comparison to the Phase 1 specimens where the average increase is 43%. A reason behind this is the difference in their average air void contents, even before the conditioning. Field specimens had in absolute numbers approximately 2% air voids, meaning that they were susceptible to more damage in their mass due to the presence of water and the increase in their pore pressure. This small difference was even more obvious at the visual inspection, where the observed damage was more clearly visible in the field specimens. Similar trend is also noticed in the dimensions change.

5.1.4 Indirect Tension Test

The testing of the Standard- and MIST-conditioned specimens took place at the same period, with a difference of two days. The Frost Damage conditioned specimen were tested when the whole conditioning cycle finished, after approximately three months.

After the necessary data transformation from Volts to Kilo-Newtons, the peak load for each specimen was recorded. Based on that the Indirect Tensile Strength was calculated as described in Paragraph 4.1.3. It should be noted that the standard's guidelines give the strength in Giga Pascal. For reasons of comfort and ease of calculations though, the strength for all specimens is given in Mega Pascal.

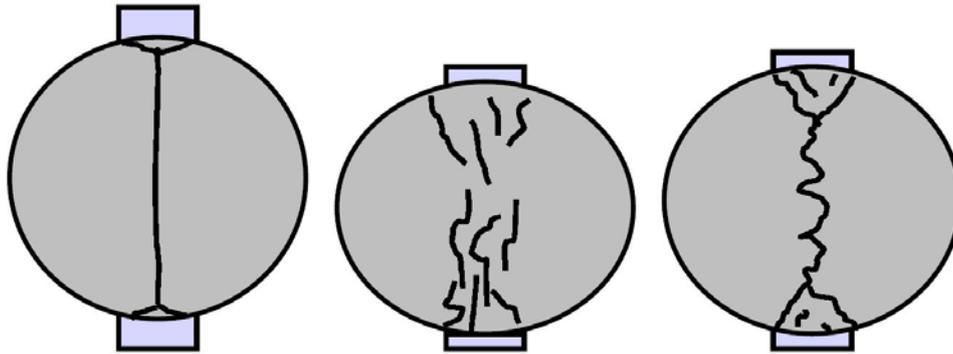


Figure 5.2 Types of failure in the ITT: a (left), b (middle), c (right) [27]

The failure pattern of all specimens followed the “Type c” failure from the standard (Figure 5.2), which is a combination of clear tensile break along the diametric line, and deformation with no visible tensile break line. They showed a limited tensile break line and larger deformed areas close to the loading strips [27].

According to the standard’s definition on the test’s precision, the obtained ITS values shall be accepted if the individual test specimens do not differ by more than $\pm 17\%$ of the mean value. The “Precision Check” column refers to this point, and all the values are accepted, with the maximum deviation being at 13,2%. This ensures that the conditioning and tests quality. In the following tables the results from all the Indirect Tension Tests collectively can be seen, divided in 3 tables based on the conditioning method (Table 5.2, Table 5.3 and

Table 5.4). The left part of the table with the 2-digit core numbers corresponds to Phase 1 lab specimens, whereas the right with the 4-digit core numbers corresponds to the Phase 3 field specimens.

Table 5.2 Standard conditioning method - ITT results - Lab (left) and Field (right)

Core	Phase	Protocol	ITS MPa	Precision Check	Core	Phase	Protocol	ITS MPa	Precision Check
40	F1	Dry	2,65	7,3%	1238	F3	Dry	2,4	0,5%
45	F1	Dry	2,25	-9,1%	1241	F3	Dry	2,4	3,6%
47	F1	Dry	2,52	1,8%	1246	F3	Dry	2,3	-4,1%
49	F1	Wet	1,94	-5,2%	1252	F3	Wet	1,7	3,6%
52	F1	Wet	2,12	3,7%	1257	F3	Wet	1,8	9,2%
56	F1	Wet	2,07	1,4%	1262	F3	Wet	1,4	-12,7%

Table 5.3 MIST conditioning method - ITT results - Lab (left) and Field (right)

Core	Phase	ITS MPa	Precision Check	Core	Phase	ITS MPa	Precision Check
34	F1	1,62	5,9%	1218	F3	0,9	4,1%
35	F1	1,41	-7,9%	1222	F3	0,7	-13,2%
36	F1	1,56	2,0%	1227	F3	0,9	10,5%
38	F1	1,53	0,0%	1231	F3	0,8	-1,4%

Table 5.4 Frost damage conditioning method - ITT results - Lab (left) and Field (right)

Core	Phase	Protocol	ITS MPa	Precision Check	Core	Phase	Protocol	ITS MPa	Precision Check
59	F1	I	2,67	8,5%	1266	F3	I	2,15	-0,2%
61	F1	I	2,37	-3,9%	1270	F3	I	2,25	4,1%
62	F1	I	2,35	-4,5%	1273	F3	I	2,07	-3,9%
63	F1	II	2,72	8,9%	1278	F3	II	2,04	0,2%
64	F1	II	2,34	-6,3%	1283	F3	II	1,96	-3,9%
66	F1	II	2,43	-2,6%	1286	F3	II	2,11	3,7%
67	F1	III	1,83	5,4%	1291	F3	III	1,28	-1,0%
68	F1	III	1,60	-8,0%	1295	F3	III	1,17	-9,3%
70	F1	III	1,78	2,6%	1299	F3	III	1,43	10,4%
72	F1	IV	1,21	1,0%	1302	F3	IV	1,10	-3,3%
75	F1	IV	1,15	-3,5%	1309	F3	IV	1,16	1,7%
79	F1	IV	1,22	2,5%	1311	F3	IV	1,16	1,7%

The mean values and standard deviations of each conditioning method and phase can be seen in Table 5.5. Also the figures plotted from these values follow the table.

Table 5.5 Descriptive statistics of ITS values by method and phase

Method	Dry		St. wet		MIST wet		Prot. I		Prot. II		Prot. III		Prot. IV	
Phase	F1	F3	F1	F3	F1	F3	F1	F3	F1	F3	F1	F3	F1	F3
Mean (MPa)	2,47	2,36	2,04	1,63	1,52	0,83	2,46	2,15	2,49	2,03	1,73	1,29	1,19	1,14
Std. Dev.	0,21	0,1	0,1	0,18	0,09	0,08	0,18	0,09	0,19	0,08	0,12	0,12	0,04	0,03

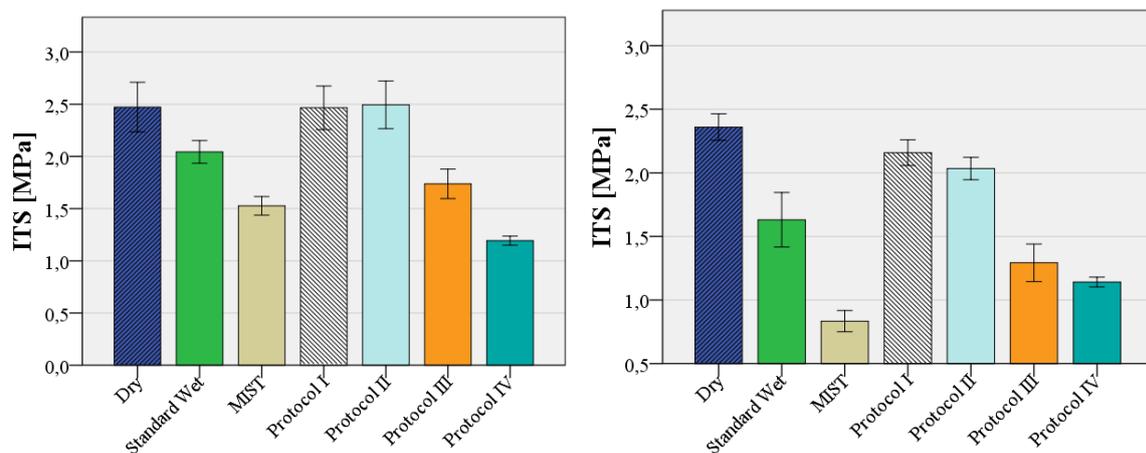


Figure 5.3 ITS distinguished by conditioning method and phase (lab/left and field/right)

The Indirect Tensile Strength Ratios calculated from the division of the wet group's average ITS by the average of the dry groups, are seen in Table 5.6, Figure 5.4 and Figure 5.5 show a graphical representation of these ITSR values, for lab and field specimens respectively.

Table 5.6 Indirect Tensile Strength Ratios (ITSR)

Method	Lab	Field
Standard	83%	69%
MIST	62%	35%
Protocol II	101%	94%
Protocol III	70%	60%
Protocol IV	48%	53%

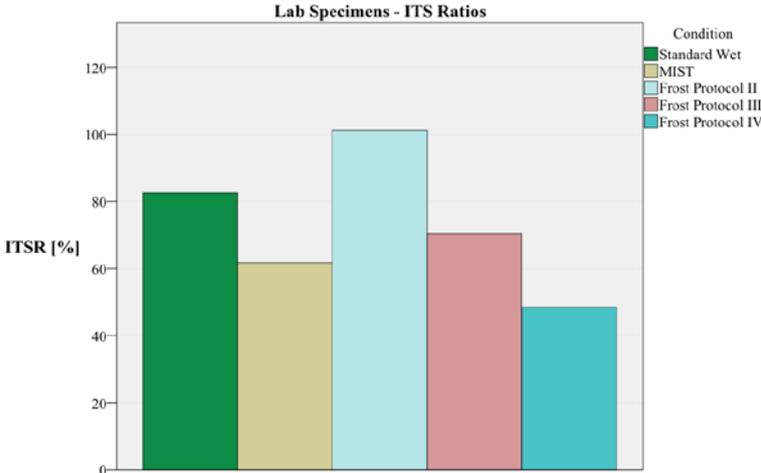


Figure 5.4 Lab specimens ITS values

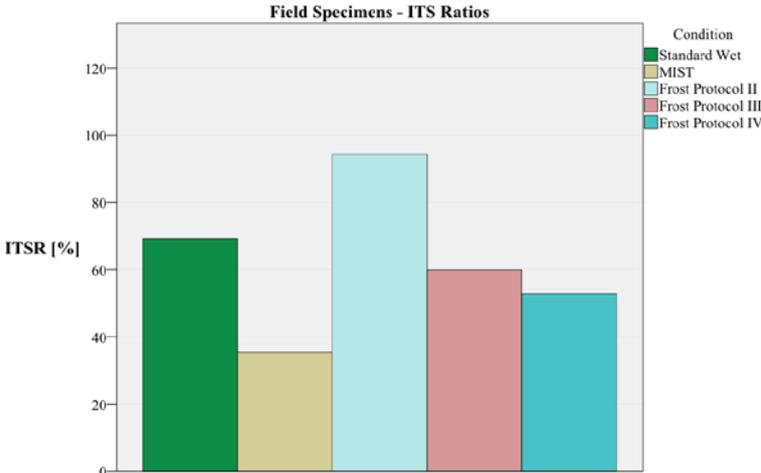


Figure 5.5 Field specimens ITS values

5.2 Conditioning methods comparison

5.2.1 Analysis of Variance (ANOVA)

The two-way ANOVA is used to determine whether there is an interaction effect between two independent variables on a continuous dependent variable. In many ways, the two-way ANOVA can be considered as an extension of the one-way ANOVA, which deals with just one independent variable rather than the two-way independent variables of the two-way ANOVA. By carrying out this analysis, we aim to identify whether the test results of different conditioning methods and phases are statistically connected in a significant way, or whether they are the result of a random relation. In this way, the precision and security in extracting conclusions on their physical relation is enhanced. The detailed methodology of ANOVA follows the step-by-step SPSS statistics guide from the online platform *Laerd Statistics* [28].

It is most often the case that the primary reason for running a two-way ANOVA is to establish whether there is an interaction effect between two independent variables (e.g. conditioning method and phase), sometimes called a two-way interaction effect. An interaction effect occurs when the effect of one independent variable on a dependent variable is different at different levels of the other independent variable. Stated another way, the effect of one independent variable (e.g. conditioning method) on a dependent variable (e.g. ITS), depends on the level of the other independent variable (e.g. phase).

Due to the size and the big number of graphs and tables generated, the representative ones that are essential for the analysis will be presented in this report. All the analysis files from SPSS can be found in the Appendices' electronic folder.

There are certain assumptions to be made and certain steps to be followed in order to reach a conclusion.

- Detect outliers or extreme cases

The first step is recognizing the presence of possible outliers. These are values of the dependent variable, that have an unusually small or large value compared to the other dependent variable values. They are recognized by their residuals, which will have large absolute values (i.e., large positive or large negative values).

There are two categories of outliers (as classified by SPSS Statistics) that can be found in a boxplot produced by SPSS Statistics: (1) outliers and (2) extreme points. Any data point that is more than 1.5 box-lengths from the edge of their box is classified by SPSS Statistics as an outlier. These data points are illustrated as circular dots. If any data points are more than 3 box-lengths away from the edge of

their box, they are classified as extreme points (i.e., extreme outliers) and are illustrated by an asterisk (*). Studying the graphs generated, no outliers were found to be present.

- Determine if the data are normally distributed

The next step is determining if the data is normally distributed. This can be checked using the Shapiro-Wilk test of normality. The results of the Shapiro-Wilk test are presented in the Tests of Normality shown in Table 5.10. If the assumption of normality has been violated, the "Sig." value will be less than 0,05 (i.e., the test is significant at the $p < 0,05$ level). If the assumption of normality has not been violated, the "Sig." value will be greater than 0,05 (i.e., $p > 0,05$).

Table 5.7 Tests of Normality for the ITS values of the Dry subset

Phase			Shapiro-Wilk		
			Statistic	df	Sig.
Lab	Dry	Residual for ITS	0,965	3	0,641
	Standard Wet	Residual for ITS	0,930	3	0,489
	MIST	Residual for ITS	0,951	4	0,724
	Protocol I	Residual for ITS	0,783	3	0,074
	Protocol II	Residual for ITS	0,918	3	0,444
	Protocol III	Residual for ITS	0,895	3	0,368
	Protocol IV	Residual for ITS	0,921	3	0,456
Field	Dry	Residual for ITS	0,987	3	0,783
	Standard Wet	Residual for ITS	0,927	3	0,477
	MIST	Residual for ITS	0,976	4	0,878
	Protocol I	Residual for ITS	0,998	3	0,918
	Protocol II	Residual for ITS	0,997	3	0,892
	Protocol III	Residual for ITS	0,992	3	0,832
	Protocol IV	Residual for ITS	0,750	3	0,000

As it can be seen, there is significant evidence that the data in their majority are normally distributed as assessed by Shapiro-Wilk's test. The only case the distribution is not normal, is at the Field specimens of Frost Protocol IV. Even though checking the normality of only 3 points is difficult in any case, this table provides an indication of it. In order to visualize these distributions and correlate them to the Shapiro-Wilk test, three representative histograms of the distributions found above will be plotted. These will include a normally distributed case (Field Protocol II), a marginally normally distributed (Lab Protocol I), and a non-normally distributed (Field Protocol IV). The histograms can be seen in Figure 5.6 below.

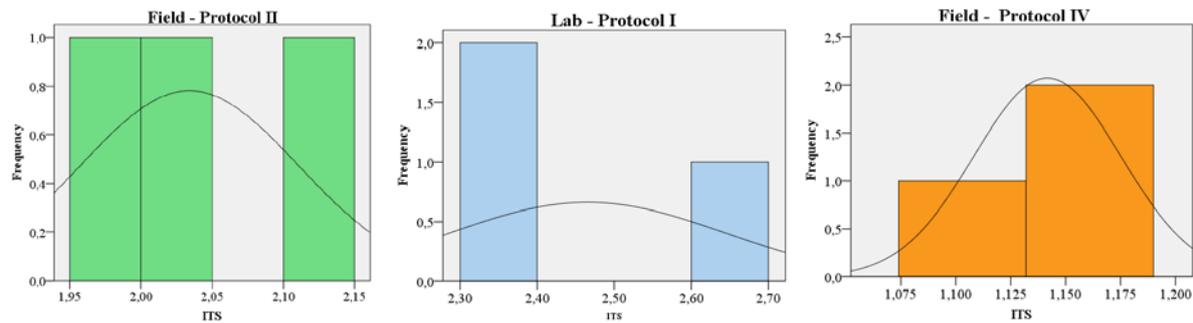


Figure 5.6 Histograms of a normal (left), almost normal (middle), and non-normal (right) distributions

The assumption of normality is necessary for statistical significance testing using a two-way ANOVA. However, the two-way ANOVA is considered "robust" to violations of normality. This means that some violation of this assumption can be tolerated and the test will still provide valid results. Therefore, there are often cases requiring only approximately normally distributed data. For this reason, because of this indication from Table 5.7, we consider the dataset approximately normally distributed in its entirety.

- Determine the homogeneity of variances

The next step is determining the homogeneity of variances. The two-way ANOVA assumes that the (population) variances of the dependent variable (or residuals) are equal in all combinations of groups of the independent variables. SPSS Statistics tests this assumption by determining if there are equal variances (called homogeneity of variances) in all combinations of groups of the two independent variables. The assumption of homogeneity of variances is tested using Levene's test of equality of variances, which is seen in Table 5.8.

Table 5.8 Levene's Test of Equality of Error Variances

F	df1	df2	Sig.
1,827	13	30	0,085

We can see that the statistical significance level is 0,085 (i.e., $p = 0,085$), which indicates that Levene's test is not statistically significant (because $p > 0,05$). A non-statistically significant result, such as this, indicates a significant evidence that we have equal (population) variances and we have not violated the assumption of homogeneity of variances.

- Determine whether an interaction effect exists

The primary goal of running a two-way ANOVA is to determine whether there is an interaction between the two independent variables: Conditioning methods and phases. As such, we are primarily interested in determining whether the effect of Phase of production is different for the 7 different protocols followed. We can determine whether we have a statistically significant interaction effect from interpreting Table 5.9. The interaction effect is represented as the product of the two independent variables in a two-way ANOVA.

Table 5.9 Tests of Between-Subjects Effects table for dry subset

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	13,090 ^a	13	1,007	64,122	0,000	0,965
Intercept	143,179	1	143,179	9117,936	0,000	0,997
Protocol	11,097	6	1,849	117,775	0,000	0,959
Phase	1,368	1	1,368	87,113	0,000	0,744
Protocol * Phase	0,489	6	0,081	5,188	0,001	0,509
Error	0,471	30	0,016			
Total	153,647	44				
Corrected Total	13,561	43				

The "Sig." column presents the significance value (i.e., p-value) of the interaction effect. We can see that the p-value for this interaction effect is .001 (i.e., $p = 0,001$). This is less than 0,05 (i.e., it satisfies $p < 0,05$), which means that there is a statistically significant interaction effect. This indicates that it was correct thinking that the effect of Phase on ITS values depends on Conditioning Method or, equivalently, the effect of Conditioning Method on ITS values depends on Phase of production. In other words, the values obtained in each different protocol and phase, are significantly different from each other, hence stronger conclusions can be drawn.

5.2.2 Comparison

The target of this paragraph is to identify the way each conditioning method affects the moisture sensitivity of asphalt concrete, and order them in terms of suitability for this purpose. This will happen through the study and comparison of the tests results, employing photos, ITS values and CT-scans.

There are some important assumptions that need to be addressed, prior to the comparison. First of all, all the specimens tested were produced from the same mixture, having uniform composition characteristics. In this way, possible observations that are the result of e.g. different density or bitumen content/quality, are excluded. Secondly, after every conditioning process, all the Indirect Tension Tests took place at the same temperature of 15°C. The sensitive factor of test temperature was in this way eliminated, making the comparison of the protocols entirely a matter of their conditions. Finally,

having only this specific mixture, the conclusions to be drawn are limited to the extent in which each conditioning method damages the specimens, and not their ability to distinguish between a good and a bad mixture. This would only be facilitated by the comparison of two mixtures, already identified as ‘good’ and ‘bad’.

Starting from the image of the specimens after the conditioning, it is a safe conclusion that the MIST conditioning results in the biggest damage of the three methods. While in the Standard and Frost Damage method no change could be noticed by naked eye, in the MIST conditioned specimens damage could easily be observed. Bitumen stains floating during the first water bath, and big adhesive and cohesive cracks after the MIST conditioning, indicated that the specimens underwent a highly damaging process. The expectation after this was a reduced strength in indirect tension; a strength much lower comparing to the other methods.

Looking at the ITT’s output, Table 5.10 arranges the conditioning methods in descending order, based on the change they caused in the ITSR values. Additionally, each group of specimens is compared to the *RAW2015* requirements for moisture sensitivity of an OL/TL (base/binder layer) mixture. For a mixture used in a base/binder layer to be considered adequate in terms of moisture sensitivity performance, these requirements set the minimum ITSR limit at 70%. In this way, the groups are distinguished by a ‘Pass’ or ‘Fail’ characterization, based on this check.

Table 5.10 Conditioning methods in descending order, based on the damage caused

Lab			Field		
Method	Change	Check	Method	Change	Check
Protocol II	+1%	Pass	Protocol II	-6%	Pass
Standard	-17%	Pass	Standard	-31%	Fail
Protocol III	-30%	Fail	Protocol III	-40%	Fail
MIST	-38%	Fail	Protocol IV	-47%	Fail
Protocol IV	-52%	Fail	MIST	-65%	Fail

The least damaging method is undoubtedly Protocol II from the Frost Damage method (Thermal Loading only). The specimens showed a very small reduction in the case of field cores, and even a small increase in the case of lab. Both phases easily pass the standard’s check, characterizing them as excellent performing in moisture sensitivity. The increase in strength at such low levels, is too small to be actually considered an improvement in the specimens performance. It is attributed to the accuracy of the measurements and the variation in the specimens. It does show however that by subjecting the mixture to pure thermal loading cycles (from -15°C to +18°C), with no moisture present, its chemical and mechanical composition is not affected. The damage caused by dry frost conditions is very limited and can even be considered negligible.

Looking at the Standard conditioning which comes second on the list, it is clear that even a short-duration water bath (3 days at 40°C), manages to change the mixture's performance in a noticeable rate. In the case of lab specimens, the strength reduction was intermediate, with a decrease of 17%, passing the standard's requirement, whereas in the field specimen's the decrease was a bit more than 30%, leading to the mixture's characterization as poor.

Protocol III (42-day water bath at 25°C) can be characterized as a method which leads to considerable damage on the specimens. The lab specimens marginally fail to satisfy the requirement, having a strength decrease of 30%, while the field specimens clearly fail, with a 40% decrease. Comparing to the standard method, one conclusion is that the combination of temperature and duration of the conditioning is not very effective in terms of damage caused. With a conditioning time 14 times bigger than the standard (not taking into account the 30-minute water bath), one would expect a radical change in performance. This change, however, was only 1,8 times higher in the lab, and 1,3 times higher in the field values, mainly because of the low conditioning temperature. For this reason, even though the long-term effects of water were magnified in a realistic way (25°C), this could be achieved in a more accelerated and time-effective way, by following a combination of medium temperatures and duration.

The MIST protocol lies in the second and first place as the most damaging protocol, in the lab and field specimens respectively. Taking into account the difference 15°C did, between Protocol III and Standard, it can be understood that the 2-week 60°C water bath led to immensely accelerated changes in the mixture's physico-chemical composition. On top of this already damaged chemical composition, the relatively high cyclic pressure led also mechanical damage and change in the specimens' structure, exposing the short-term effects of the presence of water in the mixture's mass. An exposition which happens only by this method, out of the ones studied in this thesis. Both phases clearly fail to satisfy the standard's requirements, with the field specimens being considered of the poorest quality.

The method that competes with MIST in being the most damaging, is Protocol IV, failing by far to meet the 70% requirement. Protocol IV is a combination of both water and thermal conditioning, and expectedly leads to radical changes in the specimen's performance. This impact however, can mostly be attributed to the presence of water, since Protocol III which refers to pure water conditioning, already leads to big damages. The Freeze Thaw effect that takes place results in the intense expansion and contraction of the water in the voids. Consequently, an internal damage is the outcome, being reflected in the reduced ITS values. This effect only furtherly adds up to the damage by -17% and -7% in lab and field respectively.

5.3 Comparison of lab to field determined properties

5.3.1 Tests output of this thesis

5.3.1.1 ITT

For the purpose of recognizing the differences in performance between specimens produced in lab conditions and specimens extracted from an actual pavement layer, the tests were divided in two equal batches; one with lab cores (F1) and one with field cores (F3). The same identical storing conditions, test procedures and specimen treatment was followed in both batches to allow the most accurate comparison possible. In this paragraph, the data taken into account are limited to the tests carried out for this thesis. In the next paragraph, the comparison is extended to the whole NL-Lab dataset, including all 5 works, to give a broader image.

Regarding the moisture sensitivity performance characteristic, as described in the previous paragraphs, three different conditioning protocols were followed, serving on one hand in distinguishing the way each protocol damages asphalt concrete, and on the other hand, in assessing the differences in performance between the two phases. The average values referring to the latter are in Table 5.11. They are categorized by conditioning method. A comparison of the universal average values cannot be made due to the different range of values each method leads to.

In the case of moisture sensitivity, higher ITS and ITSR values suggest a higher resistance to damage, hence a more desirable performance. For this reason, the comparison is made characterizing a phase as performing ‘better’ or ‘worse’, according to the relative comparison of the values mentioned. In the case of permanent deformation though, high values in f_c and ϵ_{1000} suggest a lower resistance to damage, hence the correspondence of ‘good’ or ‘worse’ is the opposite from this case.

Table 5.11 Average Indirect Tension Strength values, lab and field

Method	Condition	Mean Lab ITS MPa	Lab St. Dev	Mean Field ITS MPa	Field St. Dev	Difference
Standard	Dry	2,47	0,21	2,36	0,10	-4,4%
	Wet	2,04	0,10	1,63	0,18	-20,0%
MIST	MIST	1,53	0,09	0,83	0,08	-45,7%
Frost Damage	Protocol I	2,47	0,18	2,16	0,09	-12,5%
	Protocol II	2,49	0,19	2,03	0,08	-18,4%
	Protocol III	1,74	0,12	1,29	0,12	-25,8%
	Protocol IV	1,19	0,04	1,14	0,03	-4,2%

The general trend shows the field specimens' strength to be lower than the lab, in all protocols tested. The smallest deviation, around 4% lower for the field, is observed in the Standard Dry and Protocol IV (water and thermal loading) conditions. The maximum, with the field specimens being almost 46% lower than the lab, is observed in the MIST protocol. A visual representation of these values in Figure 5.7 will give a clearer image making the comparison easier. Regardless of the conditioning method, it is clear that Field produced specimens show a lower strength in Indirect Tension, for the specimens tested in this project.

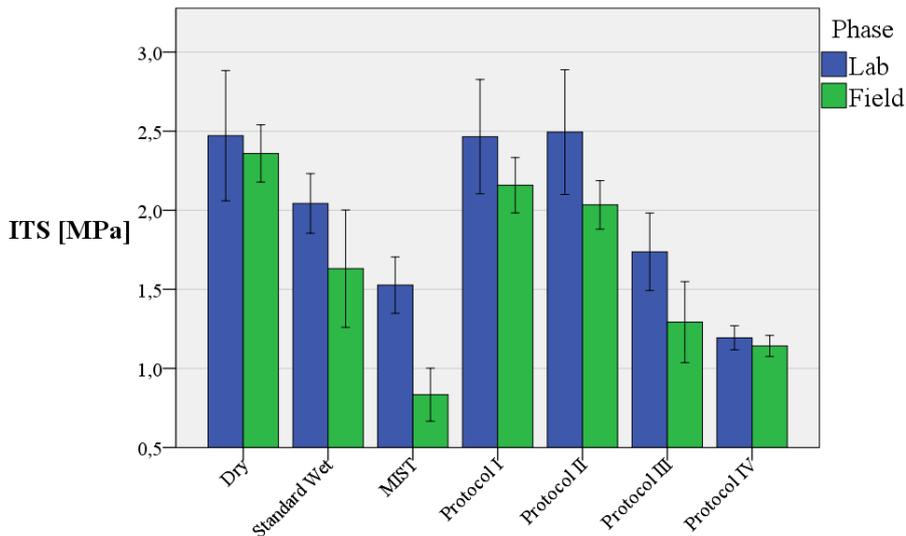


Figure 5.7 Average ITS values distinguished by phase and conditioning protocol

The next step, is comparing the Indirect Tensile Strength Ratios, resulting from the divided average values from the conditioned specimens, with the dry unconditioned ones. The image is expectedly similar to the ITS bar chart. In all the cases, besides Protocol IV, F3 cores perform worse in moisture sensitivity. The differences vary from very small, like Protocol II, to immense, like MIST Protocol.

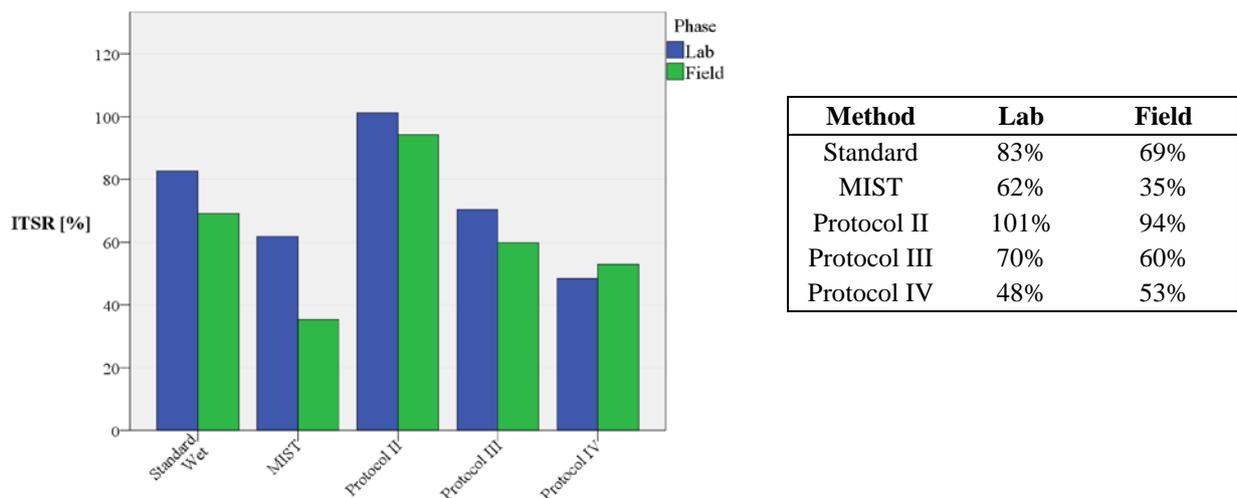


Figure 5.8 ITSR distinguished by phase and conditioning protocol

Employing the *RAW2015* standards' 70% minimum ITSR requirement, useful observations were made. In Table 5.12, the reduction in the ITSR values with the Pass/Fail check can be seen. While in the majority of the cases lab and field specimen perform similarly on the basis of this check, both passing or both failing, Standard method leads to the opposite observation. Lab specimens that were conditioned in the way described in the current standard, which is what is being followed at the moment in common practice, remained over the 70% threshold, at a satisfactory level. In contrast, field specimens failed to fulfill this requirement by 1%.

This is an important remark, since it has a direct application in the design phase as it currently is. Taking for example the case in which this mixture was designed, constructed and finally tested for its performance characterization. If it was constructed in lab conditions, the requirement would most certainly be fulfilled, meaning that the mixture is appropriate to be used at a given project. If, however, the mixture and cores were taken from the field for testing, the result would most probably not surpass the 70% ITSR value. This means, that the same mixture, differently produced, would lead to a totally different direction in the design process.

Table 5.12 RAW2015 requirements fulfilment check

Method	Lab		Field	
	Change	Check	Change	Check
Standard	-17%	Pass	-31%	Fail
MIST	-38%	Fail	-65%	Fail
Protocol II	+1%	Pass	-6%	Pass
Protocol III	-30%	Fail	-40%	Fail
Protocol IV	-52%	Fail	-47%	Fail

5.3.1.2 CT-Scans

In order to assess the effect mostly of compaction in the lab and in the field, some specimens were CT-scanned to analyze the distribution of their voids. Three sets of two specimens were chosen from each phase; one set with the maximum void contents of each phase, one set with the minimum, and one set that had similar void content. The specimens are seen in the table below.

Table 5.13 Specimen's selected for the CT-scans

	Min		Max		Similar	
	Lab	Field	Lab	Field	Lab	Field
Specimen	59	1302	61	1295	75	1238
Void content	4,1%	5,0%	5,8%	7,1%	5,2%	5,1%

After the specimens were scanned with x-rays and their cross sections were obtained, the scans were analyzed with the software *ScanIP Simpleware*. Based on the picture's color density, the software identifies the areas of voids and areas of bitumen or aggregates and calculates the voids area of each section in mm^2 . The thresholds between which a color is considered air and not asphalt is not strictly defined by the program, but depends on the user and the brightness of each image. For this reason the void calculation is quite subjective, and the absolute numbers of the area may not be perfectly representative of reality. For the sake of this research we consider that the distribution and variance along the specimen's height is accurate enough to make a comparison between the two phases.

Having obtained the void surface area of each 'slide', they were plotted against the depth z of each specimen. In some cases the height does not perfectly reach 50 mm, because some slides from the top and bottom surfaces were excluded due to their rough texture that affected the voids recognition.

Starting with the set of specimens that contained the minimum void content from each phase, the distributions seem to be comparable (Figure 5.9). Not taking into account the peaks at the edges which are most probably caused by small aggregate particles being detached, or by rough surface, the voids' surface along the height are equally distributed, and in a similar way.

The set of specimens with the maximum air void content however, shows a slightly different image (Figure 5.10). The lab specimen (61) shows an increasing void surface from top to bottom, while the field specimen's is decreasing with a hump after the middle.

In the case where the two specimens were chosen to have almost the same void content, the distributions are close to identical Figure 5.11. First of all, the fact that the two curves are in the same order of magnitude, with the field specimen showing a slightly lower void content (which coincides with the volumetric calculation's 0,1% difference), suggests that threshold choice in the software was accurate enough. Regarding the

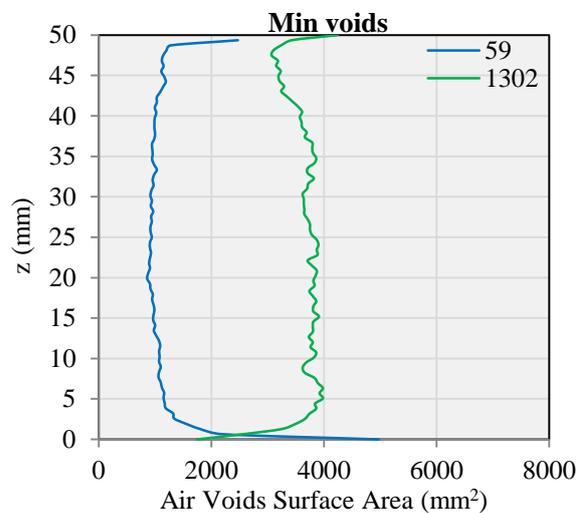


Figure 5.9 Void distribution along the specimen's height (set of specimens with the minimum void content from each phase)

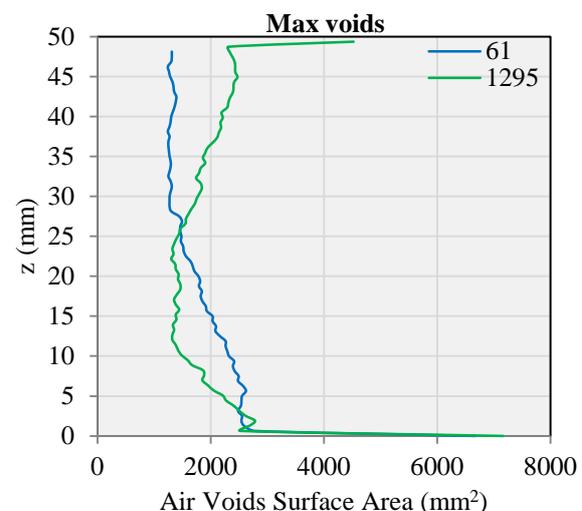


Figure 5.10 Void distribution along the specimen's height (set of specimens with the maximum void content from each phase)

distributions along the height, they can be considered identical.

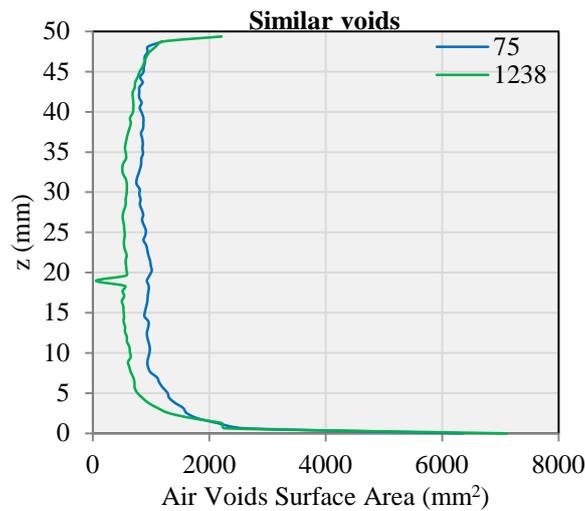


Figure 5.11 Void distribution along the specimen's height (set of specimens with similar void contents)

As it can be seen, the distribution of voids along the specimens' height does not show major differences between the lab produced and the field produced ones. Even though the number of samples is limited to safely draw a general conclusion, these results indicate that the two different compaction methods lead to similar void distributions.

5.3.2 Tests output from entire NL-Lab database

5.3.2.1 Analysis of Variance (ANOVA)

Following the same procedure as Paragraph 5.2.1, we aim to establish whether there is an interaction effect between two independent variables (Work and Phase). In other words, whether the effect of the different works on the ITS values, depends on the production phase, or the other way, whether the effect of production phase on ITS, depends on the work. From a physical point of view, the expectation is to find such a correlation. This process is done to ensure that there is also a statistical relation between the numbers, and in this way make our conclusions stronger. The steps and assumptions made are in this case done twice; once for the dry subset of specimens and once for the wet. It should be noted that since work 5 refers both to this thesis results, and the tests Dura Vermeer carried out, there is the distinction between 'W5-a' and 'W5-b'; the first referring to this thesis, and the latter to Dura Vermeer's batch.

- Detect outliers or extreme cases

Starting with the dry subset, the first step is recognizing the presence of possible outliers. These are values of the dependent variable, that have an unusually small or large value compared to the other dependent variable values. They are recognized by their residuals, which will have large absolute values (i.e., large positive or large negative values). It can be seen that in the case of the dry subset, no outliers are detected.

- Determine if the data are normally distributed

The next step is determining if the data is normally distributed. This can be checked using the Shapiro-Wilk test of normality. The results of the Shapiro-Wilk test are presented in the Tests of Normality Table 5.14. If the assumption of normality has been violated, the "Sig." value will be less than 0,05 (i.e., the test is significant at the $p < 0,05$ level). If the assumption of normality has not been violated, the "Sig." value will be greater than 0,05 (i.e. $p > 0,05$).

Table 5.14 Tests of Normality for the ITS values of the Dry subset

Work			Shapiro-Wilk		
			Statistic	df	Sig.
W1	F1	Residual for ITS	0,850	7	0,123
	F3	Residual for ITS	0,967	6	0,872
W2	F1	Residual for ITS	0,789	3	0,089
	F3	Residual for ITS	0,912	3	0,424
W3	F1	Residual for ITS	0,893	3	0,363
	F3	Residual for ITS	0,902	3	0,391
W4	F1	Residual for ITS	0,969	3	0,663
	F3	Residual for ITS	0,800	3	0,114
W5-a	F1	Residual for ITS	0,961	3	0,619
	F3	Residual for ITS	0,984	3	0,756
W5-b	F1	Residual for ITS	0,700	4	0,012
	F3	Residual for ITS	0,829	4	0,166

As it can be seen, there is significant evidence that the data in their majority are normally distributed as assessed by Shapiro-Wilk's test ($p > 0,05$). The only case the distribution is not normal, is at Phase 1 of Work 5-b (Dura Vermeer). However, as discussed before, the two-way ANOVA is considered "robust" to violations of normality and this means that some violation of this assumption can be tolerated and the test will still provide valid results. Therefore, there are often cases requiring only approximately normally distributed data. For this reason we consider the dataset normally distributed in its entirety.

- Determine the homogeneity of variances

The next step is determining the homogeneity of variances. The two-way ANOVA assumes that the (population) variances of the dependent variable (or residuals) are equal in all combinations of groups of the independent variables. SPSS Statistics tests this assumption by determining if there are equal

variances (called homogeneity of variances) in all combinations of groups of the two independent variables. The assumption of homogeneity of variances is tested using Levene's test of equality of variances, which is seen in Table 5.8.

Table 5.15 Levene's Test of Equality of Error Variances (Dry subset)

F	df1	df2	Sig.
2,486	11	33	0,021

We can see that the statistical significance level is .021 (i.e., $p = 0,021$), which indicates that Levene's test is statistically significant (because $p < 0,05$). A statistically significant result, such as this, indicates a significant evidence that we do not have equal (population) variances and we have violated the assumption of homogeneity of variances.

However, since we have a normally distributed dataset and the two-way ANOVA is somewhat robust in terms of homogeneity of variances too [28], we continue the analysis assuming homogeneous variances.

- Determine whether an interaction effect exists

The primary goal of running a two-way ANOVA is to determine whether there is an interaction between the two independent variables: Works and phases. As such, we are primarily interested in determining whether the effect of Phase of production is different for the 5 different works done. We can determine whether we have a statistically significant interaction effect from interpreting the table. The interaction effect is represented as the product of the two independent variables in a two-way ANOVA.

Table 5.16 Tests of Between-Subjects Effects table for dry subset

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	5,193	11	0,472	10,443	0,000	0,777
Intercept	299,287	1	299,287	6620,958	0,000	0,995
Work	3,188	5	0,638	14,105	0,000	0,681
Phase	0,358	1	0,358	7,910	0,008	0,193
Work * Phase	1,846	5	0,369	8,168	0,000	0,553
Error	1,492	33	0,045			
Total	327,111	45				
Corrected Total	6,684	44				

The "Sig." column presents the significance value (i.e., p-value) of the interaction effect. We can see that the p-value for this interaction effect² is .000 (i.e., $p = 0,000$). This is less than 0,05 (i.e., it satisfies $p < 0,05$), which means that there is a statistically significant interaction effect. This indicates that we were correct in thinking that the effect of Phase on ITS values depends on Work or, equivalently, the effect of Work on ITS values depends on Phase of production.

The same procedure was followed for the wet subset of specimens. The corresponding tables and figures required for the extraction of conclusions are presented below.

Table 5.17 Tests of Normality for the ITS values of the wet subset

Work			Shapiro-Wilk		
			Statistic	df	Sig.
W1	F1	Residual for ITS	0,975	7	0,935
	F3	Residual for ITS	0,899	6	0,371
W2	F1	Residual for ITS	0,999	3	0,952
	F3	Residual for ITS	0,991	3	0,817
W3	F1	Residual for ITS	0,964	3	0,637
	F3	Residual for ITS	0,980	3	0,726
W4	F1	Residual for ITS	0,980	3	0,726
	F3	Residual for ITS	0,818	3	0,157
W5-a	F1	Residual for ITS	0,938	3	0,520
	F3	Residual for ITS	0,923	3	0,463
W5-b	F1	Residual for ITS	0,998	4	0,994
	F3	Residual for ITS	0,927	4	0,577

Table 5.18 Levene's Test of Equality of Error Variances (Wet subset)

F	df1	df2	Sig.
4,605	11	33	0,000

Table 5.19 Tests of Between-Subjects Effects table for wet subset

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	9,666	11	0,879	16,444	0,000	0,846
Intercept	227,714	1	227,714	4261,370	0,000	0,992
Work	4,164	5	0,833	15,585	0,000	0,702
Phase	0,323	1	0,323	6,049	0,019	0,155
Work * Phase	5,488	5	1,098	20,540	0,000	0,757
Error	1,763	33	0,053			
Total	258,487	45				
Corrected Total	11,429	44				

² This is not interpreted as actually zero; in fact it indicates that $p < 0,0005$

The conclusions regarding the normality of the data and the homogeneity of variances are similar to the dry subset. The data are normally distributed, but not with a homogeneity in their variances. For the same reasons as in the dry subset, the analysis continued considering them as homogeneous.

It should also be noted that in the case of Work 1 – Phase 1, an outlier was detected. However, because it was not marked as an extreme, rather just as an outlier, and since it was only one case out of 45, the analysis continued without excluding it.

As in the case of the dry subset, the conclusion regarding the correlation of Works and Phases is the same. This indicates that we were correct in thinking that the effect of Phase on ITS values depends on Work or, equivalently, the effect of Work on ITS values depends on Phase of production. The fact that this statistically significant relation exists, will make the observations and conclusions stronger in terms of accuracy.

5.3.2.2 Lab to Field comparison

As it was mentioned in paragraph 1.2.5, work was distributed in two different labs to assess the differences in testing equipment. For this reason, the distinction between them during the analysis is going to be noted by W1a and W1b. Also, since work 5 is also carried by two labs (i.e. this thesis' research and Dura Vermeer), the notation W5a and W5b will serve the same purpose.

Extending the database, the image observed does not seem to be repeated. In paragraph 5.3.1, where only the test results carried out for this thesis were taken into account, lab specimens showed a superior moisture damage resistance, comparing to the field. However, this is not the case in the previous 4 works. Both in terms of ITS and ITSR, there is no consistent trend.

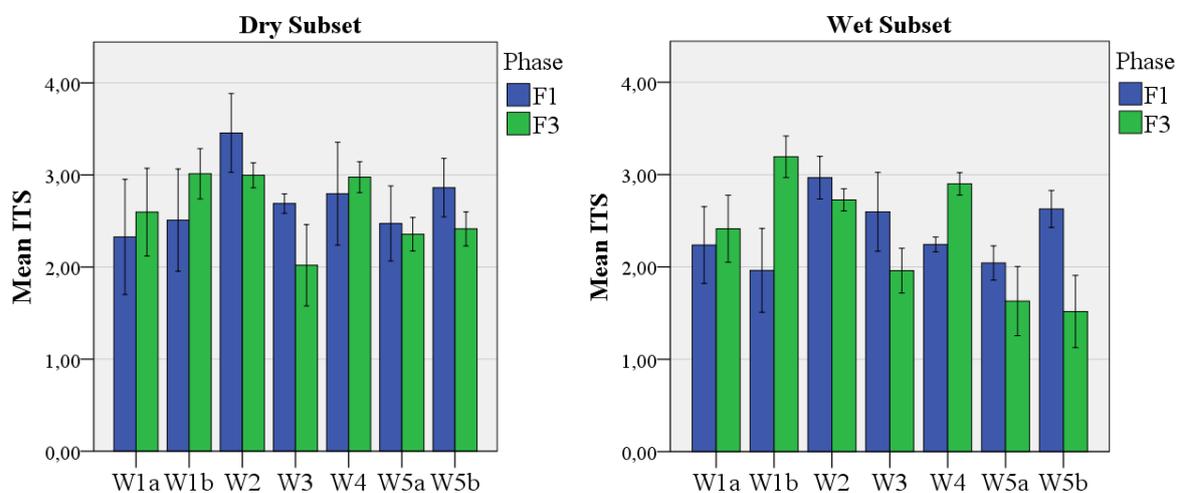


Figure 5.12 Mean ITS values for the Dry (left) and Wet subsets (right), divided in Works

Looking at Figure 5.12, both dry and wet, it is clear that the relation previously observed is altered. In Works 1 and 4 lab specimens present lower strength values than field, whereas in Works 2, 3 and 5, it is the opposite.

Comparing the ITS Ratios (Figure 5.13), the general conclusion is still the same. There are cases where lab produced mixtures perform better (Work 1a, 5a and 5b), and cases where field is better (Works 1b, 2, 3 and 4). The ITS relation in Works 2 and 3 does not seem to follow their ITS relation. Even though in terms of strength, lab values are much higher, the relative strength between wet and dry (i.e. ITS) is lower in the lab comparing to the field.

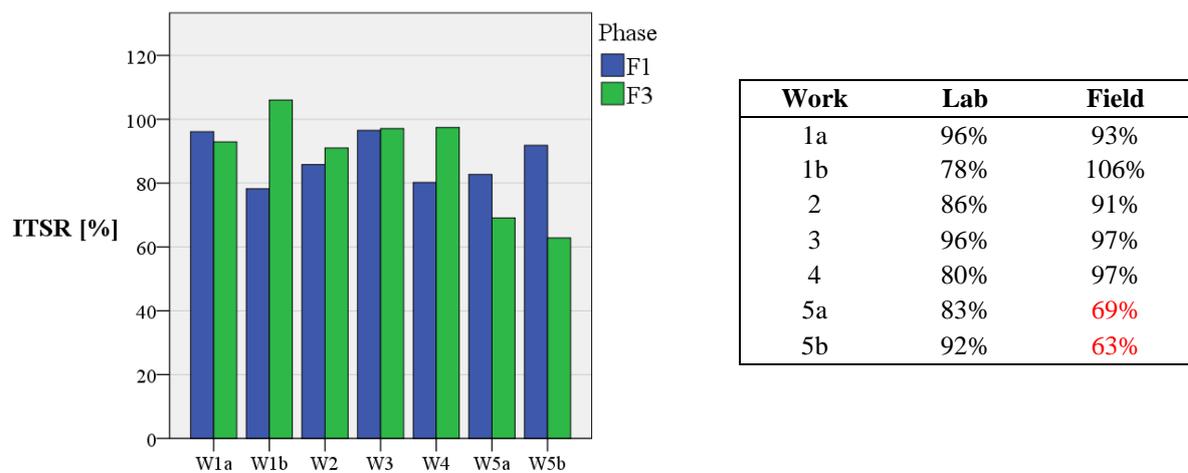


Figure 5.13 ITS values from all NL-Lab works

The field ITS values up to work 4 are consistently higher, with a very small reduction of their strength after conditioning. The corresponding lab values are lower in most cases, but at levels not very low for the mixture to be compared inadequate. In work 5 however, both labs came to very low field values, resulting in their failure, being lower than 70%. These very low values suggest a bad quality mixture in work 5, that when constructed under field conditions fails to achieve the desired resistance.

The fact that the lab values meet the requirements adequately, while the field fail to do so, is remarkable from the point that the same mixture produced in the two different ways leads to different directions in the design phase. In other words, if the design phase was based on field produced specimens, additional tests with modified composition would be required to fulfill the requirements, while according to the current procedure with lab produced specimens, the requirements would be more easily met.

5.3.3 Relation to composition parameters

5.3.3.1 Statistical Relation

The next step after defining the relation between the two phases, is tracing the explanations for the differences found. This search is done using the composition characteristics of each specimen, trying to identify trends and relations with their performance.

Initially, a common tool for detecting correlations is the Pearson product-moment correlation. It is used to determine the strength and direction of a linear relationship between two continuous variables. More specifically, the test generates a coefficient called the Pearson correlation coefficient, denoted as R , which measures the strength and direction of a linear relationship between two continuous variables. Its value can range from -1 for a perfect negative linear relationship to +1 for a perfect positive linear relationship. A value of 0 (zero) indicates no relationship between two variables. This test is also known by its shorter titles, the Pearson correlation or Pearson's correlation, which are often used interchangeably [29]. In our case we aim to study the relation between ITS values and each separate composition parameter (e.g. volumetrics and bitumen characteristics), and see how they behave relating to lab and field. The detailed methodology of calculating the Pearson's product-moment correlation follows the step-by-step SPSS statistics guide from the online platform *Laerd Statistics*.

In order to run a Pearson's correlation, there are five assumptions that need to be considered. The first two relate to the choice of study design and the measurements we chose to make, whilst the other three relate to how the data fits the Pearson correlation model. These assumptions are:

- Assumption 1: The two variables should be measured on a continuous scale (i.e., they are measured at the interval or ratio level). In our case, this assumption is satisfied for all parameters checked.
- Assumption 2: The two continuous variables should be paired (i.e., each case has two values – one for each variable). This is also satisfied, having for every specimen two paired values (ITS with composition parameters).
- Assumption 3: There needs to be a linear relationship between the two variables. The best way of checking this assumption is to plot a scatterplot and visually inspect the graph. In order to check this, the dataset was divided based on phase and condition, to exclude the conditioning effect on the results. The scatterplots of the ITS values against each parameter are presented in the set of figures in Figure 5.14.

Starting from *F1 – Dry*, we can assume that the only case where a linear relationship is detected is in the case of V_{bit} . Bitumen characteristics ($T_{R\&B}$, Pen, logA, G^* , δ and ZSV) show a vertical concentration of points at their lows or highs, depending on the case, with a small

number of points lying far from that concentration. In addition, the volumetrics (G_{mb} , VMA, VFA and V_m) have a 'cloud' distribution, with no visual trend in their scatter. For this reason, F1- Dry subset is not considered linear in its relation between ITS and the composition parameters.

In the case of *F3 - Dry* on the other hand, there is a medium linear trend in the volumetrics, while the bitumen characteristics do not have the same concentrated image, as F1 - Dry. However, they still cannot be considered linear.

The image in F1- Wet and F3 - Wet is similar to the dry subsets. While in the F1's volumetric parameters there seems to be an improvement in the linearity, the 'cloud' is still considerable to be assumed linear. The F3 subset still has a fairly good linear behavior in its volumetrics, and in this case also in the bitumen characteristics.

The general conclusion regarding the linear relationship between ITS and the composition parameters is that it is not at levels adequate enough to be considered linear. This means that assumption 3 of the Pearson correlation is violated.

Assumption 4: Assumption of bivariate normality, or in other words, there should be a normal distribution in each parameter. The Shapiro-Wilk non parametric test is recommended for testing this assumption. If the assumption of normality has been violated, the "Sig." value will be less than 0,05 (i.e., the test is significant at the $p < 0,05$ level). If the assumption of normality has not been violated, the "Sig." value will be greater than 0,05 (i.e., $p > 0,05$). This is because the Shapiro-Wilk test is testing the null hypothesis that the data's distribution is equal to a normal distribution. Rejecting the null hypothesis means that the data's distribution is not equal to a normal distribution. The test's output by SPSS is in Table 5.21 and Table 5.22 for the dry and wet subsets respectively.

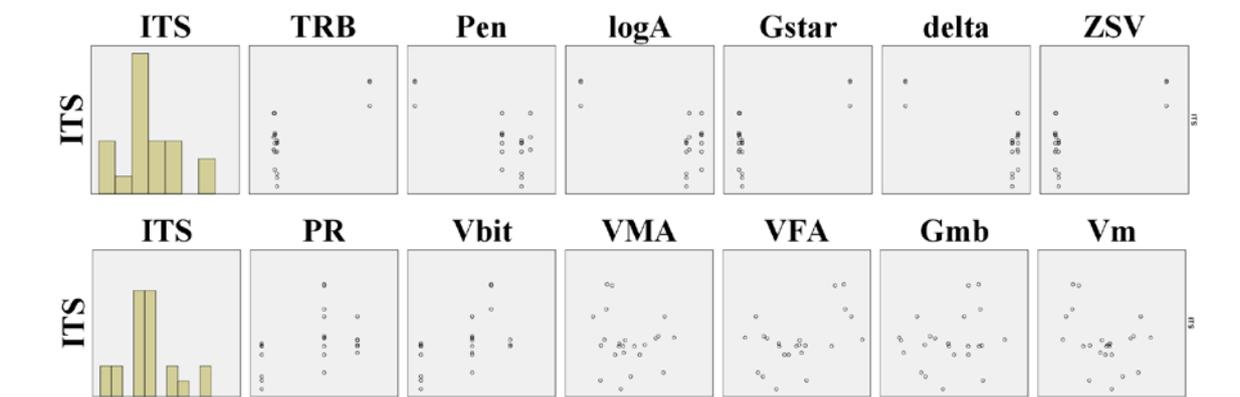
Table 5.20 Test of Normality of the data (Dry subset)

Phase	Shapiro-Wilk			Phase	Shapiro-Wilk			
	Statistic	df	Sig.		Statistic	df	Sig.	
F1	ITS	0,948	20	0,343	ITS	0,881	19	0,022
	TRB	0,463	20	0,000	TRB	0,782	19	0,001
	Pen	0,664	20	0,000	Pen	0,819	19	0,002
	logA	0,564	20	0,000	logA	0,828	19	0,003
	G*	0,462	20	0,000	G*	0,835	19	0,004
	δ	0,484	20	0,000	δ	0,808	19	0,001
	ZSV	0,462	20	0,000	ZSV	0,835	19	0,004
	PR	0,773	20	0,000	PR	0,766	19	0,000
	V_{bit}	0,742	20	0,000	V_{bit}	0,797	19	0,001
	VMA	0,935	20	0,193	VMA	0,958	19	0,534
	VFA	0,932	20	0,167	VFA	0,906	19	0,064
	G_{mb}	0,920	20	0,100	G_{mb}	0,964	19	0,662
	V_m	0,967	20	0,687	V_m	0,953	19	0,444
F3	ITS	0,948	20	0,343	ITS	0,881	19	0,022
	TRB	0,463	20	0,000	TRB	0,782	19	0,001
	Pen	0,664	20	0,000	Pen	0,819	19	0,002
	logA	0,564	20	0,000	logA	0,828	19	0,003
	G*	0,462	20	0,000	G*	0,835	19	0,004
	δ	0,484	20	0,000	δ	0,808	19	0,001
	ZSV	0,462	20	0,000	ZSV	0,835	19	0,004
	PR	0,773	20	0,000	PR	0,766	19	0,000
	V_{bit}	0,742	20	0,000	V_{bit}	0,797	19	0,001
	VMA	0,935	20	0,193	VMA	0,958	19	0,534
	VFA	0,932	20	0,167	VFA	0,906	19	0,064
	G_{mb}	0,920	20	0,100	G_{mb}	0,964	19	0,662
	V_m	0,967	20	0,687	V_m	0,953	19	0,444

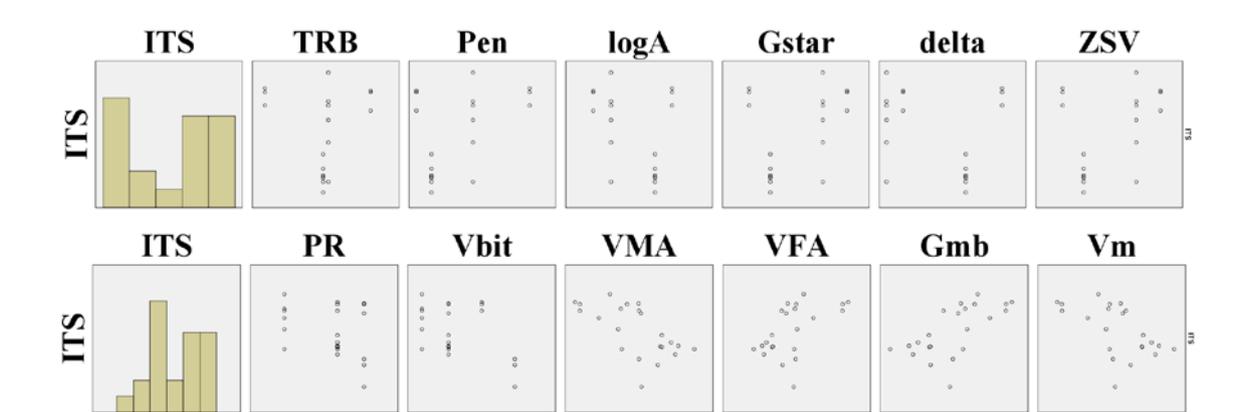
Table 5.21 Test of Normality of the data (Wet subset)

Phase	Shapiro-Wilk			Phase	Shapiro-Wilk				
	Statistic	df	Sig.		Statistic	df	Sig.		
F1	ITS	0,957	20	0,486	ITS	0,910	19	0,075	
	TRB	0,463	20	0,000	TRB	0,782	19	0,001	
	Pen	0,664	20	0,000	Pen	0,819	19	0,002	
	logA	0,564	20	0,000	logA	0,828	19	0,003	
	G*	0,462	20	0,000	G*	0,835	19	0,004	
	δ	0,484	20	0,000	δ	0,808	19	0,001	
	ZSV	0,462	20	0,000	F3	ZSV	0,835	19	0,004
	PR	0,773	20	0,000	PR	0,766	19	0,000	
	V _{bit}	0,742	20	0,000	V _{bit}	0,797	19	0,001	
	VMA	0,946	20	0,316	VMA	0,944	19	0,317	
	VFA	0,934	20	0,181	VFA	0,909	19	0,070	
	G _{mb}	0,956	20	0,472	G _{mb}	0,939	19	0,250	
	V _m	0,963	20	0,599	V _m	0,942	19	0,292	

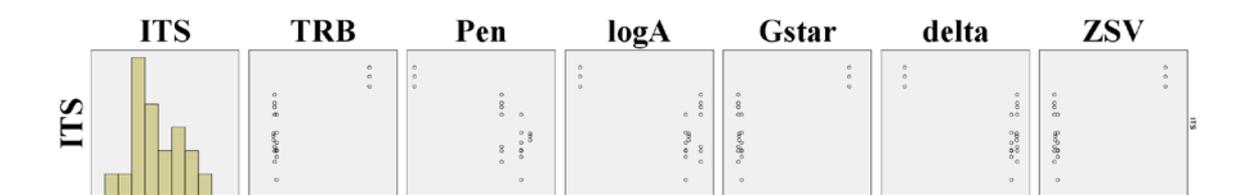
F1 – Dry



F3 - Dry



F1 – Wet



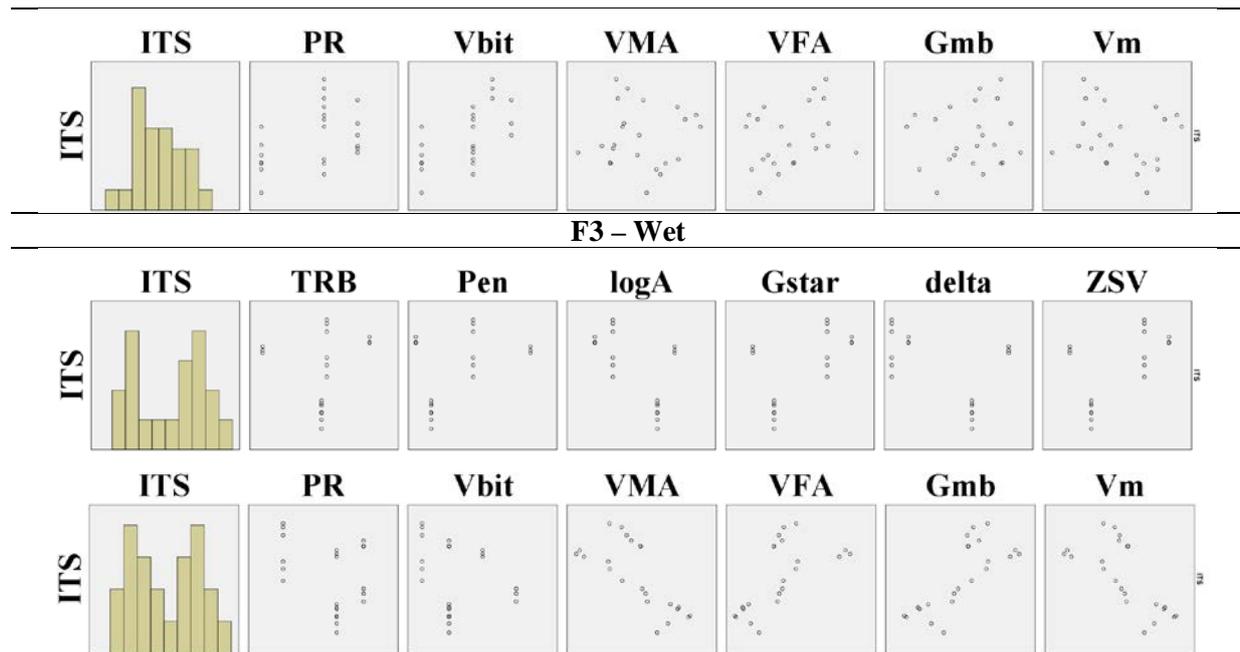


Figure 5.14 Scatterplots for the inspection of linearity between ITS and composition parameters

According to the Shapiro-Wilk test, not all variables are normally distributed. The variables that satisfy this hypothesis are marked with bold numbers. This means that the fourth assumption is also violated

- Assumption 5: There should be no significant outliers. According to SPSS there are outliers in the F1 cases of $T_{R\&B}$, Pen, logA, G^* , δ and ZSV. Even though from a physical point of view these values might be explained by other parameters, from a statistical point they have a negative effect on the Pearson correlation analysis, hence this assumption is also violated.

Having 3 out of 5 assumptions not satisfied, the conclusion is that Pearson's product-moment correlation cannot be utilized in this case. The next is following a test which is 'distribution-free' and has less assumptions; the Spearman's rank-order correlation.

The Spearman's rank-order correlation (which is the non-parametric version of Pearson product-moment correlation) calculates a coefficient, r_s or ρ , which is a measure of the strength and direction of the association/relationship between two continuous or ordinal variables [30]. The Spearman correlation evaluates the monotonic relationship between two continuous or ordinal variables. In a monotonic relationship, the variables tend to change together, but not necessarily at a constant rate. A monotonic relationship is a relationship that does one of the following: (a) as the value of one variable increases, so does the value of the other variable; or (b) as the value of one variable increases, the other variable value decreases. The Spearman correlation coefficient is based on the ranked values for

each variable rather than the raw data. In our case, the relationship between ITS and all the composition parameters available is studied, including all 5 works available.

The output of the analysis, generated by SPSS, is the correlation coefficient value r_s , which is a measure of the strength and direction of the association between the two variables. The correlation coefficient can take values from +1 to -1, which indicates a perfect positive (+1) or negative (-1) association of ranks. A correlation coefficient of zero (0) indicates no association between the ranks.

The first two assumptions needed to carry out this test are common to the Pearson's correlation assumptions, which means having continuous variables and paired observations. The third and final assumption refers to the monotonic relationship between the variables. To determine if a monotonic relationship exists, we need again to visually inspect the scatterplot of the two variables seen in the set of figures in Figure 5.14. Despite them not being explicitly monotonic by sight, we will continue with this assumption and make another check later to confirm it.

The next step in interpreting the results is to determine whether the Spearman's rank-order correlation coefficient value is statistically significant. This will allow to determine whether we can accept or reject the null hypothesis. If we set $\alpha = 0,05$ (i.e., $p < 0,05$), achieving a statistically significant Spearman rank-order correlation means that there is less than a 5% chance that the strength of the relationship found (the correlation coefficient) happened by chance, if the null hypothesis were true. Thus p-values lower than 0,05 indicate a statistically significant relation [30]. In Table 5.22 there are cases where p-value is 0,000. This is not interpreted as actually zero; in fact it indicates that $p < 0,0005$. It should be noted that statistical significance does not determine the strength of the relationship (r or ρ does that), but whether the correlation coefficient is statistically significantly different from zero. In Table 5.22 the result from the Spearman's correlation analysis on the composition parameters relating to ITS are presented, distinguished by phase and condition. In the cases where p-values are lower than the 0,05 limit, the numbers are bold.

Table 5.22 Spearman's correlation of composition parameters with ITS, distinguished by phase and condition

	Lab – F1				Field – F3			
	Dry		Wet		Dry		Wet	
	r_s	p-value	r_s	p-value	r_s	p-value	r_s	p-value
VFA	0,288	0,182	0,274	0,206	0,555	0,007	0,654	0,001
V_m	-0,257	0,237	-0,241	0,269	-0,572	0,005	-0,688	0,000
G_{mb}	0,060	0,786	0,063	0,774	0,664	0,001	0,747	0,000
VMA	-0,070	0,750	0,048	0,826	-0,654	0,001	-0,747	0,000
logA	-0,171	0,435	-0,362	0,089	-0,185	0,409	-0,545	0,009
$T_{R\&B}$	0,154	0,482	0,399	0,059	-0,232	0,298	0,293	0,185
δ	-0,079	0,720	-0,310	0,150	0,302	0,172	-0,262	0,239
V_{bit}	0,558	0,006	0,647	0,001	-0,319	0,147	-0,292	0,187
ZSV	0,122	0,607	0,189	0,424	0,210	0,389	0,498	0,030
PR	0,394	0,063	0,374	0,079	-0,257	0,248	-0,175	0,437
G^*	0,079	0,720	0,005	0,982	0,498	0,018	0,485	0,022
Pen	-0,376	0,077	-0,309	0,152	0,544	0,009	0,318	0,150

The most striking information coming out of this analysis is the fact that Lab ITS values cannot be statistically related to the mixture's composition. The statistical significance (p-value) of all the composition parameters, besides bitumen content (V_{bit}), is at levels high enough to be considered insignificant.

On the other hand, Field values, both dry and wet, show a strong and significant correlation to composition parameters. Even though some of these parameters are directly interrelated to each other in a physical sense (VFA, V_m , VMA and G_{mb}), there are also bitumen characteristics that show a strong association to ITS values.

Considering the fact that Lab and Field mixtures were produced under the same material and composition characteristics, the oddity of this observation draws the attention. Since chemical and physical interaction between the components of a mixture determine its mechanical properties, the expectation would be to find similar correlations. The fact that this did not happen however, suggests that this interaction took place in a different way for each phase. These differences could lie in the production temperature, the energy put in mixing and compacting, or factors that are unidentifiable (e.g. equipment operator's choices, etc.). The composition and bitumen parameters studied in this research, apparently were unable to capture these differences, suggesting that additional information should be included for the deeper understanding of this correlation.

Coming back to the assumption of monotonic relation, a method to confirm that this assumption was correct is doing another correlation analysis, which does not require this relation strictly. This correlation is tested with Kendall's tau-b.

Kendall's tau-b (τ_b) correlation coefficient (Kendall's tau-b, for short) is a nonparametric measure of the strength and direction of association that exists between two variables measured, on at least an ordinal scale. It is considered a nonparametric alternative to the Pearson's correlation when the data has failed one or more of the assumptions of this test. It is also considered an alternative to the nonparametric Spearman's correlation. Also, it determines whether there is a monotonic relationship between two variables. As such, it is desirable if the data would appear to follow a monotonic relationship, so that formally testing for such an association makes sense, but it is not a strict assumption, or one that we are often able to assess. The Kendall's tau-b correlation coefficients are seen in Table 5.23.

Table 5.23 Kendall's tau-b correlation coefficients

	Lab – F1				Field – F3			
	Dry		Wet		Dry		Wet	
	τ_b	p-value	τ_b	p-value	τ_b	p-value	τ_b	p-value
VFA	0,176	0,244	0,202	0,178	0,401	0,009	0,506	0,001
V _m	-0,148	0,327	-0,187	0,214	-0,410	0,008	-0,529	0,001
G _{mb}	0,032	0,832	0,083	0,579	0,471	0,002	0,567	0,000
VMA	-0,008	0,958	0,044	0,771	-0,461	0,003	-0,593	0,000
logA	-0,121	0,459	-0,275	0,089	-0,121	0,465	-0,434	0,009
T _{R&B}	0,103	0,528	0,302	0,062	-0,141	0,397	0,174	0,293
δ	-0,040	0,805	-0,195	0,228	0,219	0,189	-0,174	0,293
V _{bit}	0,448	0,007	0,502	0,003	-0,247	0,147	-0,230	0,174
ZSV	0,055	0,758	0,110	0,538	0,190	0,302	0,343	0,060
PR	0,286	0,093	0,270	0,111	-0,228	0,189	-0,095	0,583
G*	0,040	0,805	-0,036	0,827	0,413	0,013	0,318	0,054
Pen	-0,283	0,087	-0,223	0,176	0,442	0,008	0,232	0,161

Having followed the same hypothesis as the Spearman's correlation, achieving a statistically significant tau-b correlation means that there is less than a 5% chance that the strength of the relationship found (the correlation coefficient) happened by chance, if the null hypothesis were true. Thus p-values lower than 0,05 indicate a statistically significant relation. As it can be seen in Table 5.22, the result is similar to Spearman's correlation, meaning that we were safe in our assumption and conclusions from it.

5.4 Performance prediction

5.4.1 Introduction

Based on the composition parameters collected for every single specimen tested, and its measured performance in moisture damage resistance, a function describing the relation between these two will be obtained. As described in Paragraph 2.1.2 this function aims to help in the preliminary design of a mixture, when its composition has to be roughly determined in order to satisfy the certain requirements.

The tool used for this purpose is the Multiple Linear Regression Analysis. All the necessary assumptions and checks that are required in order to carry it out were described in paragraph 3.3.1 and are going to be applied in this chapter.

Since moisture sensitivity is described by the ratio of wet and dry ITS values, the target of the regression analysis will be to obtain two predicting relations, one for a dry specimen (Eq. 5.1), and one

for a wet (Eq. 5.2), and then use the predicted dry and wet strengths to calculate the predicted ratio (Eq. 5.3).

$$ITS_{dry,predicted} = \beta_o + \beta_1 \cdot y_1 + \beta_2 \cdot y_2 + \dots + \beta_i \cdot y_i \quad [MPa] \quad \text{Eq. (5.1)}$$

$$ITS_{wet,predicted} = \alpha_o + \alpha_1 \cdot x_1 + \alpha_2 \cdot x_2 + \dots + \alpha_i \cdot x_i \quad [MPa] \quad \text{Eq. (5.2)}$$

$$ITSR_{predicted} = \frac{ITS_{wet,predicted}}{ITS_{dry,predicted}} \quad [\%] \quad \text{Eq. (5.3)}$$

5.4.2 Data included

In the case of moisture sensitivity, during the course of this thesis, 3 different conditioning protocols were followed (Standard, MIST and Frost Damage). The entire NL-Lab database that is going to be used for the performance prediction, additionally to this thesis' data, was obtained on the basis of the Standard conditioning method. For this reason, the specimens from this thesis included in the regression data, will be limited to just the conditioning protocol, in order to exclude the factor of different conditioning damages that are reflected in the different ITS values.

Also, in order to extend the database, all works and phases available are going to be taken into account, including F1, F2, F3 and F3a. Ideally, the differences in their composition parameters are linked to the differences in their performance.

5.4.3 Analysis and checks

The procedure described follows the step-by-step multiple regression analysis by '*Laerd Statistics (2015). Statistical tutorials and software guides*' [21].

For the multiple linear regression to be considered appropriate for use, there are certain assumptions that need to be checked. These assumptions refer both to the nature of the dataset, and their numerical behavior in terms of distribution, linearity and collinearity. The first two of them are checked before the regression takes place, and are fulfilled for all the cases of this thesis. One refers to the dependent variable (i.e. ITS), and that it should be measured at the continuous level, and the other to the independent variables (i.e. composition parameters), also named regressors or predictors, and that they should be measured either at the continuous or at the nominal level. Both of them are satisfied.

The remaining six assumptions can only be checked after the analysis is run. The initial run, that will be used as the starting point of the analysis, will include in the relation all the available composition parameters. Based on the output and the checks, some parameters will be excluded and various combinations will be tested, to find the one that both optimally satisfies the regression requirements,

and has a physical explanation. The checks of this first run and the entire regression procedure, will be described in detail, referring to just the wet subset of specimens. After that, the results will be comparatively presented as an overview. The procedure and the necessary checks will not be presented again, but will be included in the Appendix instead.

- Independence of observations

The independence of observations is statistically tested using the Durbin-Watson test. To provide some background, the Durbin-Watson test is a test for a particular type of (lack of) independence; namely, 1st-order autocorrelation, which means that adjacent observations (specifically, their errors) are correlated (i.e., not independent). The Durbin-Watson statistic is generated by the SPSS Model Summary Table 5.24.

Table 5.24 Model Summary Table

Model	R	R ²	Adjusted R ²	Std. Error of the Estimate	Durbin-Watson
1	0,825	0,681	0,618	0,31208	1,368

The Durbin-Watson statistic for this analysis is 1,368. The Durbin-Watson statistic can range from 0 to 4, but we are looking for a value of approximately 2 to indicate that there is no correlation between residuals. We can see that our value is relatively closer to 2 than 0 or 4, so it can be accepted that there is independence of errors (residuals).

- Linearity

An assumption of multiple regression is that: (a) the independent variables collectively are linearly related to the dependent variable; and (b) each independent variable is linearly related to the dependent variable.

The first is checked by plotting a scatterplot of the studentized residuals against the (unstandardized) predicted values (Figure 5.15). If the residuals form a horizontal band, as in the scatterplot below, the relationship between your dependent variable and independent variables is likely to be linear.

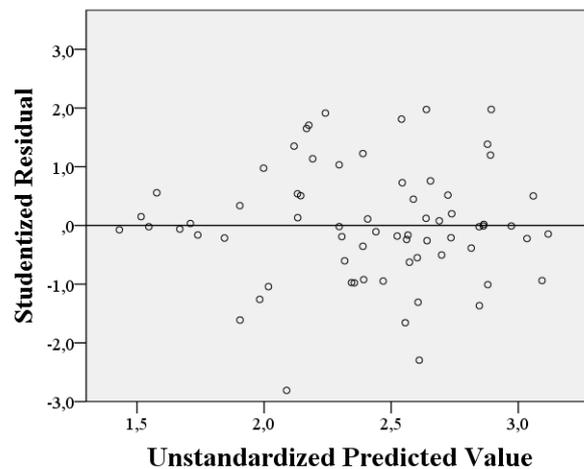


Figure 5.15 Studentized residuals against the predicted values

The linearity of each individual independent parameter with ITS was checked in paragraph 5.3.3.1, Figure 5.14, with the partial scatterplots, during the comparison between lab and field produced specimens. It was noticed that the cases where linearity was observed was mainly in the volumetric properties of F3 specimens. In cases where no linearity exists, we can consider transformations of either the independent or dependent variable or both, to achieve a linear relation. These transformations include for example the values in their logarithmic or squared form. Our scatters though seem to follow a ‘cloud’ or vertical scatter, rather than a logarithmic or square. Hence, no transformation can be applied, and the assumption of linearity between the dependent and independent variables is violated.

However, since a collective linearity exists, and there are indications of a fairly good predictability, judging from the adjusted R^2 from Table 5.24, we choose to continue with the analysis. This assumption will naturally reduce the strength of the relation produced, but it will still provide useful information.

- Homoscedasticity

The assumption of homoscedasticity is that the residuals are equal for all values of the predicted dependent variable. To check for heteroscedasticity, we can use the plot created to check linearity in the previous section, namely plotting the studentized residuals against the unstandardized predicted values (Figure 5.15).

If there is homoscedasticity, the spread of the residuals will not increase or decrease as we move across the predicted values (i.e., the points of the plot above will exhibit no pattern and will be approximately constantly spread). However, if the residuals are not evenly spread, but differ in height (e.g., a funnel shape), we do not have homoscedasticity. Instead, we have what is called heteroscedasticity (i.e., we have violated the assumption of homogeneity of variance).

The residuals in our case appear randomly scattered. On this basis, it would appear that the assumption of homoscedasticity has been met.

- Multicollinearity

Multicollinearity occurs when we have two or more independent variables that are highly correlated with each other. This leads to problems with understanding which variable contributes to the variance explained and technical issues in calculating a multiple regression model. In our case, bitumen characteristics (i.e. Pen, $T_{R\&B}$, logA, G^* , δ and ZSV) and volumetric properties (i.e. G_{mb} , Target G_{mb} , V_m , VMA, VFA, and V_{bit}) are expected to show multicollinearity, since some of them are directly related by the calculation of one from the other, and some of them are physically related. For this reason, it has to be determined which combination of them will be concluded. This multicollinearity can be recognized by the Tolerance/VIF values generated by SPSS for each regressor separately (Table 5.25).

In reality, as VIF is simply the reciprocal of Tolerance (i.e., 1 divided by Tolerance), we need only consult one of these measures. If the Tolerance value is less than 0,1 – which is a VIF of greater than 10 – we might have a collinearity problem.

Table 5.25 Coefficients statistics for the wet subset

	Unstandardized Coefficients		Sig.	Collinearity Statistics	
	B	Std. Error		Tolerance	VIF
(Constant)	-87,540	34,828	0,015	-	-
$T_{R\&B}$	0,145	0,259	0,578	0,001	906,545
Pen	-0,009	0,108	0,937	0,009	106,262
logA	-8,318	23,483	0,725	0,003	342,715
δ	-0,065	0,042	0,122	0,026	37,788
ZSV	-6,826E-06	0,000	0,002	0,011	87,307
V_{bit}	1,123	0,874	0,204	0,164	6,111
VFA	0,029	0,072	0,682	0,005	186,030
Target G_{mb}	0,005	0,009	0,546	0,377	2,653
G_{mb}	0,023	0,005	0,000	0,079	12,654
V_m	0,360	0,434	0,410	0,005	219,463

As expected, due to the inclusion of all the parameters available, a multicollinearity problem exists. Eight out of the eleven regressors exceeded the VIF limit of 10. This means that adjustments should be made on the model selected. It has to be noted, that due to extreme VIF values, the regressors VMA and G^* were automatically excluded by SPSS from the model.

- Checking for unusual points

There can be certain data points that are, in some way, classified as unusual from the perspective of fitting a multiple regression model. These data points are generally detrimental to the fit or

generalization (statistical inference) of the regression equation. There are three main types of unusual points: outliers, high leverage points and highly influential points.

a) Outliers

An outlier is an observation (data point) that does not follow the usual pattern of points (they are far away from their predicted value). There are different types of residuals that can be used to detect outliers: standardized residuals, studentized residuals or studentized deleted residuals. The default option that SPSS Statistics produces, is to use standardized residuals.

The 'Casewise Diagnostics' table of the analysis highlights any cases, where that case's standardized residual is greater than ± 3 standard deviations, which is what SPSS Statistics treats as an outlier. A value of greater than ± 3 is a common cut-off criteria used to define whether a particular residual might be representative of an outlier or not. If all cases have standardized residuals less than ± 3 , this table will not be produced as part of the SPSS Statistics output, which is what happened in this case. The next step is checking for leverage points.

b) Leverage points

An output of the SPSS's function, is the leverage of each data point separately. To determine whether any cases exhibit high leverage, a general rule of thumb is to consider leverage values less than 0,2 as safe, 0,2 to less than 0,5 as risky, and values of 0,5 and above as dangerous. Sorting the data points in a descending order based on their leverage, the maximum leverage obtained is 0,36265. In addition, there are several data points whose leverage lies between 0,2 and 0,3, considering them as slightly risky. In this case, the decision of treating them or not, will be made in combination with their influence, which is checked below by Cook's distance. If they lead to high influence, then these cases need to be filtered-out and the regression re-run.

c) Influential points

Another output of the SPSS regression function, is each data point's Cook Distance. As a rule of thumb, if there are Cook's Distance values above 1, they should be investigated. Sorting these values in a descending order, the highest Cook's Distance obtained was 0,12119, which indicates that no point is in risk of being considered highly influential.

In combination with the leverage points, we conclude that no data point is risky enough to be removed.

- Normality

In order to be able to run inferential statistics (i.e., determine statistical significance), the errors in prediction – the residuals – need to be normally distributed. A common method we can use to check for the assumption of normality of the residuals is a histogram with superimposed normal curve (Figure 5.16) and a P-P Plot (Figure 5.16), which were both produced by the Linear Regression function.

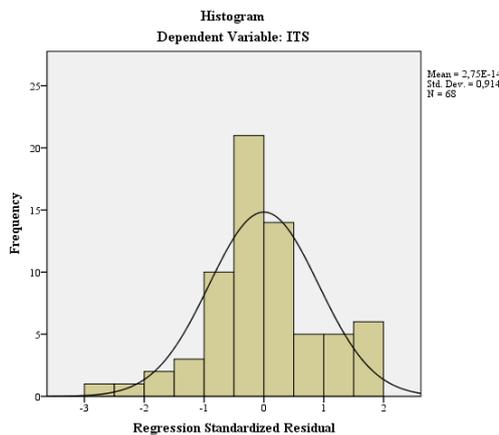


Figure 5.16 Distribution histogram of residuals

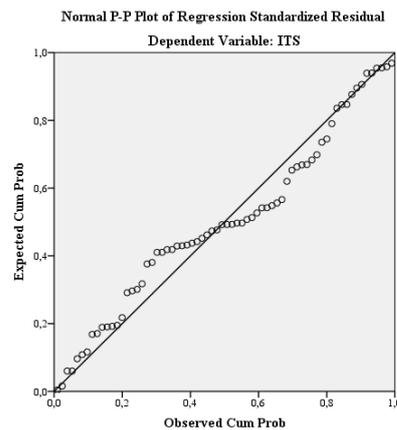


Figure 5.17 P-P plot

We can see from the histogram that the standardized residuals appear to be approximately normally distributed. The mean and standard deviation should have values of approximately 0 (zero) and 1, respectively, which is also satisfied. To confirm the findings we should really look at a P-P Plot.

If the residuals are normally distributed, the points will be aligned along the diagonal line. In reality, the points will never be perfectly aligned along the diagonal line. Moreover, we only need the residuals to be approximately normally distributed because regression analysis is fairly robust to deviations from normality.

We can see from the P-P Plot above that although the points are not aligned perfectly along the diagonal line (the distribution is somewhat peaked), they are close enough to indicate that the residuals are close enough to normal, for the analysis to proceed. As multiple regression analysis is fairly robust against deviations from normality, we can accept this result as meaning that no transformations need to take place.

After checking the assumptions we concluded that, firstly adjustments in the parameters included should be made due to collinearity problems, and secondly, we neglect the linearity problem between the dependent and independent values. The next step is to determine whether the multiple regression model is a good fit for the data. There are a number of statistics you can use to determine whether the multiple regression model is a good fit for the data. These are: (a) the multiple correlation coefficient,

R, (b) the percentage (or proportion) of variance explained, R^2 and Adjusted- R^2 , (c) the statistical significance of the overall model and the predictors, and (d) the precision of the predictions from the regression model.

a. Multiple Correlation Coefficient (R)

Since the multiple correlation coefficient R, or simply the Pearson correlation coefficient between the ITS values predicted by the regression and the actual values of the dependent variable, is not a common measure used to assess goodness of fit, we will only go into detail at the other statistics.

b. Total variation explained (R^2 and Adjusted R^2)

The coefficient of determination – more commonly known as R^2 – is a measure of the proportion of variance in the dependent variable that is explained by the independent variable. More specifically, it is the proportion of variance in the dependent variable that is explained by the independent variables over and above the mean model. It is computed by Eq. 3.5, and is seen in Table 5.26.

Table 5.26 Wet ITS model's quality assessment parameters

R	R^2	Adjusted R^2
0,825	0,681	0,618

We can see that R^2 is equal to 0,681. This means that the addition of all our independent variables into a regression model explained 68,1% (i.e., $0,681 \times 100 = 68,1\%$) of the variability of our dependent variable. However, R^2 is based on the sample and is considered a positively-biased estimate of the proportion of the variance of the dependent variable accounted for by the regression model (i.e., it is larger than it should be when generalizing to a larger population. There is another measure called adjusted R^2 , which corrects for this positive bias in order to provide a value that would be expected in the population (Eq. 3.6). We can see that adjusted R^2 is 0,618, which is a quite accurate prediction. Adjusted R^2 will always be smaller than R^2 , but it is preferable that we use this value to report the proportion of variance explained (i.e., report 61,8% rather than 68,1%).

A useful visual representation of the model fit, is plotting the measured dependent value against the predicted one, in a scatterplot (Figure 5.18). The closer the points fall to the equality line (i.e. line where measured and predicted are perfectly equal), the more accurate the fit.

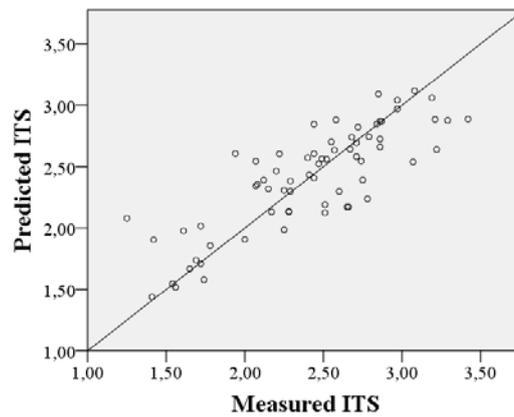


Figure 5.18 Equality scatterplot

We can say that the fairly good Adjusted R^2 values, are also confirmed by the image of the equality plot, with a large proportion of the points lying very close to the equality line.

c. Statistical Significance of the Model

The statistical significance of the overall model (i.e., the model containing all independent variables) is presented in the "Sig." column of Table 5.27.

Table 5.27 ANOVA table of the model tested

	Sum of Squares	df	Mean Square	F	Sig.
Regression	11,623	11	1,057	10,848	0,000
Residual	5,454	56	0,097		
Total	17,077	67			

We can see that the "Sig." value is 0,000, which actually means that $p < 0,0005$. If $p < 0,05$, we have a statistically significant result. On the other hand, if $p > 0,05$, we do not have a statistically significant result. This means that the addition of all our independent variables (i.e., our overall model) leads to a model that: (a) is statistically significantly better at predicting the dependent variable than the mean model; and (b) is a statistically significantly better fit to the data than the mean model.

Regarding the statistical significance of the predictors, this can be determined by the 'Sig.' column of Table 5.28. Sig. which refers to p-value, has to be smaller than 0,05 for the coefficient to be considered statistically significant.

Table 5.28 Coefficients statistics for the wet subset

	Unstandardized Coefficients		Sig.	Collinearity Statistics	
	B	Std. Error		Tolerance	VIF
(Constant)	-87,540	34,828	0,015	-	-
T _{R&B}	0,145	0,259	0,578	0,001	906,545
Pen	-0,009	0,108	0,937	0,009	106,262
logA	-8,318	23,483	0,725	0,003	342,715
δ	-0,065	0,042	0,122	0,026	37,788
ZSV	-6,826E-06	0,000	0,002	0,011	87,307
PR	-0,004	0,012	0,750	0,267	3,748
V _{bit}	1,123	0,874	0,204	0,164	6,111
VFA	0,029	0,072	0,682	0,005	186,030
Target G _{mb}	0,005	0,009	0,546	0,377	2,653
G _{mb}	0,023	0,005	0,000	0,079	12,654
V _m	0,360	0,434	0,410	0,005	219,463

We can see that the only regressors which are considered to be statistically significant are the Zero-Shear viscosity and the bulk density. This observation is not surprising, having in mind the collinearity problem that exists in this specific model. The choice of a new model, with parameters that do not show collinearity, and at the same time, are statistically significant, is necessary.

5.4.4 Predictors combinations

Several regression analyses were ran, each time with a different combination, based on the parameter's significance and collinearity. Since Bulk Density, G_{mb}, showed the best linearity with ITS, and also a high level of significance in the first trial (p-value = 0,000), the combinations were focused on it. Also, Zero Shear Viscosity, ZSV, showed a good significance level in the first try (p-value = 0,002), so it was the second regressor that the combinations evolved around.

All the assumptions were also checked after each combination. Except for the linearity problem of the individual independent variables with the dependent, which was addressed previously, all the assumptions were satisfied, hence the final choice of them was only a matter of their statistical significance in the model (Sig.), and their collinearity factor (VIF). Some representative sets of parameters, including the regressor's significance and collinearity, the model's coefficients of variation and significance, and the independence of the values (Durbin Watson), are presented in Table 5.29. All the sets that were investigated can be found in the Appendix. The coefficients (i.e. multiplier of each parameter in the model) are going to be presented for the final model chosen.

Table 5.29 Representative combinations of predictors for the wet subset

Wet Subset												
Set	1		2		3		4		5		6	
Pred.	Sig.	VIF	Sig.	VIF	Sig.	VIF	Sig.	VIF	Sig.	VIF	Sig.	VIF
(const.)	0,015	-	0,000	-	0,000	-	0,000	-	0,000	-	0,000	-
T _{R&B}	0,578	906,5			0,000	37,142	0,000	37,151			0,001	153,9
Pen	0,937	106,2									0,001	29,30
logA	0,725	342,7									0,000	73,07
G*												
δ	0,122	37,780							0,002	9,605		
ZSV	0,002	87,307	0,002	1,010	0,001	37,453	0,001	37,758	0,065	9,567		
V _{bit}	0,204	6,111					0,315	1,210				
VMA												
VFA	0,682	186,0										
Target G _{mb}	0,546	2,653										
G _{mb}	0,000	12,654	0,000	1,010	0,000	1,130	0,000	1,132	0,000	1,016	0,000	1,184
V _m	0,410	219,46										
R ²	0,681		0,524		0,628		0,634		0,588		0,583	
Adjusted R ²	0,618		0,510		0,610		0,610		0,569		0,560	
Model Sig.	0,000		0,000		0,000		0,000		0,000		0,000	
Durbin Wats.	1,368		1,738		1,441		1,248		1,689		1,814	

Starting from the initial model of all the parameters and the 0,618 of the Adjusted R², no combination managed to exceed this number. In terms of statistical significance and collinearity however, the models tested showed a much improved image. Also, regarding the Durbin Watson coefficient, we see that the values lie between 1 and 2 which is the desired.

Preferably, the number of parameters included should not be more than 4, excluding the constant. Also, it was aimed that these 4 parameters should equally include 2 bitumen characteristics and 2 volumetric. Having G_{mb} and ZSV in each category, the combination of the remaining 2 parameters was searched.

As it seems, including only G_{mb} and ZSV (Set 2), drops the Adj-R² to 0,51, with increased however significance and decreased collinearity. Adding T_{R&B} to that set (Set 3), brought the Adj-R² value to 0,61, bringing at the same time small collinearity problems. In Set 6, including all three bitumen characteristics referred to in Paragraph 3.1.4, expectedly leads to big collinearity problems, with high statistical significance though.

After doing several combinations, the conclusion is to follow Set 3, which includes T_{R&B}, ZSV, and G_{mb}. Despite the slightly increased collinearity levels, the statistical significance of the parameters, in combination with the higher model accuracy achieved, comparing to the other combinations, leads us to compromise with some collinearity.

Following the same procedure, assumptions and checks, the optimum combination of parameters aimed at predicting the dry strength of the specimens was investigated. A representative overview of the combinations tested can be seen in Table 5.30. Similar regressors were found to have a significant

influence in the performance prediction, namely G_{mb} and ZSV and $T_{R\&B}$ (Set 6). An additional parameter that showed a statistical significance and also led to an increase in the Adj-R2, was V_{bit} (Set 2). The problem in this set is again a small collinearity between ZSV and TR&B, but again for the sake of accuracy and statistical, but also physical, significance, Set 2 is chosen as the predicting relation for the dry strength. Physical significance refers to the amount of bitumen in a mixture which is known to have an important effect in moisture sensitivity, as discussed in Paragraph 3.1.4.

Table 5.30 Representative combinations of predictors for the dry subset

Set	Dry Subset											
	1		2		3		4		5		6	
Pred.	Sig.	VIF	Sig.	VIF	Sig.	VIF	Sig.	VIF	Sig.	VIF	Sig.	VIF
(const.)	0,100	-	0,000	-	0,000	-	0,000	-	0,000	-	0,000	-
$T_{R\&B}$	0,043	883,70	0,001	35,964							0,002	35,961
Pen	0,073	93,74										
logA	0,202	306,20										
G^*	-	-										
δ	0,242	38,550										
ZSV	0,003	85,600	0,005	36,630	0,009	1,229			0,000	1,019	0,016	36,367
V_{bit}	0,012	6,107	0,004	1,209	0,007	1,209	0,024	1,000				
VMA	-	-										
VFA	0,324	158,56										
Target G_{mb}	0,099	2,550										
G_{mb}	0,035	11,564	0,000	1,104	0,000	1,019	0,000	1,000	0,000	1,019	0,000	1,104
V_m	0,372	183,28										
R^2	0,663		0,593		0,515		0,280		0,456		0,533	
Adjusted R^2	0,597		0,567		0,492		0,260		0,440		0,511	
Model Sig.	0,000		0,000		0,000		0,000		0,000		0,000	
Durbin Wats.	1,644		1,404		1,228		1,038		1,038		1,149	

In order to have two comparable relations between the dry and wet strength, instead of Set 6 from the wet subset, it is preferable to follow Set 4 which additionally includes V_{bit} . Even though, the statistical significance of V_{bit} in that relation is not high, for the sake of equal comparison and physical explanation, this compromise can be made. The coefficients of the regressors for the two sets concluded, along with the predicting relations generated, are seen below.

Table 5.31 Regressor multipliers

	Dry	Wet
(Constant)	-27,541	-46,818
$T_{R\&B}$	0,1329	0,2242
ZSV	-0,000002926	-0,000005129
G_{mb}	0,0079	0,0145
V_{bit}	0,8716	0,3975

$$ITS_{dry,pred} = -27,541 + 0,1329 \cdot T_{R\&B} - 2,926 \cdot 10^{-6} \cdot ZSV + 0,0079 \cdot G_{mb} + 0,871 \cdot V_{bit} \quad [MPa] \quad \text{Eq. (5.4)}$$

$$ITS_{wet,pred} = -46,818 + 0,2242 \cdot T_{R\&B} - 5,129 \cdot 10^{-6} \cdot ZSV + 0,0145 \cdot G_{mb} + 0,397 \cdot V_{bit} \quad [MPa] \quad \text{Eq. (5.5)}$$

5.4.5 ITSR prediction

As described in the introduction of this chapter, the predicted ITSR is going to be calculated by the division of wet and dry predicted ITS (Eq. 5.3). In order to assess the quality of this prediction, each specimen's predicted ITS, from all works, phases and conditions was calculated. The predicted value of each work and phase ITSR was then computed and compared to the measured one. The quality of this prediction was then assessed on the basis of R^2 and Adjusted- R^2 (Table 5.32).

Table 5.32 ITSR model's quality assessment parameters

R	R²	Adjusted R²
0,781	0,609	0,588

Being at levels similar to the individual wet and dry relations, the ITSR prediction relation can be considered of medium quality. This can be mostly attributed to the absence of strong linear relationships between strength and composition parameters, and the participation of many different labs and researchers in obtaining the dataset.

5.4.6 Risks

An important remark at this point regards the risks that should be taken into account when conducting such a regression research. Even though regression analysis is a powerful tool for a research, the fact that it is based on limited dataset, even though it might be big and diverse, can lead to misinterpretations and miscalculations. Having a relation of good predicting quality, will return with high certainty its prediction if the independent values lie within the limits of the ones used to obtain this relation. If extreme independent values are used, their effect was not known during the regression analysis, and even though the chances of them leading to a representative results are good, there are still chances that the behavior of the dependent value radically changes after a certain point. This poses the risk of extending the application of the relation to cases that it does not represent, thus overestimating its capabilities.

In addition, the risk of including a parameter that shows a good statistical relation with the dependent variable, but does not strongly relate from a physical point of view, always exists, and should carefully be considered. This raises the attention to the researchers using the relations obtained, which only give a good indication of what to expect, and not what to confidently consider as being the case.

5.4.7 Example of application

A brief example of how these relations can be used in application is presented in this paragraph. Having the 70% ITSR requirement as the requirement, a series of different composition combinations will be made to show the way the relations could be used in the preliminary design of a mixture.

Starting with some representative values for each of the 4 properties chosen, their effect on the mixture's performance on moisture damage will be seen by adjusting them (Table 5.33). With a 'strength' of 58,8%, the mixture designed will perform according to the rough ITSR value calculated.

Table 5.33 Various design parameters and the predicted performance they lead to

Parameter	Unit	Combination							
		1	2	3	4	5	6	7	8
T_{R&B}	°C	58	58	60	60	60	60	62	62
ZSV	Pa.s	70000	70000	70000	70000	200000	200000	200000	200000
G_{mb}	kg/m ³	2360	2400	2400	2400	2400	2340	2340	2340
V_{bit}	%	4,5	4,5	4,5	4	4	4	4	5
ITS Wet	MPa	1,88	2,46	2,90	2,71	2,04	1,17	1,62	2,01
ITS Dry	MPa	2,58	2,90	3,16	2,73	2,35	1,87	2,14	3,01
ITSR	%	73%	85%	92%	99%	87%	62%	76%	67%

6 PERMANENT DEFORMATION ANALYSIS

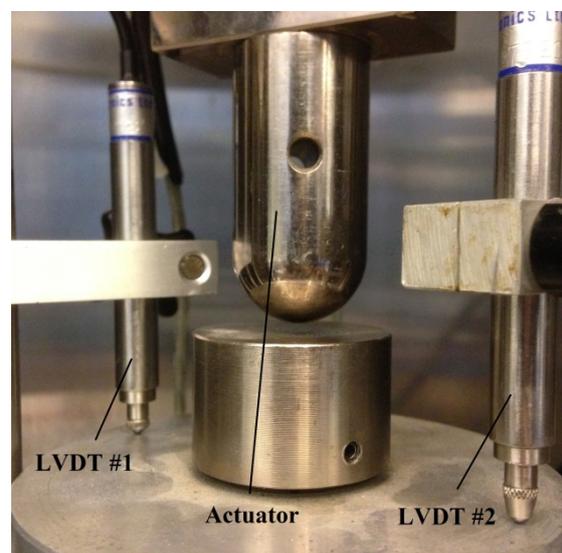
The results of the lab phase and the analysis of the permanent deformation related questions, is the subject of this chapter. Initially, all the individual results and curves are presented, with no further comment or explanation. Each individual research question is then answered in a separate paragraph, referring to the corresponding results or curves.

6.1 Test results

6.1.1 Creep curves

The output of the Triaxial Cyclic Compression Test is the creep curve, which is the accumulated permanent deformation (strain), against the number of cycles. During a test, permanent deformation is measured through 3 sensors (Picture 6.1); 2 LVDT's with an accuracy of 0,001 mm, and 1 sensor that measures the actuator's 'travel' in each pulse, with an accuracy of 0,1 mm. In the end, the creep curve can be generated by the mean value of the 2 LVDT's, by each individual LVDT's measurement, or by the actuator's strain.

Initially, the mean value of the 2 sensors was chosen. After some tests however, it was noticed that the two individual measurements did not correspond accurately, with some cases being extremely out of range. This implied that there was a calibration error in the sensors and that some adjustment was necessary in order to bring the two measurements at the same levels. This required a considerable amount of time and it was decided that for the sake of time economy and consistency, all



Picture 6.1 Measuring and loading equipment

the curves would be generated from the actuator's measurement. Even though there was clearly some 'noise' in each measurement, in the end the curve fit would accurately follow the mean value of that noise, which is the actual creep curve.

In this way, after the end of each test, each corresponding creep curve was reproduced in Excel. Figure 6.1, Figure 6.2, Figure 6.3 and Figure 6.4, represent all the curves generated for Testing Stage 1, divided respectively by test protocol. Protocols I and II consist of 4 curves for each phase of production, whereas Protocols III and IV include 5 curves for each phase. The overview of the different stages, phases and protocols is presented again in Table 6.1. The reason for the extra specimens for these protocols, was that during the planned tests, some specimens reached excessively high deformations, in both phases. Therefore, in order to check whether this was incidental, or indeed followed a trend, one extra specimen was tested for each phase and protocol, hence the 10 curves in protocols III and IV. The actuator's noise, referred to in the previous paragraph, is the reason that the curves are not perfectly smooth. It is obvious though that in the wider view it is negligible. The different phases are distinguished by color; black refers to F1 specimens from the lab, and red to F3 from the field.

Table 6.1 Overview of triaxial's testing stage 1

	Phase	Temperature °C	Maximum Stress kPa	Friction Reduction Method	Number of tests
Stage 1	F1	40	450	Plastic + Soap	4
		40	750		4
		50	450		4
		50	750		4
	F3	40	450	Plastic + Soap	4
		40	750		4
		50	450		4
		50	750		4

6.1.1.1 Stage 1

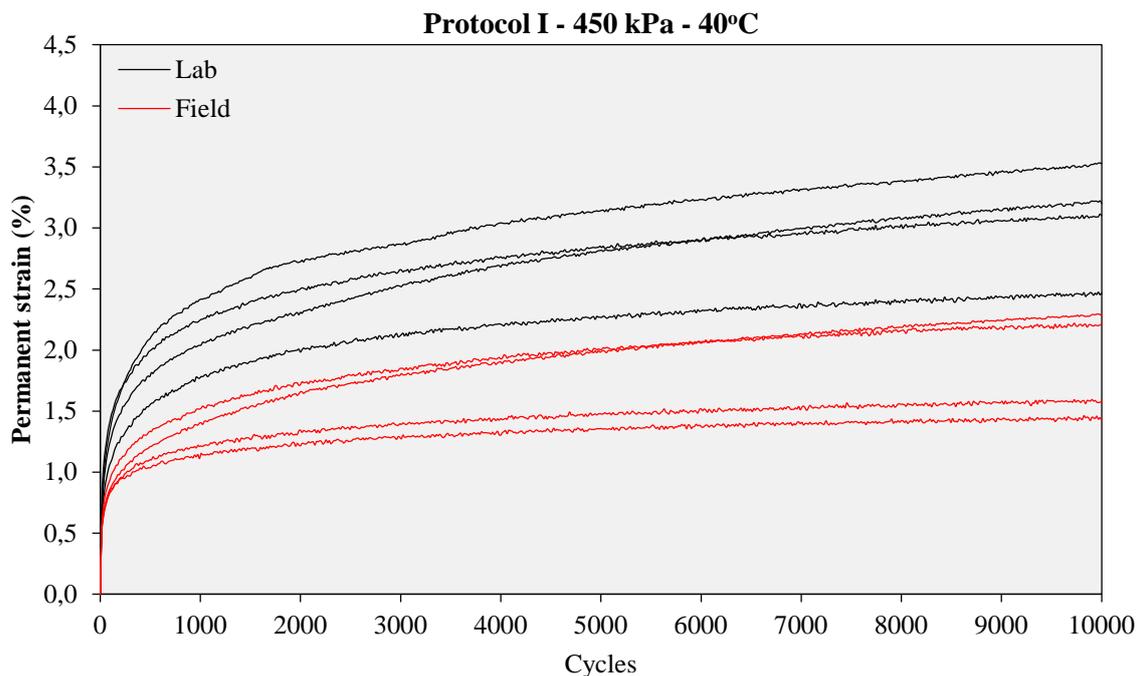


Figure 6.1 Creep curves for Protocol I specimens

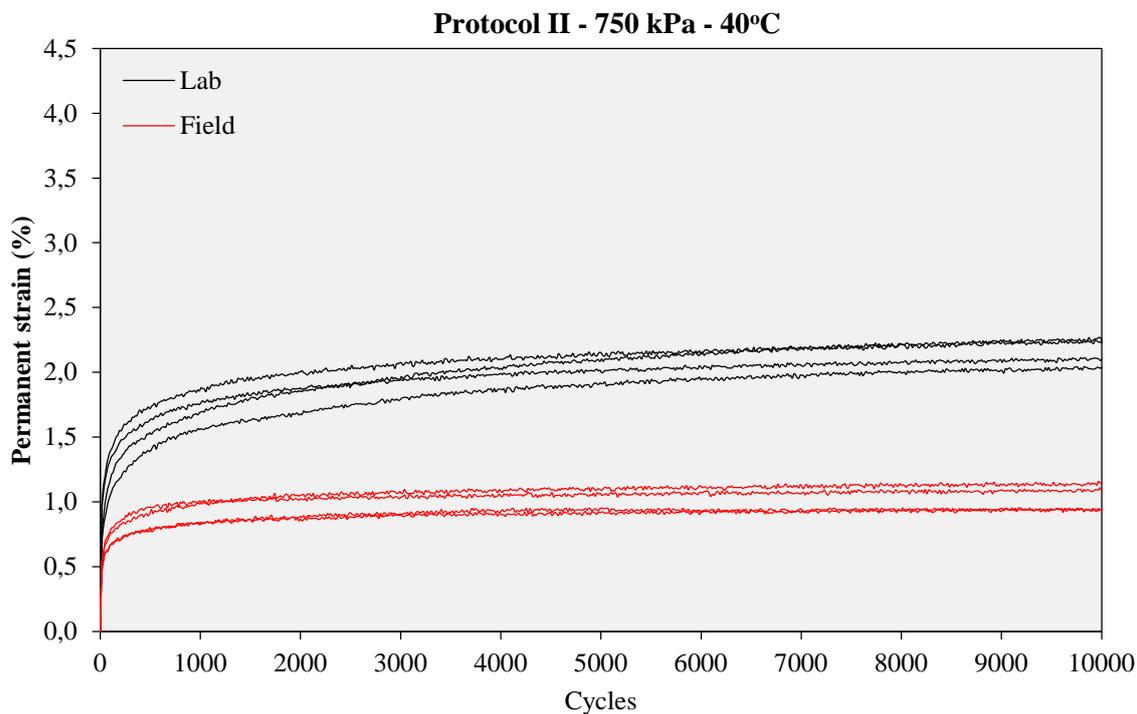


Figure 6.2 Creep curves for Protocol II specimens

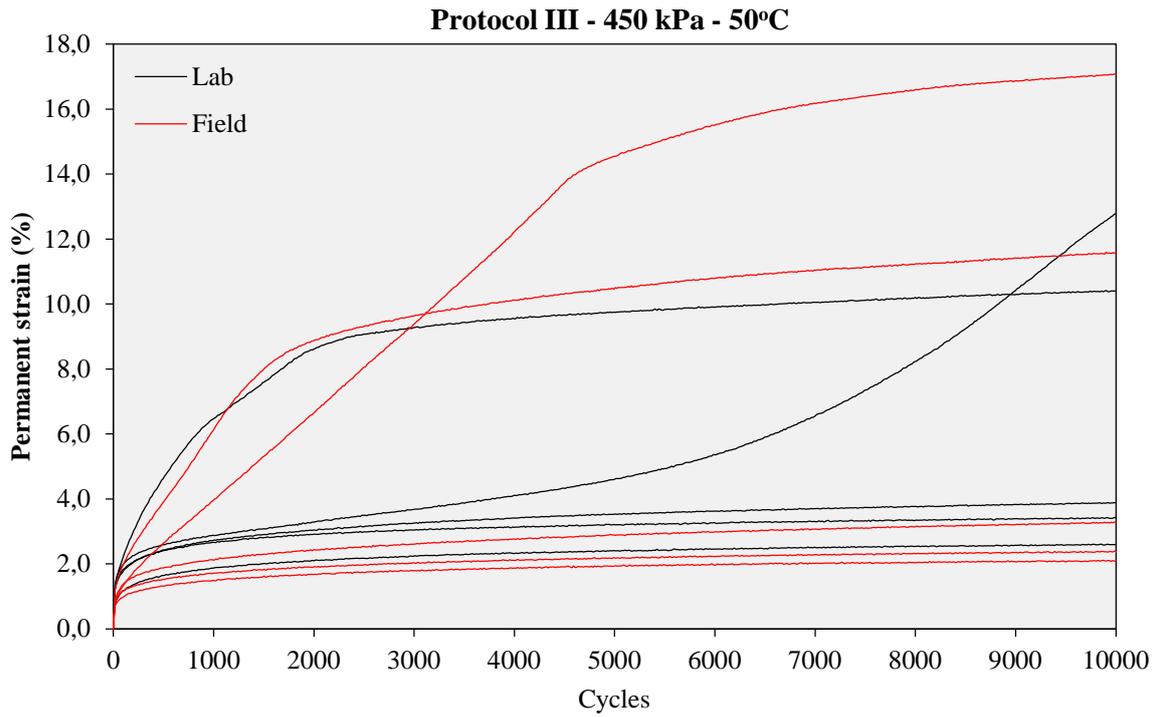


Figure 6.3 Creep curves for Protocol III specimens

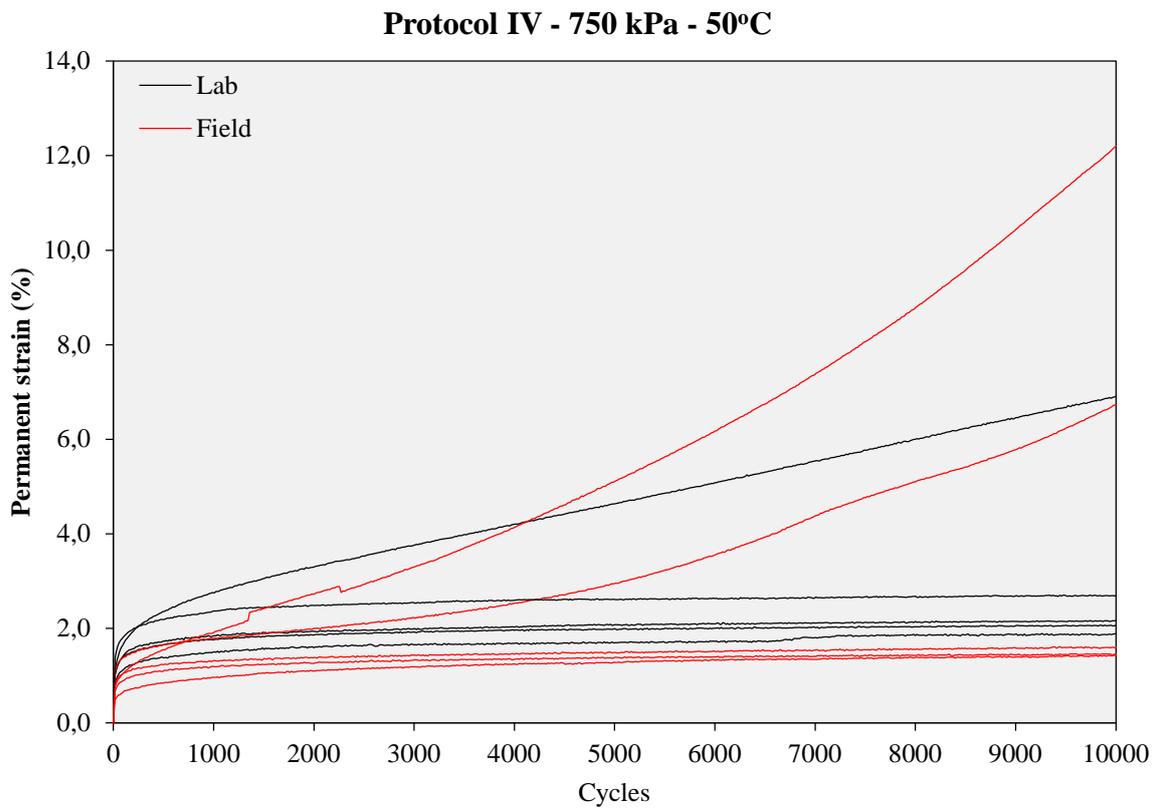


Figure 6.4 Creep curves for Protocol IV specimens

To facilitate the comparison between the all the protocols, an overview is presented in Figure 6.5. The permanent deformation axis in protocols III and IV was limited to 4,5%, so as to comply with the other protocols and make the comparison in a uniform scale.

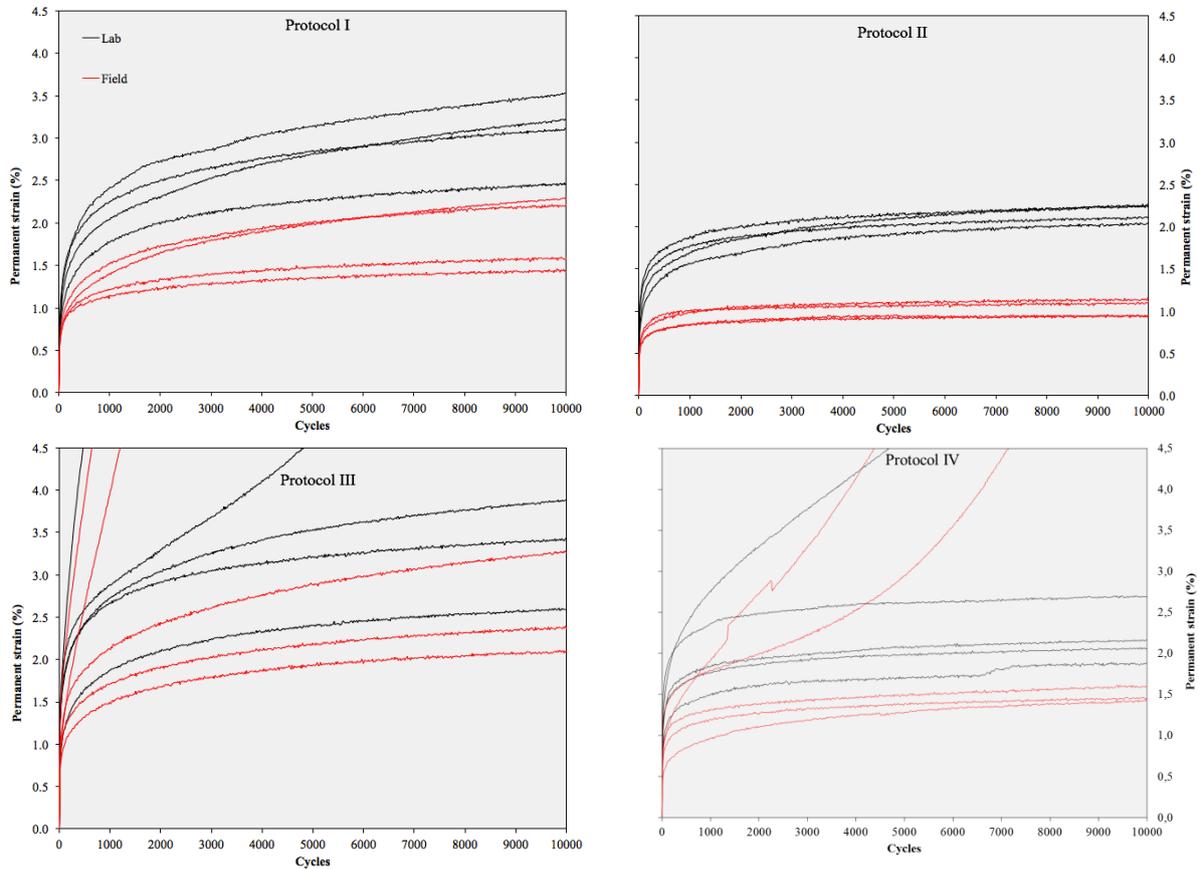


Figure 6.5 Creep curves comparative overview

6.1.1.2 Stage 2

Testing Stage 2 referring to the friction reduction methods is seen in Figure 6.6 and Figure 6.7. The first figure gives the concentrated curves from all 4 different methods tested, whereas the second gives the specimens which were tested using the same Latex membrane twice and thrice. This is done in a separate graph due to the range of values they reached which is not comparable to the rest. An overview of the number of tests in each friction reduction methods is presented in Table 6.2.

Table 6.2 Friction reduction methods testing overview

	Phase	Temperature °C	Maximum Stress kPa	Friction Reduction Method	Number of tests
Stage 2	F2	40	450	No reduction	4
				Plastic + Soap	4
				Teflon	4
				Latex	6

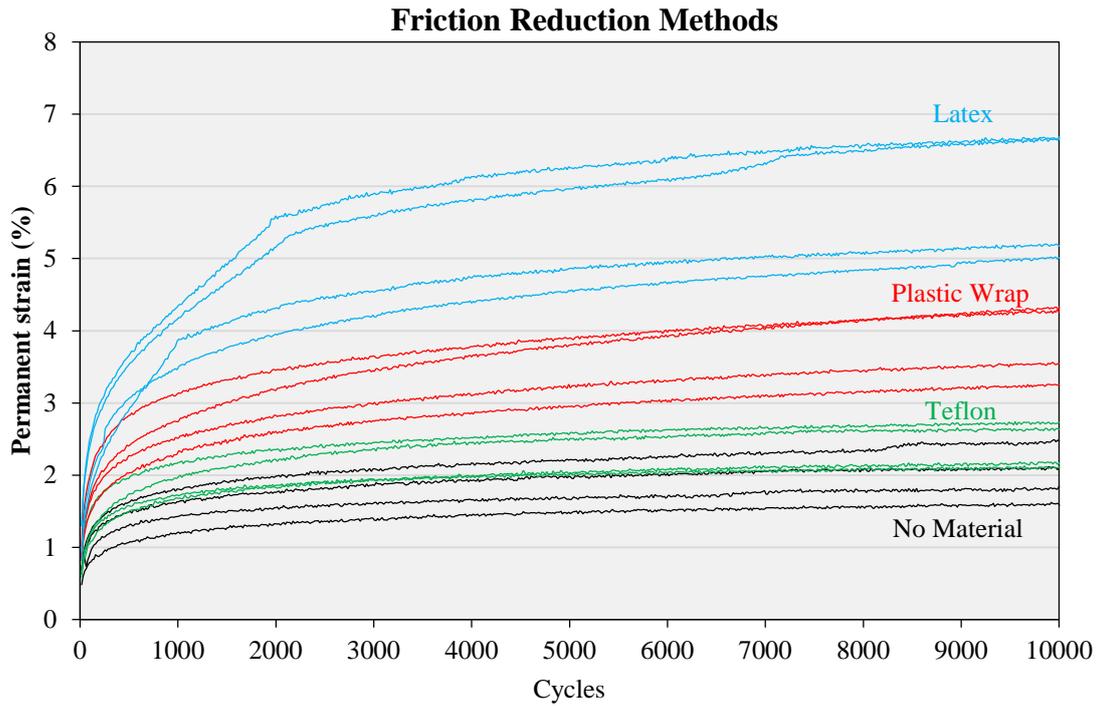


Figure 6.6 Triaxial test curves from friction reduction methods testing stage

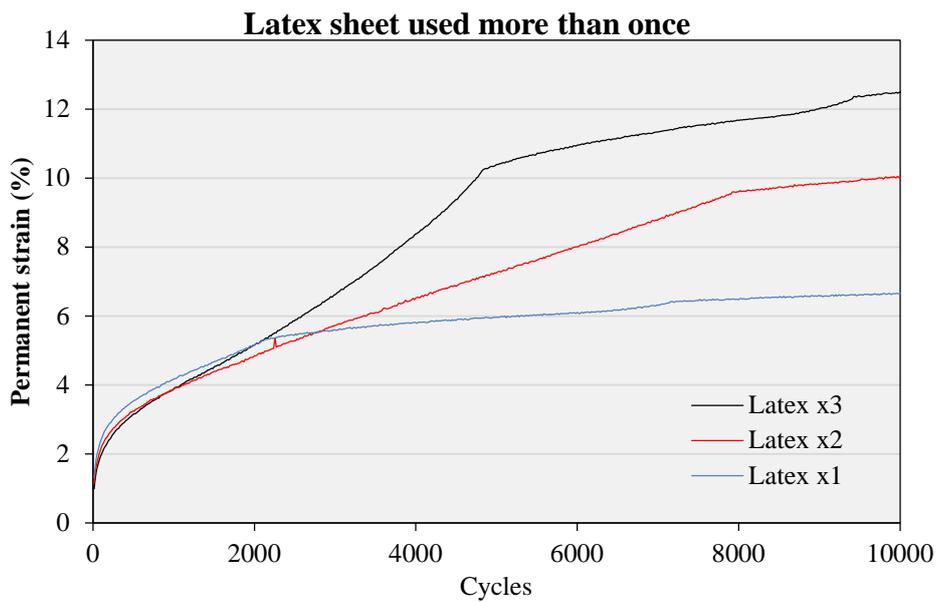


Figure 6.7 Triaxial test curves from specimens tested with Latex membranes that were used more than once after the first test

6.1.2 Linear and logarithmic curve fit

The next step after the generation of the creep curves, is to represent them in a quantitative way, to be able to analyze and compare using numbers. This is done in the two ways defined in the standard (a linear fit and a logarithmic one), described in Paragraph 4.2.2.1. Both relations were fitted using Excel's trend line function.

The general practice followed in the Netherlands for the linear fit, regarding the interval to which the curve is fit, is between 4000 and 10000 cycles, hence this was also the case in this thesis. In the cases where the curve shows tertiary response (i.e. increasing deformations after the initial decrease), and the tertiary stage is reached (stage 3), the NEN standard defined the determination of f_c , at the inflection point.

The logarithmic fit was done in a similar manner. In the cases where only stage 2 was reached, the whole curve was taken into account for the fit. When stage 3 was reached, the fit was limited to the curve up to the inflection point, neglecting in this way the tertiary part of the curve. Examples representing the way the curve fitting was carried out, are seen in Figure 6.8. The first two graphs depict the creep curve of specimen number 1228, where only stage 2 was reached. The second two graphs show the creep curve of specimen number 50, where the tertiary was reached within then 10000 cycles.

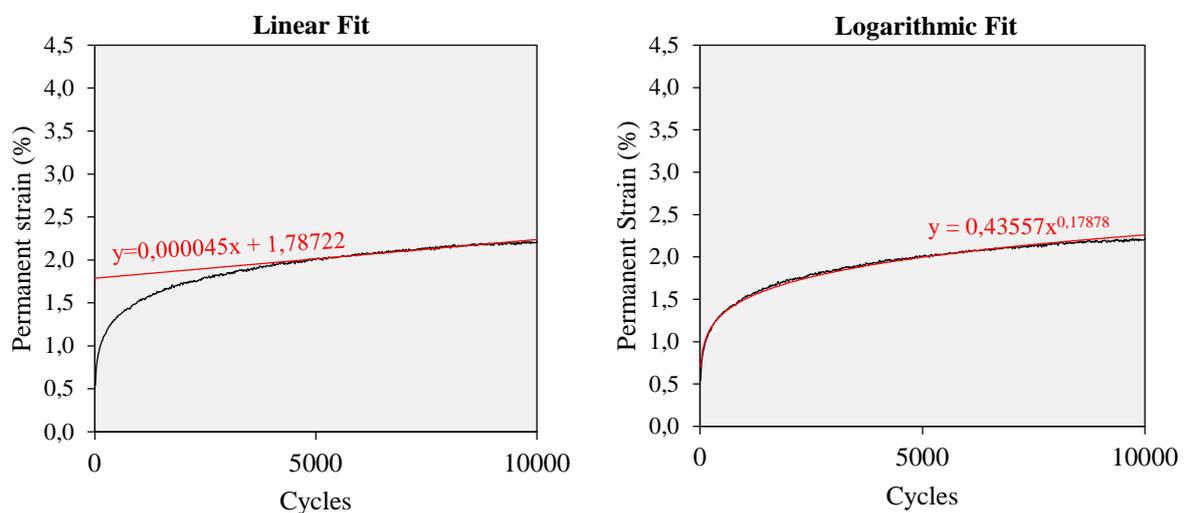


Figure 6.8 Representative linear fit without stage 3 (left), and logarithmic fit (right), from Core #1228

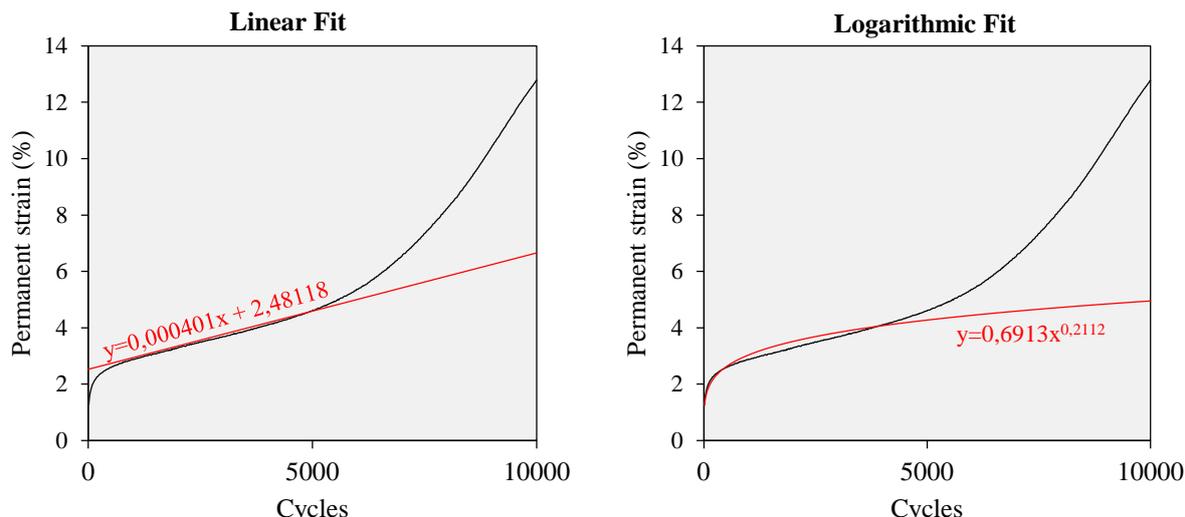


Figure 6.9 Representative linear fit with stage 3 (left), and logarithmic fit (right), from Core #50

Having obtained the fit lines for all the specimens tested, the parameters describing the curves were extracted, in the way described in Paragraph 4.2.2.1. The strain rates f_c and the permanent deformation after 1000 cycles ϵ_{1000} are concentrated in the tables below. Similarly with the moisture sensitivity tables, the left part of Table 6.3 refers to the lab specimens, and the right to the field. Also, Table 6.3 corresponds to the first testing stage where lab specimens are compared to field, whereas Table 6.4 to the second stage where the friction reduction methods are the subject of comparison. The red numbers indicate that the test exceeded the permanent deformation limit and the specimen is considered as failed.

Table 6.3 Concentrated Triaxial test results - Stage 1: different temperatures and stress conditions, Lab (left) and Field (right)

Core	Protocol	f_c $\mu\text{m/m/pulse}$	f_c avg	ϵ_{1000} %	ϵ_{1000} avg	Core	Protocol	f_c $\mu\text{m/m/pulse}$	f_c avg	ϵ_{1000} %	ϵ_{1000} avg
7	I	0,79		2,32		1223	I	0,64		1,41	
10	I	0,86	0,66	2,02	2,07	1226	I	0,24	0,38	1,19	1,30
12	I	0,56		2,19		1228	I	0,45		1,50	
21	I	0,41		1,73		1237	I	0,19		1,12	
33	II	0,21		1,83		1239	II	0,13		0,83	
37	II	0,19	0,26	1,73	1,69	1242	II	0,06	0,08	0,97	0,89
42	II	0,35		1,67		1253	II	0,06		0,82	
44	II	0,27		1,52		1255	II	0,08		0,96	
48	III	0,75		2,70		1260	III	0,35		1,48	
50	III	4,01		2,97		1269	III	2,34		5,88	
53	III	0,43	1,41	1,83	3,27	1274	III	0,83	1,25	2,13	3,16
55	III	0,46		2,61		1276	III	0,43		1,70	
76	III	1,39		6,26		1306	III	2,28		4,63	
58	IV	0,34		1,45		1287	IV	0,16		1,18	
60	IV	0,19		1,81		1290	IV	0,23		1,30	
71	IV	0,17	1,08	1,74	2,03	1292	IV	2,39	1,98	1,86	1,49
74	IV	4,54		2,84		1301	IV	0,29		0,97	
80	IV	0,17		2,31		1303	IV	6,81		2,16	

Table 6.4 Concentrated triaxial test results – Stage 2: different friction reduction methods

Core	Method	f_c $\mu\text{m}/\text{mm}/\text{pulse}$	f_c avg	ϵ_{1000} %	ϵ_{1000} avg
1125	No Reduction	0,24	0,34	1,17	1,48
1127		0,29		1,38	
1132		0,27		1,60	
1141		0,54		1,76	
1146	Teflon	0,32	0,29	2,12	1,83
1148		0,33		1,91	
1157		0,28		1,62	
1162		0,21		1,67	
1164	Plastic + Soap	0,63	0,81	2,26	2,63
1173		1,10		2,74	
1178		0,69		2,48	
1180		0,80		3,05	
1191	Latex	0,97	2,63	3,36	3,96
1194		0,72		3,47	
1196		0,91		4,40	
1205		1,54		4,21	
1210	Latex x2	6,35	-	4,05	-
1212	Latex x3	5,27	-	4,30	-

6.1.3 Cases with difficulty in slope definition and cases of failure

It should be noted that there were some cases where the definition of these parameters could not happen in such a clear way, due to their 'obscure' shape, that did not comply with any of the norm's cases. This happened for example in Protocol's III specimen number 1306, seen in Figure 6.10.

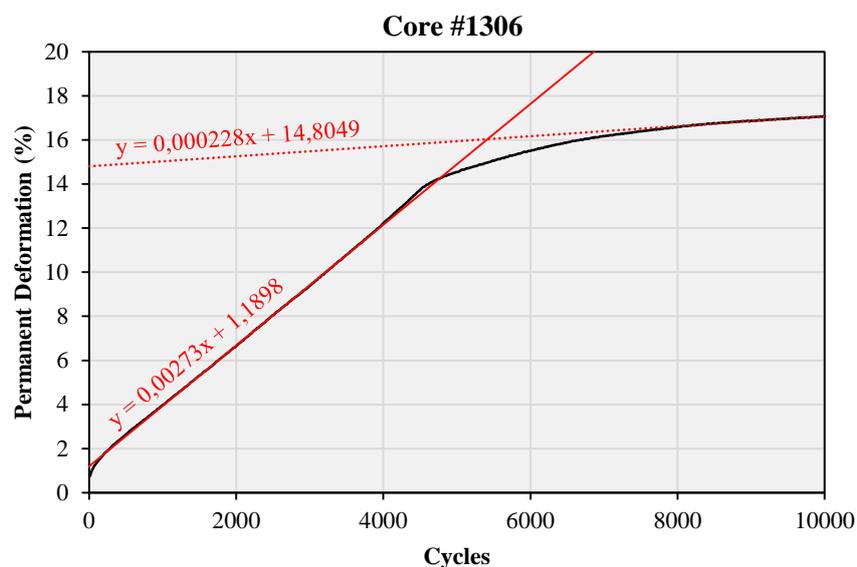


Figure 6.10 Creep curve with difficulty in defining its rutting parameters in protocol III

Up to 4000 cycles and a little bit after that, the curve has a constant slope. If the linear fit is done in this interval, the f_c value is measured at the level of 27,3 $\mu\text{m}/\text{mm}/\text{pulse}$, which is a considerably high

value. This indicates a very poor rutting resistance of the mixture under the test conditions (Protocol III, which is more severe than the standard Protocol I); an indication which seems to be realistic considering the permanent strain levels (after 6% a specimen is considered as failed) and the image of the core after the test, which had clearly failed.

If however, the slope is defined at the secondary stage of the curve, which is usually the part that follows the initial high deformation and the curve approaches a linear trend, the f_c values are naturally completely different. In this case, it is 10 times smaller, being 2,28 $\mu\text{m}/\text{mm}/\text{pulse}$. Such a value is still considerably big for a mixture, and is big enough to be considered as highly rutting susceptible. It is however, completely different from the case where somebody would define it as 27,3, and one could argue that it overestimates the mixture's resistance.

The problem lies in dividing the curve in the common creep curve stages (I, II and III). There could be an argument that the secondary stage in this case lies between 1000 and 4500 cycles, and that after that the specimen has failed and reached the tertiary stage. There could also be arguments though, that up to 4500 cycles, the specimen's deformation has not yet stabilized and that after that where the slope changes, we consider the secondary stage.

The fact that this type of curve is not common, and not described in the standard, may lead to ambiguities that leave it up to the researcher's judgement, which way to follow. For this reason, it was decided that regarding the question of performance prediction via the regression analysis, these specimens would not lead to significant contributions in the quality of the prediction, since they reached failure. This means that failure came by parameters other than the composition or bitumen characteristics, and cannot be linked to performance. These parameters are mostly the result of high temperature testing (50°C instead of 40°C), which is not the common practice, and leads to highly sensitive and unstable performance.

Regarding the questions where the comparison is qualitative and does not include statistical functions, these values will be taken into account since the conclusions will mostly be drawn from a physical explanation.

6.2 Comparison of lab to field determined properties

6.2.1 Tests output from this thesis

6.2.1.1 ANOVA

As in the case of moisture sensitivity, a statistical analysis prior to the comparison was made, to strengthen the correlations and conclusions drawn. This analysis is the two-way ANOVA, aiming to establish whether there is an interaction effect between two independent variables, in this case Test Protocol and Phase. From a physical point of view, the expectation is to find such a correlation. This process is done to ensure that there is also a statistical relation between the numbers. The assumptions and procedure are identical to Paragraph 5.3.2.1, thus only the necessary tables and comments will be presented, without elaborating further on their background.

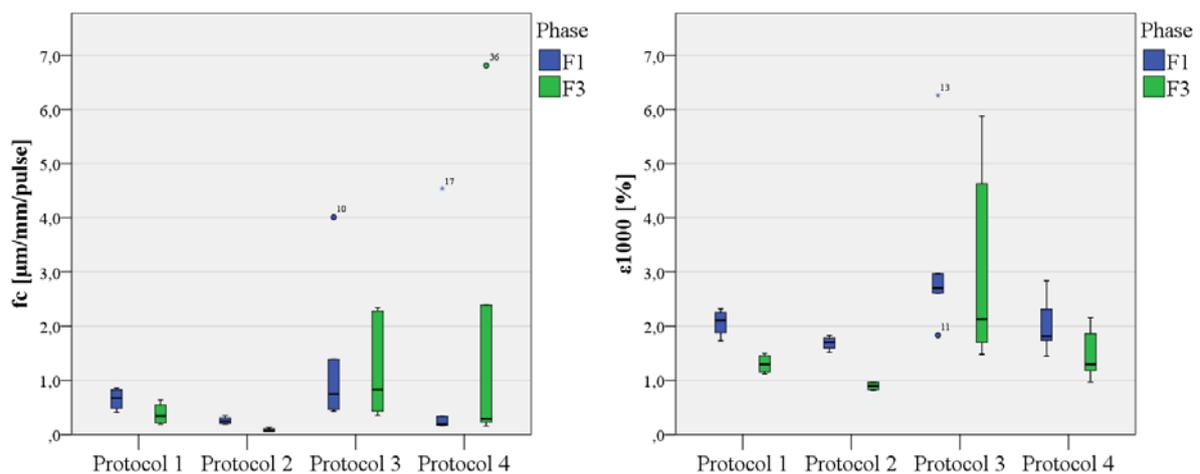


Figure 6.11 Boxplots for the detection of outliers in the case of f_c (left) and ϵ_{1000} (right)

- Detect outliers or extreme cases

As it appears from Figure 6.11, there are two cases in f_c and one case in ϵ_{1000} considered as outliers (circle markers), and one extreme case in each (star markers). One option available in this case is keeping the outliers. This requires a lot of confidence, but can be a perfectly acceptable strategy in dealing with outliers. Ideally, we are looking to find a method that evaluates whether the outlier has an appreciable effect on your analysis. One method is to run the two-way ANOVA with and without the outlier(s) included in the analysis. We can then compare the results and decide whether the two results differ sufficiently for different conclusions to be drawn from the data. If the conclusions are essentially the same (e.g., both result in a statistically significant result, confidence intervals are not appreciably

different, etc.), we might keep the outlier(s) in the data. For this reason, both analyses were done, with and without the outliers.

- Determine if the data are normally distributed

The next step is determining if the data is normally distributed. This can be checked using the Shapiro-Wilk test of normality. The results of the Shapiro-Wilk test are presented in the Tests of Normality Table 5.14. If the assumption of normality has been violated, the "Sig." value will be less than 0,05 (i.e., the test is significant at the $p < 0,05$ level). If the assumption of normality has not been violated, the "Sig." value will be greater than 0,05 (i.e., $p > 0,05$). Table 6.5 presents the test's output for f_c and ϵ_{1000} values, including the outliers, and Table 6.6 excluding them.

Table 6.5 Test of normality (including outliers)

Phase		Shapiro-Wilk			Shapiro-Wilk		
		f_c			ϵ_{1000}		
		Statistic	df	Sig.	Statistic	df	Sig.
F1	Protocol 1	,934	4	,618	,965	4	,813
	Protocol 2	,928	4	,584	,986	4	,933
	Protocol 3	,748	5	,029	,774	5	,049
	Protocol 4	,580	5	,000	,936	5	,639
F3	Protocol 1	,924	4	,560	,919	4	,530
	Protocol 2	,808	4	,117	,790	4	,085
	Protocol 3	,801	5	,082	,847	5	,185
	Protocol 4	,746	5	,027	,923	5	,549

Table 6.6 Test of normality (excluding outliers)

Phase		Shapiro-Wilk			Shapiro-Wilk		
		f_c			ϵ_{1000}		
		Statistic	df	Sig.	Statistic	df	Sig.
F1	Protocol 1	,934	4	,618	,965	4	,813
	Protocol 2	,928	4	,584	,986	4	,933
	Protocol 3	,841	4	,199	,923	3	,463
	Protocol 4	,712	4	,016	,936	5	,639
F3	Protocol 1	,924	4	,560	,919	4	,530
	Protocol 2	,808	4	,117	,790	4	,085
	Protocol 3	,801	5	,082	,847	5	,185
	Protocol 4	,676	4	,006	,923	5	,549

First of all, it can be seen that ϵ_{1000} has a bigger normality in the distribution of its values, both with the outliers, and even more excluding them. This is not a surprising fact, considering the aforementioned difficulty and vagueness in determining the f_c values, and the sensitivity of the test itself. In addition, it is obvious that the problematic distributions mostly appear at protocols 3 and 4, which were the most unstable in performance. This was also more or less expected from the big deviations that were noticed in the creep curves, with half of the specimens failing, and half not.

Despite this assumption not being entirely fulfilled, the analysis will continue, keeping it in mind when drawing the conclusions.

- Determine the homogeneity of variances

The assumption of homogeneity of variances is tested using Levene's test of equality of variances, which is seen in Table 6.7, including the outliers.

Table 6.7 Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
ϵ_{1000}	6,752	7	28	0,000
f_c	3,643	7	28	0,006

We can see that the statistical significance level is lower than 0,05 in both cases, which indicates that Levene's test is statistically significant (because $p < 0,05$). A statistically significant result, such as this, indicates that we do not have equal (population) variances and we have violated the assumption of homogeneity of variances. Even doing the test excluding the outliers, the outcome is still the same, having non-homogeneous variances.

However, as mentioned in previous cases, since the two-way ANOVA is somewhat robust in terms of homogeneity of variances too [28], we continue the analysis assuming homogeneous variances.

- Determine whether an interaction effect exists

We can determine whether we have a statistically significant interaction effect from interpreting Table 6.8 and Table 6.9. The interaction effect is represented as the product of the two independent variables in a two-way ANOVA.

Table 6.8 Tests of Between-Subjects Effects table for ϵ_{1000} values

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	23,325	7	3,332	3,111	0,015	0,437
Intercept	140,706	1	140,706	131,363	0,000	0,824
Phase	2,685	1	2,685	2,507	0,125	0,082
Protocol	20,166	3	6,722	6,276	0,002	0,402
Phase * Protocol	0,684	3	0,228	0,213	0,887	0,022
Error	29,992	28	1,071			
Total	203,870	36				
Corrected Total	53,317	35				

Table 6.9 Tests of Between-Subjects Effects table for f_c values

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	13,427	7	1,918	0,881	0,534	0,181
Intercept	27,883	1	27,883	12,812	0,001	0,314
Phase	0,045	1	0,045	0,021	0,887	0,001
Protocol	11,152	3	3,717	1,708	0,188	0,155
Phase * Protocol	2,177	3	0,726	0,333	0,801	0,034
Error	60,938	28	2,176			
Total	106,571	36				
Corrected Total	74,365	35				

Talking about both parameters, there is not statistically significant interaction between phase of production and testing protocol. Excluding the possible outliers, even though it leads to an improvement in the interaction's significance (from 0,887 to 0,473, and from 0,801 to 0,350 for ϵ_{1000} and f_c respectively), it is still considered not significant.

There are several conclusions that can be drawn out of this. First of all, the result is not surprising, due to the violation of the assumptions reducing the test's strength and giving us an idea of what to expect. Also, this expectation derived also from the image we already had about the dataset. It was clear that Protocols 3 and 4 with their high test temperature led to unstable and unpredictable behavior. This resulted in inconsistencies in the tests results, which are depicted in the non-normal distribution of their values and the inhomogeneity of their variances. Another conclusion is the better 'quality' of data in ϵ_{1000} comparing to f_c values, in terms of outliers and distribution.

The final and general conclusion coming from the Analysis of Variances, is that the strength of our conclusions to be drawn regarding the lab to field comparison, is reduced. Even though we might observe differences in the values, that from a physical point of view are explained or expected, from a statistical point of view they will not be strong enough for our conclusions to be additionally supported.

6.2.1.2 Comparison

The comparison will happen on the basis of Figure 6.11 from Paragraph 6.2.1.1.

Starting from Protocol 1, lab values seem to be higher than the field, with an average f_c of 0,66 $\mu\text{m}/\text{mm}/\text{pulse}$ comparing to 0,38. There seems to be an overlap in some of their values, however in the case of permanent deformation the two phases are more clearly separated. The average permanent deformation value in the lab is 2,07%, while in the field 1,30%. This indicates a higher rutting susceptibility for the lab produced specimens.

Protocol 2 shows a similar image to the first, with the absolute values being relatively decreased due to the increased confining stress. Lab f_c values with an average of 0,66 are clearly higher than the corresponding 0,08 from the field, with the values, also of the permanent deformation, not overlapping in their range. Again, the indication of better performing field specimens is strong. Taking into consideration that Protocols 1 and 2 performed better in the ANOVA of Paragraph 6.2.1.1, comparing to the other two protocols which were the main source of the problematic analysis, this conclusion can be more safely drawn.

The comparison expectedly becomes more difficult in Protocol 3. Due to the increased test temperature and the low confining pressure during the test, the behavior of the specimens was unstable and the capturing of their performance less precise. This is depicted in the data scatter. With extreme outlying cases both in f_c and in ϵ_{1000} , the distribution of the data is very widely spread around the mean, and as a result the two phases overlap with each other. Almost no conclusion can be drawn regarding the lab to field comparison.

In Protocol 4 where the confining pressure increased and the specimen were more stable during the test, the image is slightly clearer regarding ϵ_{1000} . Their distribution, also according to Table 6.5 Shapiro-Wilk's test of normality, is fairly normal, and makes the comparison easier. Lab values show a trend of being slightly higher than the field, with a small overlap in their values. On the other hand, f_c values show a very big scatter and their distribution of values cannot facilitate a comparison.

If we can make a rough deduction from these data, is that there is an indication of field specimens performing better in terms of rutting resistance. This means that testing a lab produced specimen in permanent deformation, underestimates the mixture's performance, and for this reason the estimation is considered safe and conservative. However, this conclusion is limited to the dataset of this thesis only, and because of the aforementioned problematic distribution and correlation of cases, it is not definitive. In the following paragraph where the entire NL-Lab database is taken into account, the possibility of deducting information about the lab-field relation is higher.

6.2.2 Tests output from entire NL-Lab database

6.2.2.1 ANOVA

Following the same procedure as in the previous analyses of variance, the interaction effect of phase and work was examined, taking into account the entire NL-Lab database from the triaxial tests. Having no outliers detected, the analysis of normality, homogeneity and interaction was carried out by SPSS. As in the other cases, the number of data points in each case is not wide enough to extract secure conclusion, and it will just give an indication of the trend.

Table 6.10 Test of Normality for each work

Phase		Shapiro-Wilk			Shapiro-Wilk		
		f_c			ε_{1000}		
		Statistic	df	Sig.	Statistic	df	Sig.
W1	F1	0,879	8	0,186	0,892	8	0,243
	F3	0,917	8	0,409	0,950	8	0,711
W2	F1	0,787	5	0,063	0,828	5	0,135
	F3	0,958	5	0,794	0,973	5	0,896
W3	F1	0,877	3	0,315	0,893	3	0,363
	F3	0,821	4	0,145	0,902	4	0,442
W4	F1	0,876	4	0,320	0,942	4	0,666
	F3	0,942	4	0,668	0,989	4	0,951
W5a	F1	0,934	4	0,618	0,965	4	0,813
	F3	0,924	4	0,560	0,919	4	0,530
W5b	F1	0,882	5	0,319	0,815	5	0,107
	F3	0,791	5	0,068	0,844	5	0,176

Starting by examining the distribution of the data among each work, we can see from the Shapiro-Wilk test in Table 6.10, that both f_c and ε_{1000} seem to have normally distributed values (i.e. Sig.>0,05). The first prerequisite is satisfied.

The next step is checking the homogeneity among the variances. This assumption is tested using Levene's test of equality of variances, which is seen in Table 6.11.

Table 6.11 Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
f_c	13,639	11	47	,000
ε_{1000}	1,888	11	47	,066

We can see that the statistical significance level of ε_{1000} is 0,066 μ which indicates that Levene's test is not statistically significant (because $p > 0,05$). A non-statistically significant result, such as this, indicates a significant evidence that we have equal (population) variances and we have not violated the assumption of homogeneity of variances. In the case of f_c however this is not satisfied, meaning we do not have homogeneity of variances. This again confirms the already known 'problem' of the f_c values and their difficult repeatability.

However, since the values are normally distributed and we have equally sized subsets, we continue with the two-way ANOVA, because it is somewhat robust to heterogeneity of variance in these circumstances.

Finally, the interaction effect of phases and works is eventually tested. This is seen in Table 6.12 and Table 6.13, in the 'Sig.' column of 'Phase*Work' row. If the value is smaller than 0,05 then a statistically significant interaction exists and the strength of our conclusions is increased.

Table 6.12 Tests of Between-Subjects Effects table for f_c

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1,353	11	0,123	12,375	0,000	0,743
Intercept	3,385	1	3,385	340,619	0,000	0,879
Phase	0,000	1	0,000	0,041	0,841	0,001
Work	1,055	5	0,211	21,240	0,000	0,693
Phase * Work	0,294	5	0,059	5,908	0,000	0,386
Error	0,467	47	0,010			
Total	5,014	59				
Corrected Total	1,820	58				

Table 6.13 Tests of Between-Subjects Effects table for ϵ_{1000}

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	15,443	11	1,404	21,014	0,000	0,831
Intercept	58,776	1	58,776	879,767	0,000	0,949
Phase	1,875	1	1,875	28,064	0,000	0,374
Work	6,511	5	1,302	19,491	0,000	0,675
Phase * Work	6,668	5	1,334	19,962	0,000	0,680
Error	3,140	47	0,067			
Total	76,026	59				
Corrected Total	18,583	58				

We can see that in both cases there is indeed a statistically significant interaction between phase and work. This means that the values obtained in each different work and phase, are significantly different from each other, and the connection between them and the physical meaning can be made more safely.

A notable fact is in the f_c check, where its relation with phase led to a non-significant statistical result (Sig = 0,841). This means that the values do not show a consistent relation among the various phases. However, when work also is taken into account simultaneously, the significance is high, exposing the big variation of f_c values between the different works.

6.2.2.2 Comparison

After the conclusion that rutting resistance is statistically dependent on the phase of production in combination with work, we can get an all-rounded view by looking at the numbers comparatively. Since work 1 tests were carried out in two different labs, and it was noticed that the results did not match, they will be treated as two separate works. 'W1a' referring to the first lab, and 'W1b' to the second. Also, since the target is to make a comparison as accurately as possible, meaning that it should happen under the same terms for all works, W5a (this thesis), where various test protocols were used, will be limited only to the standard protocol (Protocol I with 40°C and 450kPa maximum stress applied).

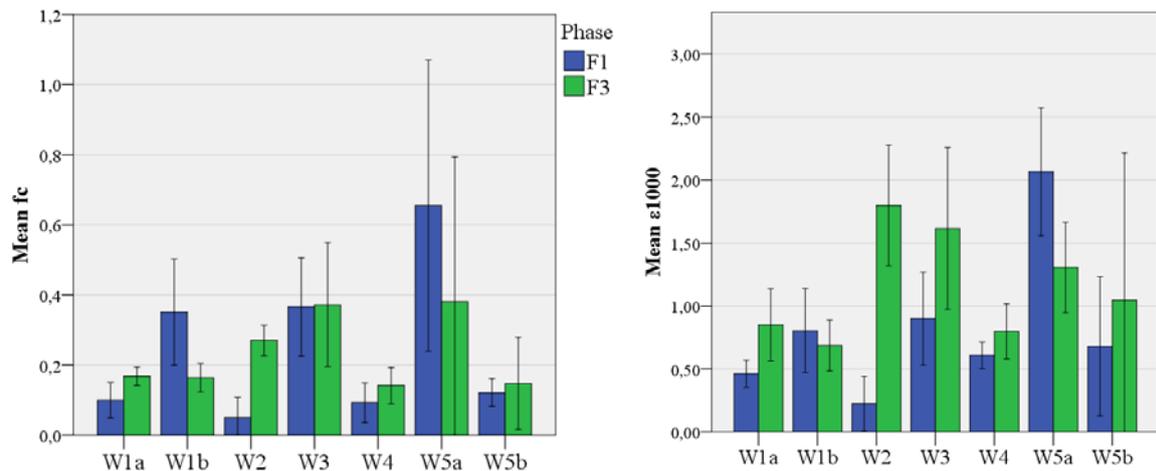


Figure 6.12 Average f_c (left) and ϵ_{1000} (right) values from every NL-Lab work separately

As it can be seen in Figure 6.12, lab and field values do not coincide in the entirety of the project. The only case where two phases match, is the f_c values at work 3. The ϵ_{1000} values though are still far from similar. In 5 out of 7 works, rutting levels in the lab are lower than in the field (W1a, W2, W3, W4 and W5b). In the rest 2 works the situation is inverted (W1b and W5a). This means that there is the possibility of a conservative estimation of rutting resistance, that works on the safe side, but also the possibility of overestimating a mixture's ability to resist rutting.

6.2.3 Relation to composition parameters

6.2.3.1 Statistical relation

Following the same principle as in the case of moisture sensitivity, the next step is to trace back and identify the reasons behind the lab and field performance correlation. This is done by investigating the composition parameters and bitumen characteristics.

Same as described in Paragraph 5.3.3.1, the first tool used for this purpose is Pearson's product-moment correlation, used to determine the strength and direction of a linear relationship between two continuous variables. The assumptions necessary for the employment of this test are listed below:

- Assumption 1: The two variables should be measured on a continuous scale (i.e., they are measured at the interval or ratio level). In our case, this assumption is satisfied for all parameters checked.
- Assumption 2: The two continuous variables should be paired (i.e., each case has two values – one for each variable). This is also satisfied, having for every specimen two paired values (f_c or ϵ_{1000} with composition parameters).

- Assumption 3: There needs to be a linear relationship between the two variables. The best way of checking this assumption is to plot a scatterplot and visually inspect the graph. The scatterplots of the f_c and ϵ_{1000} values against each parameter are presented in the set of figures in Figure 6.13. The top two rows of each phase refer to the bitumen characteristics and the bottom to the volumetric and test conditions.

The scatter of f_c and ϵ_{1000} is similar, thus the comparison with the parameters refers to both. Starting from F1, the bitumen characteristics have a mostly vertical concentration in their highs or lows, due to the limited variance in their values chosen. The distribution of the other volumetrics looks like a ‘cloud’, with VMA, VFA and V_m approaching a weak linear distribution.

In the case of F3, the linearity of the graphs is inferior. The bitumen parameters values are mostly concentrated in vertical columns, which cover all the range of f_c or ϵ_{1000} . The volumetrics are again in a ‘cloud’ pattern, showing no linearity.

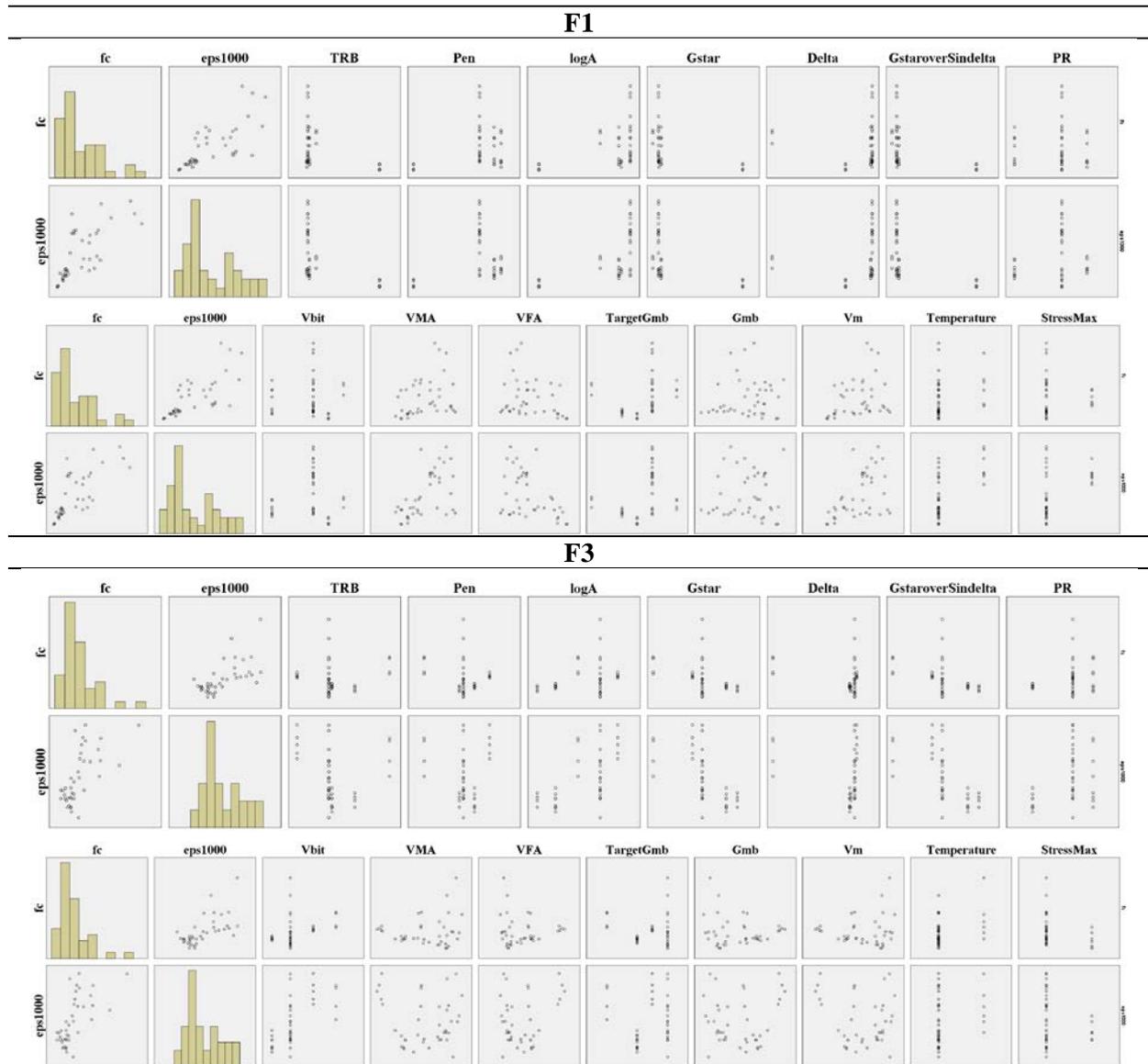


Figure 6.13 Scatterplots for the inspection of linearity between permanent deformation and composition parameters

- Assumption 4: Assumption of bivariate normality, or in other words, there should be a normal distribution in each parameter, test by the Shapiro-Wilk non parametric test. If the assumption of normality has been violated, the "Sig." value will be less than 0,05. If the assumption of normality has not been violated, the "Sig." value will be greater than 0,05. The test's output by SPSS is in Table 6.14.

Table 6.14 Test of Normality of the data

	F1			F3		
	Shapiro-Wilk			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
$T_{R\&B}$	0,449	40	0,000	0,766	40	0,000
Pen	0,660	40	0,000	0,788	40	0,000
logA	0,584	40	0,000	0,830	40	0,000
G^*	0,446	40	0,000	0,836	40	0,000
δ	0,456	40	0,000	0,426	40	0,000
$G^*/\sin\delta$	0,438	40	0,000	0,839	40	0,000
PR	0,732	40	0,000	0,736	40	0,000
V_{bit}	0,767	40	0,000	0,784	40	0,000
VMA	0,947	40	0,062	0,928	40	0,014
VFA	0,924	40	0,010	0,874	40	0,000
Target G_{mb}	0,855	40	0,000	0,779	40	0,000
G_{mb}	0,968	40	0,306	0,937	40	0,027
V_m	0,955	40	0,114	0,918	40	0,006
Temperature	0,462	40	0,000	0,428	40	0,000
σ_{max}	0,491	40	0,000	0,462	40	0,000
f_c	0,860	40	0,000	0,830	40	0,000
ϵ_{1000}	0,906	40	0,003	0,925	40	0,011

According to the test, the only values that follow a normal distribution are the voids in the mineral aggregate, density and air voids in F1. This means that the assumption of normality is also violated.

- Assumption 5: There should be no significant outliers. This assumption is satisfied since the outliers observed were not included for the reason explained in Paragraph 6.1.3.

With two very important assumptions for the Pearson's test being violated, this means that as in the case of moisture sensitivity (Paragraph 5.3.3.1), a 'distribution-free' test should be followed; the Spearman's rank-order correlation. Its assumption of monotonic relationships cannot be again covered with certainty, hence following the same process as moisture sensitivity, the Kendall's tau-b test will also be carried out for a confirmation. The Spearman's correlation coefficients along with their statistical significance are seen in Table 6.15. As it compares to Kendall's tau-b Table 6.16, the image is similar, so we can draw some conclusions out of this correlation.

Table 6.15 Spearman's correlation of f_c and ϵ_{1000} to composition parameters

Spearman's	F1				F3			
	f_c		ϵ_{1000}		f_c		ϵ_{1000}	
	r_s	Sig.	r_s	Sig.	r_s	Sig.	r_s	Sig.
f_c	1,000	-	0,841	0,000	1,000	-	0,706	0,000
ϵ_{1000}	0,841	0,000	1,000	-	0,706	0,000	1,000	-
$T_{R\&B}$	-0,200	0,216	-0,481	0,002	-0,099	0,543	-0,387	0,014
Pen	0,128	0,432	-0,012	0,940	-0,040	0,807	-0,004	0,978
logA	0,456	0,003	0,710	0,000	0,295	0,064	0,599	0,000
G^*	-0,607	0,000	-0,689	0,000	-0,539	0,000	-0,731	0,000
δ	0,377	0,017	0,653	0,000	0,072	0,657	0,395	0,012
$G^*/\sin\delta$	-0,607	0,000	-0,689	0,000	-0,539	0,000	-0,731	0,000
PR	-0,022	0,894	0,112	0,492	0,204	0,207	0,325	0,040
V_{bit}	-0,119	0,466	-0,079	0,629	0,468	0,002	0,683	0,000
VMA	-0,402	0,010	0,552	0,000	-0,001	0,994	0,029	0,858
VFA	-0,301	0,060	-0,364	0,021	0,046	0,776	0,010	0,952
Target G_{mb}	0,470	0,002	0,471	0,002	-0,079	0,628	0,142	0,381
G_{mb}	0,068	0,675	-0,016	0,924	-0,007	0,968	-0,047	0,775
V_m	0,371	0,019	0,476	0,002	-0,048	0,771	-0,008	0,960
Temperature	0,299	0,061	0,542	0,000	0,334	0,035	0,270	0,092
σ_{max}	0,127	0,434	0,433	0,005	-0,339	0,032	-0,111	0,495

First of all, regardless of the composition parameters, it is obvious that f_c and ϵ_{1000} are stronger correlated in F1 than F3 (0,841 instead of 0,706). Ideally these two parameters would be equally related, since they both describe the rutting resistance of a mixture. However due to the computing inaccuracies of converging the curves into these parameters, there is this diversion from being perfectly related. In addition to that, it is generally observed that ϵ_{1000} values show higher correlation coefficients to the composition parameters, comparing to f_c .

In general terms, we can say that there is a slightly better correlation between volumetrics and performance of F1. Bitumen characteristics though, show similarly strong correlations with the performance, regardless the phase of production.

Table 6.16 Kendall's tau-b correlation of f_c and ϵ_{1000} to composition parameters

Kendall's tau-b	F1				F3			
	f_c		ϵ_{1000}		f_c		ϵ_{1000}	
	τ_b	Sig.	τ_b	Sig.	τ_b	Sig.	τ_b	Sig.
f_c	1,000	-	0,652	0,000	1,000	-	0,478	0,000
ϵ_{1000}	0,652	0,000	1,000	-	0,478	0,000	1,000	-
$T_{R\&B}$	-0,136	0,268	-0,377	0,002	-0,088	0,470	-0,314	0,010
Pen	0,084	0,496	-0,036	0,772	-0,006	0,960	-0,020	0,871
logA	0,358	0,004	0,562	0,000	0,234	0,055	0,460	0,000
G^*	-0,487	0,000	-0,543	0,000	-0,422	0,001	-0,575	0,000
δ	0,311	0,011	0,516	0,000	0,033	0,784	0,333	0,006
$G^*/\sin\delta$	-0,487	0,000	-0,543	0,000	-0,422	0,001	-0,575	0,000
PR	-0,017	0,892	0,071	0,578	0,178	0,161	0,258	0,042
V_{bit}	-0,097	0,440	-0,045	0,720	0,359	0,004	0,550	0,000
VMA	0,288	0,010	0,413	0,000	-0,022	0,843	0,047	0,675
VFA	-0,205	0,064	-0,276	0,012	0,055	0,616	-0,021	0,852
Target G_{mb}	0,344	0,005	0,282	0,021	-0,053	0,669	0,099	0,428
G_{mb}	0,058	0,600	-0,022	0,843	0,017	0,880	-0,057	0,608
V_m	0,254	0,022	0,348	0,002	-0,054	0,624	0,022	0,843
Temperature	0,248	0,062	0,448	0,001	0,276	0,037	0,224	0,092
σ_{max}	0,106	0,427	0,359	0,007	-0,281	0,034	-0,092	0,488

In conclusion, comparing F1 to F3 through a statistical correlation of their composition parameters did not lead to striking results, as in the case of moisture sensitivity, and no significant conclusions can be drawn out of this.

6.2.3.2 Comparison

Trying to connect the known performance indicators to the parameters of the mixture is highly complex process. Asphalt concrete is the product of many parts, that all of them as a whole define its behavior when it is eventually compacted into a pavement layer. Even though there is a direct relation between some parameters, one factor alone cannot fully explain performance. It is their combination that defines it. Due to this complexity, there is the risk of reaching false conclusions and observations that might seem to hold, while other additional factors also lie behind it and are neglected. For this reason, the observations made are not definitive and care should be taken in their interpretation. They consist of indications and trends that can possibly be confirmed if more elaborate research is done on them. This task is even more difficult considering the bad relation between composition and performance seen in Figure 6.13 of the previous paragraph.

A useful tool to facilitate the lab to field comparison is to investigate the accuracy with which the target density was reached in each case. This is done by dividing the density achieved (G_{mb}) by the target density of every work. The closer the number obtained is to +1, the bigger the accuracy in reaching the target density. Values greater than 1 suggest an overcompacted specimen, while smaller than 1 suggest poor compaction. How each work and phase correspond to this accuracy is seen in Figure 6.14, where the mean precision values are plotted. The red horizontal line represents the perfect equality between target and achieved bulk density.

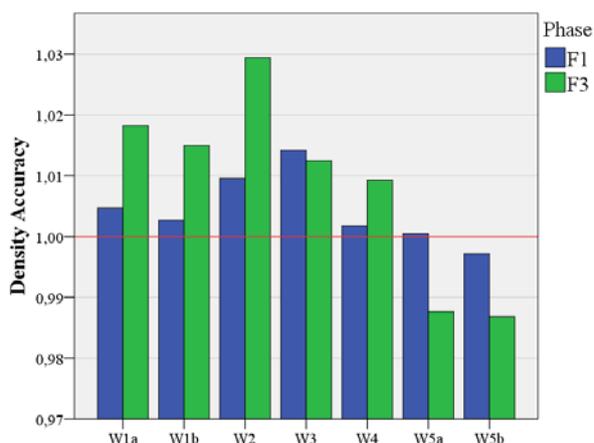


Figure 6.14 Accuracy in achieving the target density

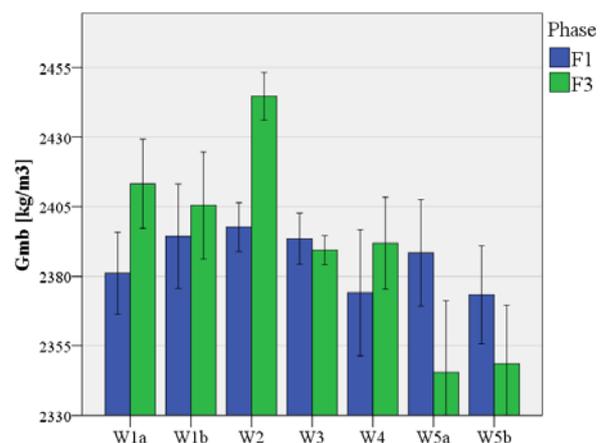


Figure 6.15 Average densities

First of all, we can certainly say that lab specimens show better accuracy in achieving their target density. The majority of the works, besides 2 and 3, lie very close to the red line. On the other hand, field specimens tend not to achieve their target density very accurately. Up to work 4, they are overcompacted, while work 5 is poorly compacted. Talking in density terms, the conclusion is the same, since the densities follow the accuracy's relations (Figure 6.15).

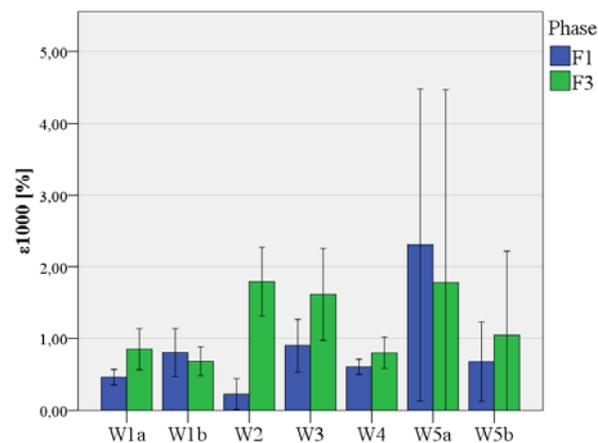


Figure 6.16 Permanent strain values from all works

Trying to match this observation with performance however, does not lead to any connection with performance. Looking at the permanent deformation values from Figure 6.16, there are cases where higher density accuracy leads to better performance (W1a, W4 and W5b), and cases where it leads to poorer performance (W1b and W5a). As explained at the beginning of this paragraph, performance is not the result of just one parameter, but the combination of many, and this is the reason we cannot trace any relation here.

6.3 Performance prediction

6.3.1 Introduction

Based on the composition parameters collected for every single specimen tested, and its measured performance in permanent deformation, a function describing the relation between these two will be obtained. As described in paragraph 2.4.3, this function aims to help in the preliminary design of a mixture, when its composition has to be roughly determined in order to satisfy the certain requirements, and the identification of the most important parameters that play a role in that.

The tool used for this purpose is the Multiple Linear Regression Analysis. A detailed description of the checks and assumptions required for this analysis is included in paragraph [ref] from the moisture sensitivity description. For space economy, not all of these steps will be elaborately presented in this chapter. The remarks and observations that were made before and during the analysis, and are of specific interest to the permanent deformation case, will be fully explained.

Since permanent deformation is described both by the strain slope (f_c) and by the permanent strain after 1000 cycles (ε_{1000}), the target of the regression analysis will be to obtain two predicting relations, one for each property.

$$f_{c,predicted} = \alpha_o + \alpha_1 \cdot x_1 + \alpha_2 \cdot x_2 + \dots + \alpha_i \cdot x_i \quad [\mu m / mm / pulse] \quad \text{Eq. (6.1)}$$

$$\varepsilon_{1000,predicted} = \beta_o + \beta_1 \cdot y_1 + \beta_2 \cdot y_2 + \dots + \beta_i \cdot y_i \quad [\%] \quad \text{Eq. (6.2)}$$

6.3.2 Data included

During the course of this thesis, 4 different testing protocols were followed, both for F1 and F3, distinguished by the combination of test temperature and maximum stress applied. The entire NL-Lab database, was obtained on the basis of only one protocol, the one prescribed for a base/binder layer (Protocol 1). All of these cases are going to be included, since this variation in test conditions was intentionally chosen to serve the purpose of bigger variety in the parameters.

As far as F2 specimens from this thesis are concerned (i.e. friction reduction methods research), only the 4 specimens tested with the same material as F1 and F3 are going to be included. The reason for this is the conclusion of paragraph [ref], that the friction reduction method chosen radically changes the test's output. In the previous NL-Lab works, not all the friction reduction methods were reported, making it impossible to use it as a categorical variable in explaining the variance in the performance. Due to some of them being reported however, it is clear that not a consistent method was used

between the various works, meaning that there will be a considerable variance in the data, that will not be able to be explained by the predicting relation.

Finally it should be noted that the tests that led to the specimen's failure, are not included in the regression analysis, for the reason described in paragraph 6.1.3.

6.3.3 Analysis and checks

The procedure described follows the step-by-step multiple regression analysis by 'Laerd Statistics (2015). *Statistical tutorials and software guides*' [21].

For the multiple linear regression to be considered appropriate for use, there are certain assumptions that need to be checked. These assumptions, described in detail in paragraph 3.3.1 refer both to the nature of the dataset, and their numerical behavior in terms of distribution, linearity and collinearity. The first two of them are checked before the regression takes place, and are fulfilled for all the cases of this thesis. One refers to the dependent variable (i.e. f_c and ϵ_{1000}), and that it should be measured at the continuous level, and the other to the independent variables (i.e. composition parameters), also named regressors or predictors, and that they should be measured either at the continuous or at the nominal level. Both of them are satisfied.

The remaining six assumptions can only be checked after the analysis is run. The initial run, that will be used as the starting point of the analysis, will include in the relation all the available composition parameters, and will use f_c as its target; the ϵ_{1000} analysis will follow next. Based on the output and the checks, some parameters will be excluded and various combinations will be tested, to find the one that both optimally satisfies the regression requirements, and has a physical explanation. The checks of this first run and the entire regression procedure, will be described in detail, referring to just the wet subset of specimens. After that, the results will be comparatively presented as an overview. The procedure and the necessary checks will not be presented again, but will be included in the Appendix instead.

- Independence of observations

The independence of observations is statistically tested using the Durbin-Watson test. The Durbin-Watson statistic is generated by the SPSS Model Summary Table 6.17.

Table 6.17 Model Summary Table

R	R²	Adjusted R²	Std. Error of the Estimate	Durbin-Watson
0,707	0,500	0,446	0,14878	0,975

The Durbin-Watson statistic for this analysis is 0,975. The Durbin-Watson statistic can range from 0 to 4, but we are looking for a value of approximately 2 to indicate that there is no correlation between

residuals. Our value is not very close to 2, however since there is no reason for our observations to be related, we can interpret the Durbin-Watson test as if there is independence of errors (residuals).

- Linearity

An assumption of multiple regression is that: (a) the independent variables collectively are linearly related to the dependent variable; and (b) each independent variable is linearly related to the dependent variable.

The first is checked by plotting a scatterplot of the studentized residuals against the (unstandardized) predicted values (Figure 6.17). If the residuals form a horizontal band, the relationship between our dependent variable and independent variables is likely to be linear. However, as it can be seen, a funnel is created, where residual values are increasing with the increasing predicted values. This indicates a collectively non-linear relationship, a fact which was expected from the low correlation values found in Table 6.17 Model Summary Table.

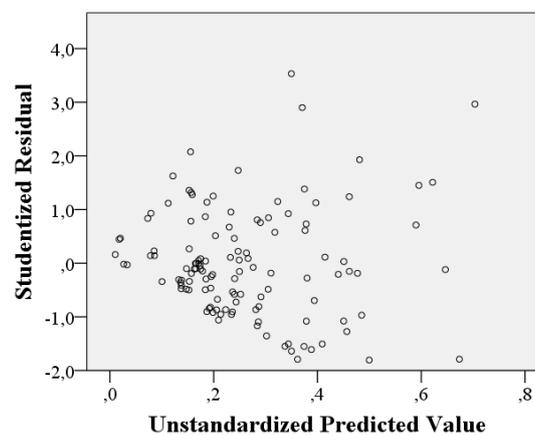


Figure 6.17 Studentized residuals against the predicted values

The linearity of each individual independent parameter with ITS was checked in paragraph 6.2.3.1, Figure 6.13, with the partial scatterplots, during the comparison between lab and field produced specimens. Almost no linearity was noticed between all the parameters and the performance. Only VMA, VFA and V_m showed a poor trend. So, in general terms, neither a collective nor an independent linear relation exists between composition and performance. This will have the consequence of a reduced predicting quality, which was already obvious from Table 6.17.

- Homoscedasticity

The assumption of homoscedasticity is that the residuals are equal for all values of the predicted dependent variable. To check for heteroscedasticity, we can use the plot created to check linearity in the previous section, namely plotting the studentized residuals against the unstandardized predicted values Figure 6.17.

If there is homoscedasticity, the spread of the residuals will not increase or decrease as we move across the predicted values (i.e., the points of the plot above will exhibit no pattern and will be approximately constantly spread). However, if the residuals are not evenly spread, but differ in height (e.g., a funnel shape), we do not have homoscedasticity. Instead, we have what is called heteroscedasticity (i.e., we have violated the assumption of homogeneity of variance).

The residuals in our case show a strong funnel shape. On this basis, it would appear that the assumption of homoscedasticity has been violated.

- Multicollinearity

Multicollinearity occurs when we have two or more independent variables that are highly correlated with each other. This multicollinearity can be recognized by the Tolerance/VIF values generated by SPSS for each regressor separately (Table 6.18). If VIF is greater than 10, we might have a collinearity problem.

Table 6.18 Coefficients statistical parameters

	Unstandardized		Sig.	Collinearity Statistics	
	B	Std. Error		Tolerance	VIF
(Constant)	-3,709	7,399	0,617		
T _{R&B}	0,048	0,073	0,515	0,002	560,600
Pen	0,005	0,031	0,881	0,014	71,935
logA	2,854	6,154	0,644	0,005	195,820
δ	0,000	0,002	0,921	0,141	7,087
ZSV	-9,733E-07	0,000	0,073	0,023	43,591
V _{bit}	-2,181	0,592	0,000	0,043	23,063
VMA	0,991	0,251	0,000	0,002	543,070
VFA	0,027	0,025	0,279	0,006	173,945
G _{mb}	0,001	0,002	0,716	0,073	13,607
V _m	-0,737	0,328	0,027	0,001	1037,583
Temperature	0,008	0,005	0,131	0,727	1,376
σ _{max}	-0,001	0,000	0,000	0,682	1,466

As expected, due to the inclusion of all the parameters available, a multicollinearity problem exists. Eight out of the twelve regressors exceeded the VIF limit of 10. This means that adjustments should be made on the model selected. It has to be noted, that due to extreme VIF values, G* was automatically excluded by SPSS from the model.

- Checking for unusual points

There can be certain data points that are, in some way, classified as unusual from the perspective of fitting a multiple regression model. These data points are generally detrimental to the fit or generalization (statistical inference) of the regression equation. There are three main types of unusual points: outliers, high leverage points and highly influential points.

d) Outliers

An outlier is an observation (data point) that does not follow the usual pattern of points (they are far away from their predicted value). We decided already exclude the failed specimens from the analysis considering them as outliers, but not exclude the other cases, aiming for connecting their performance to composition. For this reason we continue to the next step.

e) Leverage Points

To determine whether any cases exhibit high leverage, a general rule of thumb is to consider leverage values less than 0,2 as safe, 0,2 to less than 0,5 as risky, and values of 0,5 and above as dangerous. Sorting the data points in a descending order based on their leverage, the maximum leverage obtained is 0,2435, which is slightly above the safe limit. No other data point exceeded 0,2, so this check is satisfied.

f) Influential points

Another output of the SPSS regression function, is each data point's Cook Distance. As a rule of thumb, if there are Cook's Distance values above 1, they should be investigated. Sorting these values in a descending order, the highest Cook's Distance obtained was 0,1611, which indicates that no point is in risk of being considered highly influential.

In combination with the leverage points, we conclude that no data point is risky enough to be removed.

- Normality

A common method we can use to check for the assumption of normality of the residuals is a histogram with superimposed normal curve (Figure 6.16) and a P-P Plot (Figure 6.17), which were both produced by the Linear Regression function.

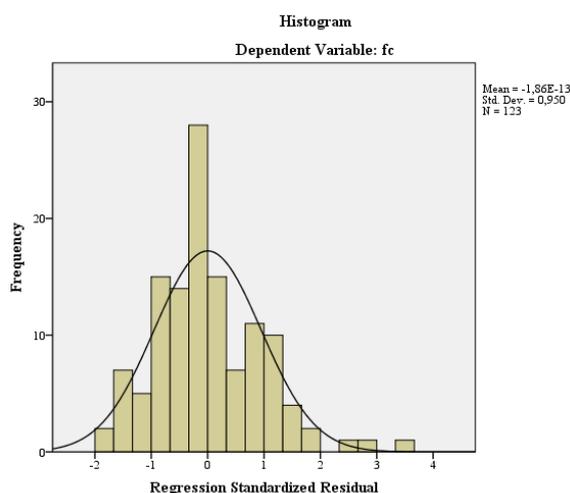


Figure 6.18 Histogram of standardized residuals

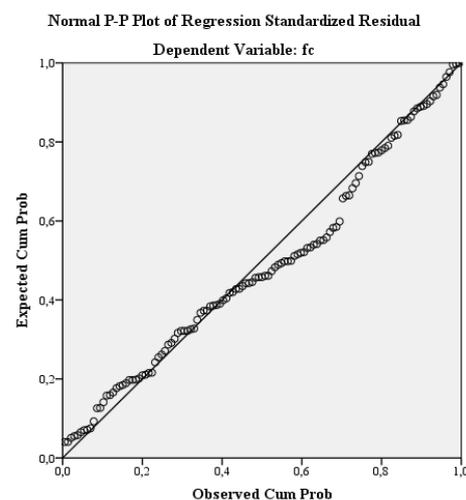


Figure 6.19 Normal P-P Plot

We can see from the histogram that the standardized residuals appear to be approximately normally distributed. The mean and standard deviation should have values of approximately 0 (zero) and 1, respectively, which is also satisfied. From the P-P Plot, we can see that although the points are not aligned perfectly along the diagonal line, they are close enough to indicate that the residuals are close enough to normal, for the analysis to proceed. As multiple regression analysis is fairly robust against deviations from normality, we can accept this result as meaning that no transformations need to take place.

After checking the assumptions we accept that the accuracy and quality of the prediction is going to be very poor. With two basic prerequisites being violated (linearity and homoscedasticity), the maximum Adjusted- R^2 achieved was below 0,5. As it will later be seen, the same applies in the prediction of ε_{1000} , even though the Adjusted- R^2 values are slightly improved. As a consequence, the target of obtaining a predictive relation for the permanent deformation resistance cannot be met.

This finding is not surprising since we already knew that there is big variability in the test results among the 5 NL-Lab works that were carried out in different labs. This variability comes from the sensitivity of the triaxial test in the friction reduction method chosen and the data acquisition (curve fitting). Things that are not explicitly defined in the standard's procedures, and consist one of the problems that triggered this research.

Despite this problem however, the investigation of different parameter combinations will still be useful in getting an indication of which of them plays an important role in the triaxial test. This analysis was also made in paragraph 6.2.3.1, where the comparison of lab to field specimens was the subject. In that case however, the collinearity factor was not taken into account, and the correlations were only looked from their monotonicity point of view (Spearman's correlation).

6.3.4 Predictors combinations

6.3.4.1 Strain slope (f_c)

The choice of parameter combinations will be a matter of their statistical significance in the model (Sig.), and their collinearity factor (VIF). Some representative sets of parameters, are presented in Table 6.19.

Table 6.19 Parameter combination sets for f_c

Set	1		2		3		4		5		6	
Pred.	Sig.	VIF	Sig.	VIF	Sig.	VIF	Sig.	VIF	Sig.	VIF	Sig.	VIF
(Constant)	0,617	-	0,979	-	0,000	-	0,011	-	0,002	-	0,674	-
$T_{R\&B}$	0,515	560,60									0,303	162,39
Pen	0,881	71,94									0,121	26,350
logA	0,644	195,82									0,675	80,857
δ	0,921	7,087										
ZSV	0,073	43,591	0,000	1,152	0,000	1,090	0,000	1,101	0,000	1,098		
V_{bit}	0,000	23,063	0,001	19,942	0,001	16,050	0,002	1,053	0,002	1,053		
VMA	0,000	543,07	0,000	539,78	0,000	381,54						
VFA	0,279	173,95	0,694	82,380								
G_{mb}	0,716	13,61	0,498	11,354			0,000	1,059	0,000	1,032		
V_m	0,027	1037,5	0,006	914,07	0,000	386,48						
Temperature	0,131	1,376	0,069	1,375	0,081	1,352	0,003	1,253				
σ_{max}	0,000	1,470	0,000	1,413	0,000	1,359	0,000	1,251	0,007	1,050		
R^2	0,500		0,489		0,486		0,366		0,314		0,134	
Adjusted R^2	0,446		0,453		0,459		0,339		0,291		0,112	

Starting from the initial model of all the parameters (Set 1) and the expected big collinearity problem, the combinations evolved around the parameters that showed a good significance. Excluding all the bitumen characteristics besides ZSV (Set 2), not only decreased the collinearity levels, but also slightly increased the Adjusted- R^2 , still at low levels though. Then, taking out the parameters with the very low significance levels (i.e. high Sig. values) (Set 3), the collinearity was improved and slightly the quality as well. Replacing the void content and VMA with density (Set 4), we see that the collinearity reached acceptable levels for all the parameters, decreasing however the relation strength. Set 5 examines the effect of excluding temperature from the relation. It is obvious that the strength significantly decreased, exposing the temperature's role in the triaxial test. Finally, Set 6 shows that by just including the bitumen characteristics, not only we expectedly have collinearity problems, but also the strength of the relation is the poorest.

In total, the composition parameters that show the biggest relation to permanent deformation are the ZSV, V_{bit} , VMA, V_m , Temperature and σ_{max} , all found in Set 3. VMA and V_m show a correlation, however they entail physical aspects that are not included in both of them at the same time, and this is the reason they were chosen. The test conditions naturally play an important role and this was the reason for their addition in the NL-Lab database.

6.3.4.2 Permanent strain (ϵ_{1000})

Following the same procedure, assumptions and checks, the optimum combination of parameters aimed at predicting ϵ_{1000} and identifying the parameters with the biggest influence was investigated. A representative overview of the combinations tested can be seen in Table 6.20.

Table 6.20 Parameter combination sets for ϵ_{1000}

Set	1		2		3		4		5		6	
Pred.	Sig.	VIF										
(Constant)	0,225	-	0,001	-	0,000	-	0,000	-	0,000	-	0,002	-
$T_{R\&B}$	0,507	564,07									0,310	162,38
Pen	0,911	72,438					0,006	2,312	0,007	2,310	0,914	26,350
logA	0,394	196,73									0,048	80,857
δ	0,180	7,123										
ZSV	0,171	43,749	0,000	1,152	0,000	1,049	0,000	2,390	0,000	2,380		
V_{bit}	0,969	23,063	0,908	19,942								
VMA	0,374	544,60	0,502	539,78								
VFA	0,045	175,48	0,000	82,380	0,000	41,142	0,000	41,152	0,000	41,003		
G_{mb}	0,187	13,789	0,042	11,354	0,033	9,658	0,011	9,854	0,020	9,423		
V_m	0,700	1041,7	0,316	914,07	0,000	62,867	0,000	62,912	0,000	61,551		
Temperature	0,033	1,381	0,028	1,375	0,021	1,339	0,018	1,339	0,039	1,189		
σ_{max}	0,070	1,479	0,157	1,413	0,246	1,332	0,211	1,333				
R^2	0,547		0,501		0,485		0,518		0,511		0,288	
Adjusted R^2	0,497		0,465		0,458		0,488		0,485		0,270	

It is clear that quality of prediction is marginally higher than the f_c . This possibly relates to the fact that ϵ_{1000} is more strict and accurate in its extraction from the creep curve. However, the levels of accuracy of the relation are still low enough to be used as a performance indication.

Regarding the parameters though, some indications can be extracted on their relationship to permanent deformation. In specific, the parameters from Set 5 seem to give the best combination of significance, collinearity and relation strength. They include both bitumen characteristics (Pen and ZSV), volumetrics (VFA, G_{mb} and V_m), and test conditions (temperature). It is notable that the maximum stress applied (σ_{max}) does not have a significant contribution to the strength of the relation (comparing Set 4 to 5), while in the case of f_c , it showed a strong one.

6.4 Friction reduction methods

The output of the testing stage 2 gave the curves, strain slopes, and permanent deformation values given in paragraph 6.1. The initial question was to examine the differences in the test's output that might result from the different friction reduction methods. Even though one might think that there should not be major differences, since the standard does not explicitly define this, the true fact, that can even be noticed at the first view of the curves, is that it does make a big difference. The creep curves are presented again below, in Figure 6.20.

The general image shows the 'No material' specimens resulting in the lowest strain levels, with values comparable to the 'Teflon' specimens, which resulted in slightly higher strain levels. The 'Plastic Wrap' specimens were the next, in terms of both strain levels and strain slope, with values almost double the numbers of Teflon and No Material. This behavior was also expected due to the clearly less friction between the specimen and the plates, which was already felt during the mounting of the specimens. Finally, the highest values were obtained for the specimens tested with Latex sheets and vacuum grease.

All the methods except for the Latex, consist of smooth curves, with no kinks and a relatively constant slope in the secondary creep stage. The 'Latex' curves though, contain sudden changes in their slope with kinks, during both at the first, and the second stage.

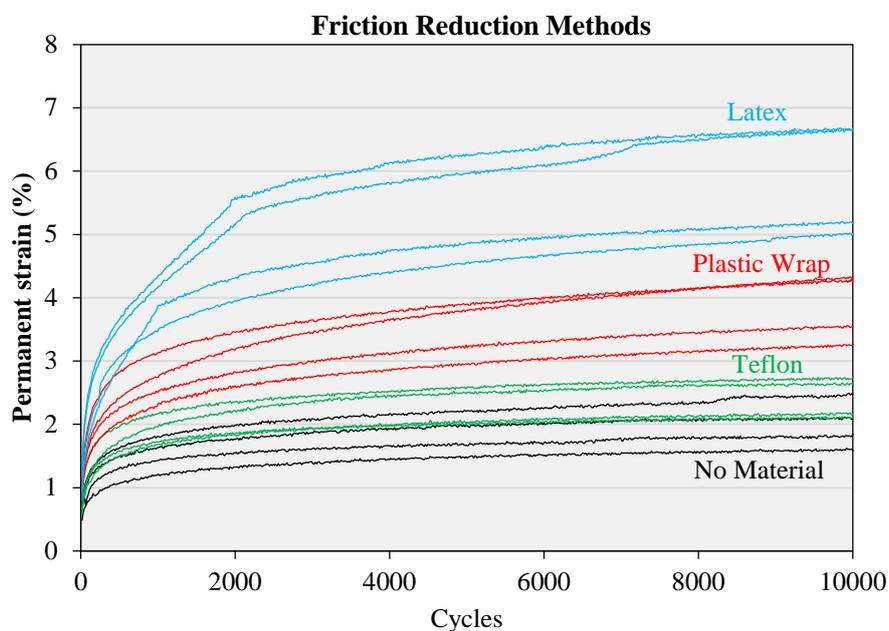


Figure 6.20 Creep curves distinguished by the different friction reduction method used

The boxplots representing the distribution of the permanent deformation levels both after 1000 cycles, and at the end of the test at 10000 cycles, are given below, along with the boxplot referring to the strain slopes, f_c .

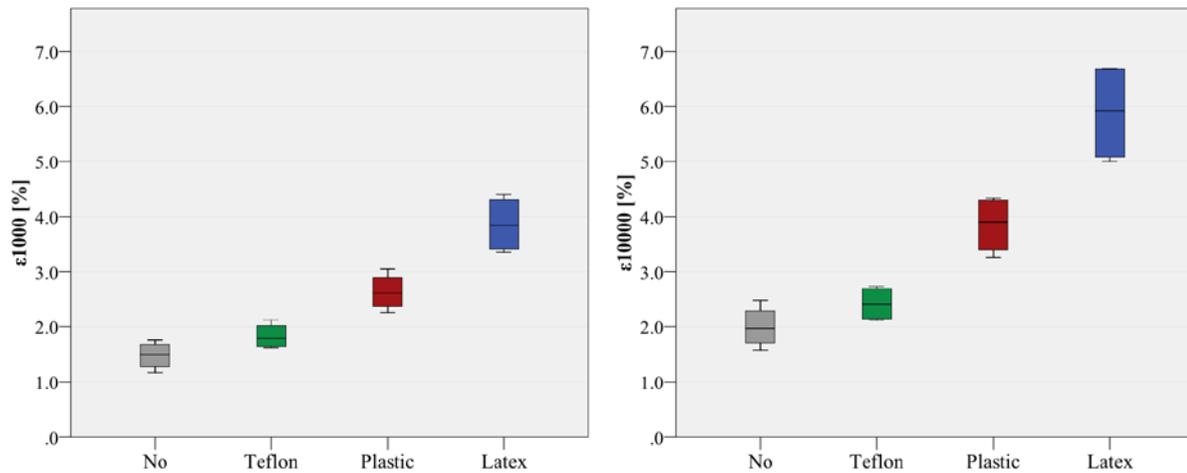


Figure 6.21 Permanent strain levels after 1000 cycles (left) and after 10000 cycles (right)

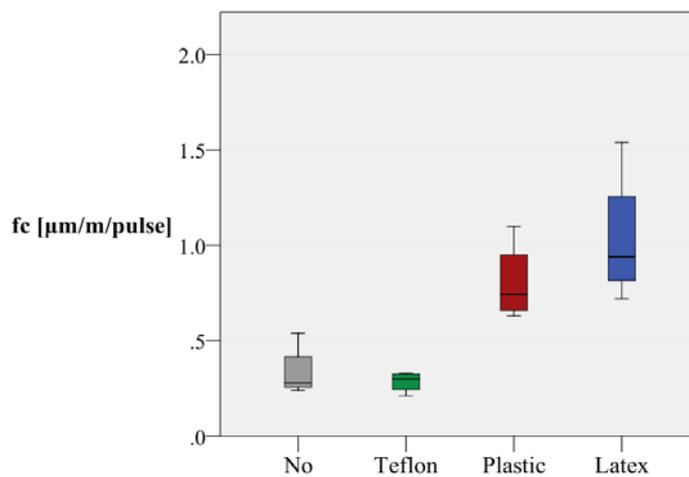


Figure 6.22 Strain slopes

Starting from the batch that was tested without any friction reduction means, with pure contact between the specimen and the loading plates, the result was the expected, and also the desired, in order to use it as a reference. Due to the high friction coefficient, the total pressure applied was not entirely transferred to the specimen, preventing it from deforming freely. As a result, part of the loading was translated into heat and minor surface wear of the loading plates, instead of the specimen fully uniform lateral expansion. Because asphalt concrete is stress state sensitive (i.e. stronger and stiffer if confined), this results in the lowest strain levels of all methods. The relatively good distribution of the strain values around their mean value, suggest a fairly good repeatability. The distribution in the f_c values is less accurate, being the second comparing to other methods.

Having already described the Teflon's behaviour during the unmounting of the specimens after each test (Paragraph 4.2.2.2), the curves and values came to confirm the expectations. The two Teflon sheets were tightly attached to each other, and required a considerable amount of sliding force to be detached. This 'glue' effect, even though it was initially thought to be only because of the vacuum grease used in the first test, was also observed in the next tests, meaning that it also resulted from the high load applied. In this way, friction was apparently not effectively reduced during the test, leading to results similar to the cases where no friction reduction was used, not fulfilling their target. These comparable strain levels are clearly seen in the boxplots, contrary to the f_c values which were slightly lower. However, this mismatch is not significant in this order of magnitude, and does not lead to some conclusion.

The group of specimens tested with plastic wrap and soap showed a very effective reduction of friction. With strain levels approximately two times higher, and a quite good distribution around the mean value, they performed in a stable manner. This stability is not seen in the strain slope though, where values have a medium spread around the mean. This can be explained by the sensitivity of this parameter; sensitivity that seems to increase exponentially with the decrease in friction levels, and also applies to the other methods.

Regarding the Latex membranes, this group of specimens with the highest deformation of all, showed the lowest accuracy in the distribution of values. Starting with a fairly good spread during the first 1000 cycles, this spread widened with the further loading. This widening follows the poor distribution of the f_c values, which is the worst of all methods. However it follows the trend described previously, that lower friction leads to bigger sensitivity in the strain slopes.

Despite the very effective reduction of friction in terms of numbers, the general image of this method is the poorest of all. Combining the medium repeatability, with the highly unstable curves, the Latex membranes seem to sacrifice accuracy in results with reduction in friction. The kinks observed in the curves, relate probably to the nature of the material. Latex being very elastic with a high Poisson's ratio (Rubber's Poisson Ratio = ~ 0.5), on one hand it allows almost a free deformation, but on the other hand the more it deforms, the more it tends to return to its initial state. This tendency creates high stresses that intervene in the axial loading of the specimen. In addition, because of the high pressure they are subjected to, in between the specimen and loading plate, the return in their initial state does not happen instantly, during every pulse, but at random points, when they find the 'freedom' to do so. This random subtraction is what disturbs the loading, and creates the kinks in the curves.

As far as the Latex membranes that were re-used after their first test are concerned (Figure 6.23), the image is far from comparable to the other methods. With permanent deformation levels 2 times higher than the membranes used only once, and 4 to 6 times than the plastic or Teflon sheets, the specimen's performance is characterized as inadequate, overpassing the failure thresholds.

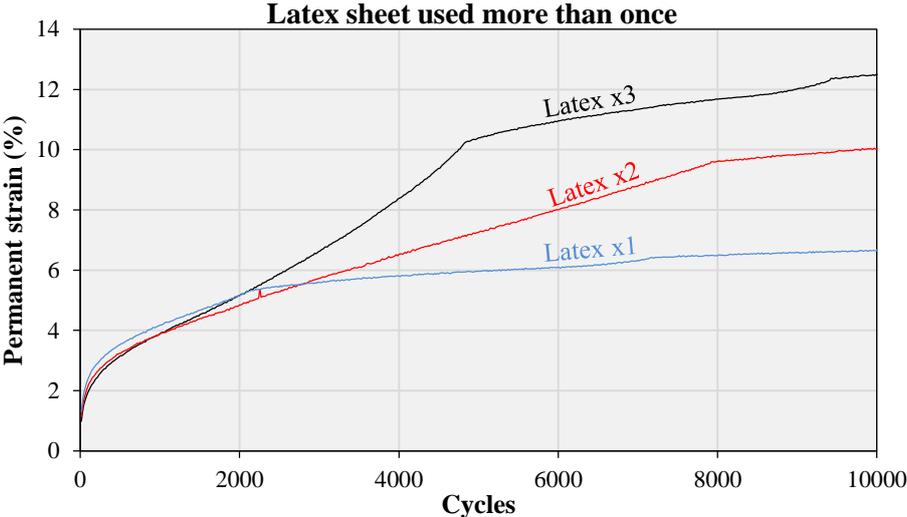


Figure 6.23 Curves of specimens tested with the same Latex sheet for more than one test

7 CONCLUSIONS & RECOMMENDATIONS

In this chapter, the collective conclusions drawn throughout this research project are listed and commented. They are divided in the two functional characteristics the research focused on; Moisture sensitivity and Permanent deformation. Finally, the recommended additions and adjustments for further research are presented.

7.1 Conclusions

7.1.1 Moisture Sensitivity

7.1.1.1 Conditioning methods

Starting from the conditioning methods comparison, we can say that the MIST protocol is the most damaging of all. It leads not only to physico-chemical changes, but also to changes in the specimen's internal structure. This is the result of the high pressure utilized during the conditioning. It was visible in the CT-scans and by naked eye, that big cohesive and adhesive types of failure took place in the specimen's mass, something that was not observed in the other conditioning methods. The disadvantage of these big structural changes was the big shape deformation that made the specimens' alignment in the ITT test impossible, meaning that there was an undesirable stress distribution during the test. In Frost Protocol IV, where the specimen was subjected in both water and thermal loading, the reduction in strength was similar to the MIST, however the visual inspection did not show major cracks or changes in the structure. Most probably, the frost effect damaged the structure internally and on a micro-scale, not being visible by naked eye. In addition, the damage reached is mostly attributed to the presence of water alone, since Protocol III already led to big reductions.

The conclusion, comparing these methods to the standard, is that even though the standard conditioning protocol may lead to considerable changes, after only 3 days of conditioning, these changes are only restricted to the chemical properties of the mixture. Additional loadings, affecting

also the physical state of the mixture, would however be beneficial in characterizing a mixture more elaborately. The damages would be magnified and the performance indicated in a more sensitive way. This is because the physico-chemical reaction of the mixture's components will be exposed, and the way they bond to form a structure (i.e. asphalt concrete) will 'violently' be changed.

7.1.1.2 Comparison of lab to field determined properties

Despite the indications from this research's results, that lab produced specimens perform better than the field, extending the database to all NL-Lab's works, the image becomes vague. In general, the conclusion that can be safely drawn, is that lab and field produced specimen perform differently, and their behavior does not match. There are cases where indeed lab strength values are higher, but also cases where they are lower. Since in each work the specimens were produced under different conditions, depending on the lab facilities and operators, the inability to draw a conclusion on how lab compares to field, is mostly attributed to this fact. Asphalt concrete is very sensitive to production conditions, for example mixing and compaction temperatures and energy, and this is the reason composition parameters did not show a direct correspondence to performance. This made it difficult to base the lab and field comparison on their mixture properties. Regarding the void distribution inside the specimen's mass, CT-scans also did not show major differences between lab and field, at least in the small number of samples scanned.

An observation that could possibly be made was that field specimens were found to be statistically related to composition parameters in a monotonic manner, while lab specimens did not. This confirms the fact that differences in the production process in the lab and field indeed exist, and that additional parameters that are better related to performance should be utilized.

7.1.1.3 Performance prediction

Even though a strong linear relationship between performance and composition parameters did not exist, the combinations of the parameters led to a fairly good relation. With an adjusted-R² of 0,61 for the dry subset, and 0,56 for the wet, both including statistically significant parameters, the predicting possibilities are not at very high levels, but are still considered acceptable, taking into account the big variability and the number of data points used. By using the relations suggested below, the ITSR value of a mixture can be obtained with an adjusted-R² of 0,58.

$$ITS_{dry, pred} = -27,541 + 0,1329 \cdot T_{R\&B} - 2,926 \cdot 10^{-6} \cdot ZSV + 0,0079 \cdot G_{mb} + 0,871 \cdot V_{bit} \quad [MPa]$$

$$ITS_{wet, pred} = -46,818 + 0,2242 \cdot T_{R\&B} - 5,129 \cdot 10^{-6} \cdot ZSV + 0,0145 \cdot G_{mb} + 0,397 \cdot V_{bit} \quad [MPa]$$

The intervals that were included, in which these parameters are related to ITS are:

Parameter	Minimum	Maximum
$T_{R\&B}$	56,2°C	76,2°C
ZSV	28700 Pa.s	712240 Pa.s
G_{mb}	2320,5 kg/m ³	2451,4 kg/m ³
V_{bit}	4,03%	4,50%

7.1.2 Permanent deformation

7.1.2.1 Comparison of lab to field determined properties

Considering the statistical analyses on the dataset, and the problems encountered during the determination of the performance parameters, the strength of the conclusions was expected to be reduced. Medium correlations between f_c and ϵ_{1000} , vagueness in the standard's description of their computation, and unstable performance due to the adjusted test conditions, led to difficulties in comparing lab to field values. Also, the participation of many research labs and the variability in the test setups they used, made the comparison between all NL-Lab's works inconsistent.

The general conclusion on lab to field comparison is, as in the case of moisture sensitivity, that there is not a consistent trend. There are works where lab performs worse than the field, and works with the opposite observation. This is the result of all the aforementioned reasons, making the correspondence of performance to composition parameters almost inexistent. The certain conclusion nevertheless, is that lab produced and field produced specimens perform in a different manner and do not match in any case.

7.1.2.2 Performance prediction

With all the inconsistencies and problems mentioned in the previous paragraph, predicting the permanent deformation performance also proved to be of low quality. Certain assumptions that are necessary for the regression analysis were not met, and this was depicted in the low adjusted-R² values obtained. This means that the possibilities of obtaining a predicting relation that outputs a rough expectation of a mixture's performance on permanent deformation are very reduced based on the current dataset.

Some conclusions can however still be drawn. It is confirmed that indeed using f_c as a representative factor is less descriptive and accurate comparing to ϵ_{1000} . Also, regarding the composition parameters that show possibilities of being related to performance, not only from a physical point which is expected, but also in a statistical manner, these include: bitumen characteristics (Pen and ZSV),

volumetrics (VFA, G_{mb} and V_m), and test conditions (Temperature). A notable observation was that the maximum stress applied (σ_{max}), contrary to expectations, did not show a big statistical correlation to the obtained ϵ_{1000} , while in the case of f_c it did. Probably the root of this unexpected observation is the instability of the test when the temperature and stress parameters were adjusted differently from the standard.

7.1.2.3 Friction reduction methods

It is an obvious conclusion that the choice of friction reduction method plays a major role in the triaxial test's output. The same material, tested in 4 different ways, led to different results in each case. With the Teflon sheets, that are currently the material of prescription in the standard, leading to the least effective friction reduction, and the latex sheets, that were used in the previous standard version, leading to the highest friction reduction, it is clear that this transition resulted in inconsistencies. Taking into account the fact that the requirements remained the same, the problem increases in importance.

A material that is not, and was never, suggested by the standard, is a plastic foil in combination with conventional soap. This combination led to a very effective friction reduction, which was also relatively stable and consistent. The fact that it is also easy and economical to apply, makes it a possible alternative.

7.1.2.4 General conclusions

The fact that triaxial test is very sensitive to small adjustments in its setup, and as a result has a poor repeatability, leads to the conclusion that with the current EN standard and the way it prescribes the test, inconsistencies in the results are highly potential. Referring to the determination of the f_c values, the fact that they give a poorer representation of the creep curve comparing to ϵ_{1000} , was confirmed by the big deviations and bad statistical relations found.

7.2 Recommendations

7.2.1 Moisture sensitivity

- Now that some indications on the destructibility of the conditioning methods have been obtained, the next step is to define their accuracy. This will be achieved by an additional research on conditioning methods, having three different mixtures, and not just one. The mixtures should be designed to have a good, medium and bad behavior in moisture damage respectively, and will all be conditioned and tested in the same manner. In this way the conditioning methods will be compared based on their ability to categorize these mixtures.
- Adjustment of the MIST conditioning parameters is suggested to fit the case of dense asphalt mixtures. This adjustment includes a reduction in the pore pressure applied, to avoid the excessive deformation observed. In addition, a new mounting setup that accommodates any misalignments and leads to uniform stress distribution will also be beneficial.
- Since the potential for predicting moisture sensitivity performance was fairly good, it is suggested that it will be extended to also include surface layers and porous mixtures, additionally to the base/binder layers and dense mixtures that were tested so far.
- Additional research on lab to field comparison, this time including parameters that were taken into consideration so far. One parameter that could describe the differences between lab and field is the energy input during the mixing process.
- Since the CT-scans gave some small indication on how the void distribution looks like in the lab and in the field, this can be a starting point to further look into it, to confirm or reject the relation found from the small sample number.

7.2.2 Permanent deformation

- Establish the use of ϵ_{1000} in place of f_c in the mixture characterization. Also, come up with a new parameter that combines both and gives more information.
- Adjust the standard so that it explicitly prescribes a friction reduction method, that is universally used in the same way, in all the lab tests.
- Suggested material to be used as a friction reduction method: soap and plastic foil (similar to Luflexen® from Basf)
- In order to obtain an accurate and descriptive prediction relation, a consistent research has to be made. This research should consistently use one friction reduction material. Also, it should strictly take place under uniform production, storing and testing conditions, to prevent any additional factors from altering the composition-performance relation.

8 BIBLIOGRAPHY

- [1] S. Erkens, „Relating lab properties of high percentage RAP mixtures to field performance - The NL-LAB program,” in *TRB*, Washington, 2015.
- [2] J. Molenaar, Performance related characterization of the mechanical behaviour of asphalt mixtures, Delft, 2003.
- [3] G. Seleridis, „Rijkswaterstaat Internship - Technical Report,” Delft, 2013.
- [4] N. R. C. Transportation Research Board, Hot-Mix Asphalt Paving Handbook, Washington, DC, 2000.
- [5] A. Varveri, S. Avgerinopoulos, A. Scarpas, A. Collop en S. Erkens, „On the combined effect of moisture diffusion and cyclic pore pressure generation in asphalt concrete,” in *TRB*, Washington, 2014.
- [6] „Pavement Interactive,” 22 April 2011. [Online]. Available: <http://www.pavementinteractive.org/article/moisture-susceptibility/>. [Geopend 18 November 2016].
- [7] R. Terrel en S. Al-Swailmi, „Water sensitivity of Asphalt-aggregate mixes: Test selection,” Washington, DC, 1994.
- [8] FHWA, Distress identification manual for the long-term pavement performance program, Fourth Revised Edition FHWA-RD-03-031, Washington, 2003.
- [9] R. Carvalho, Prediction of permanent deformation in asphalt concrete, Maryland, 2012.
- [10] „Pavement Interactive,” 2008. [Online]. Available: <http://www.pavementinteractive.org/article/rutting/>. [Geopend November 2016].
- [11] M. Miljkovic, „Rutting mechanisms and advanced laboratory testing of asphalt mixtures resistance against permanent deformation,” *Facta Universitatis, Series: Architecture and Civil Engineering*, vol. 9, nr. 3, pp. 407-417, 2011.
- [12] McGennis, Anderseon, Kennedy en Solaimanian, Introduction to Superpave asphalt mixture design, Washington, DC: Federal Highway Administration, Office of Technology Applications, 1994.
- [13] E. Remisova, „Resistance to permanent deformation in binder content and film thickness viewpoint,” *ACEE*, January 2013.
- [14] J. Sousa, J. Craus en C. Monismith, „Summary report on permanent deformation in asphalt concrete,” Institute of Transportation Studies, University of California Berkeley, 1991.
- [15] „NCHRP Synthesis 274: Methods to Achieve Rut-resistant Durable Pavements,” Transportation Research Board, 1999.
- [16] R. Izzo en K. Stuart, „Correlation of superpave $G^*/\sin\delta$ with rutting susceptibility from laboratory mixture tests,” Transportation Research Board, 1995.

- [17] „NPR-CEN/TS 15325: Bitumen and bituminous binders - Determination of Zero-Shear Viscosity (ZSV) using a Shear Stress Rheometer in creep mode,” 2008.
- [18] P. M. en R. C., „Binder Rheology and Asphaltic Pavement Permanent Deformation; The Zero-Shear Viscosity,” in *Eurasphalt & Eurobitume Congress*, Strasbourg, 1996.
- [19] D. Anderson, Y. Le Hir, J. Planche en D. Martin, „Zero Shear Viscosity of Asphalt Binders,” Transportation Research Board, 2002.
- [20] J. De Visscher en A. Vanelstraete, „Practical test methods for measuring the zero shear viscosity of bituminous binders,” vol. 37, nr. 5, 2004.
- [21] Laerd Statistics, „Multiple Regression in SPSS,” [Online]. Available: <https://statistics.laerd.com/>. [Geopend May 2017].
- [22] „Wikipedia,” [Online]. Available: <https://en.wikipedia.org/wiki/P-value>. [Geopend 06 2017].
- [23] Wikipedia, „Box plot,” [Online]. Available: https://en.wikipedia.org/wiki/Box_plot. [Geopend May 2017].
- [24] Wikipedia, [Online]. Available: <https://en.wikipedia.org/wiki/Quartile>. [Geopend May 2017].
- [25] C. Giezen, D. V. Vliet, S. Valcke, S. Erkens, J. Voskuilen, P. Kuijper, A. Scarpas, A. Varveri en S. Avgerinopoulos, „TNO 2013 R11050, IQ-2012-64 Frost Damage Research,” InfraQuest, Delft, 2012.
- [26] „NEN-EN 12697-25: Bituminous mixtures - Test methods - Part 25: Cyclic compression test,” 2016.
- [27] „NEN-EN 12697-23: Bituminous mixtures - Test methods for hot mix asphalt - Part 23: Determination of the indirect tensile strength of bituminous specimens,” 2003.
- [28] Laerd Statistics, „Two-Way ANOVA in SPSS,” [Online]. Available: <https://statistics.laerd.com/>. [Geopend June 2017].
- [29] Laerd Statistics, „Pearson's Product-Moment Correlation,” [Online]. Available: <https://statistics.laerd.com/>. [Geopend June 2017].
- [30] Laerd Statistics, „Spearman's Correlation,” [Online]. Available: <https://statistics.laerd.com/>. [Geopend June 2017].
- [31] Florio, Berti, Kasbergen, Villani, Scarpas, Erkens, Sangiorgi en Lantieri, „American and European Mix Design Approaches Combined,” in *Transportation Research Board*, Washington, 2014.
- [32] M. Phillips en C. Robertus, „Binder Rheology and Asphaltic Pavement Permanent Deformation; The Zero-Shear Viscosity,” in *Eurasphalt & Eurobitume Congress*, Strasbourg, 1996.