Dynamic Wheel/Rail Rolling Contact at Singular Defects with Application to Squats

Xin Zhao

Dynamic Wheel/Rail Rolling Contact at Singular Defects with Application to Squats

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 29 juni 2012 om 10:00 uur

door Xin ZHAO
Master of Engineering, Southwest Jiaotong University, P.R. China
geboren te Taian, Shandong, P.R. China

Dit proefschrift is goedgekeurd door de promotoren:
Prof.dr.ir. A.A.A. Molenaar
Prof.dr.ir. R.P.B.J. Dollevoet

Copromotor: Dr.ir. Z.Li

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof.dr.ir. A.A.A. Molenaar,	Technische Universiteit Delft, promotor
Prof.dr.ir. R.P.B.J. Dollevoet,	Technische Universiteit Delft, promotor
Dr.ir. Z.Li,	Technische Universiteit Delft, copromotor
Prof.dr.ir. D.J. Schipper,	Technische Universiteit Twente
Prof.ir. A.Q.C. van der Horst,	Technische Universiteit Delft
Ir. T. Sysling,	ProRail
Dr. A. Ekberg,	Chalmers University of Technology, Sweden
Prof.dr.ir. J.G. Rots,	Technische Universiteit Delft, reservelid

To my wife, Zhu Yan, the best woman in the world

Acknowledgements

Last but not the least, this dissertation is dedicated to my wife, Zhu Yan, the best woman in the world.

Xin Zhao
24th May 2012, Delft

Summary

Squats, as a kind of short wavelength rail surface defects, have become one of the main rolling contact fatigue problems in railways worldwide. The purpose of this work is to better understand the squatting phenomenon, contribute to reduction and even prevention of squat occurrence, and thereby reduce the related costs.

To such an end, a new modeling approach has been developed in this dissertation, i.e. a three-dimensional finite element (FE) model of the vehicle–track interaction. Both wheelset and rail are simulated as three-dimensional continua. A detailed surface-to-surface contact algorithm is integrated within the FE model in order to solve the frictional rolling contact between the wheel and rail. Different traction/braking efforts are simulated. Detailed modeling of the wheel and rail ensures the consideration of important eigen-modes related to squats, mainly in the high frequency range. Other structures of the vehicle–track system are also modeled to appropriate extents.

An estimate of contact stresses with sufficient accuracy is the basis of further dynamic, stress, and fatigue analyses of squats. The FE model has been validated for both normal and tangential contact solutions by comparing it to the widely accepted Hertz theory and Kalker's CONTACT program. Due to the fact that Hertz theory and CONTACT are only applicable to static contact problems, the steady-state rolling contact between smooth wheels and smooth rails, with the contact occurring in the middle of the rail top, is simulated by the FE model for the purpose of validation. The results show that the FE model is reliable for the solution of frictional rolling contact. On the other hand, the FE model can also take into account actual contact geometry, material non-linearity, and transient effects, which are required for more complicated cases like the wheel–rail rolling contact at a squat. Therefore, the newly developed modeling approach provides a valid and promising tool to solve the problem of rolling contact in the presence of friction.

With the validated FE model, the influence of plastic deformation on the solution of frictional wheel–rail rolling contact is further investigated. A bi-linear elasto-plastic material model is employed. It is found that the contact geometry change caused by plastic deformation can significantly modify both the normal and the tangential solutions.

When squat type defects are added to the rail top, the calculated dynamic contact forces show a good agreement in wavelength with observed squats in the field. This means that vibrations related to squats are captured by the FE model, proving the applicability of the FE modeling in treating the high frequency dynamics of a system containing rolling contact. Furthermore, based on the simulations and field observations, a growth process of squats from light to mature state has been postulated. This postulation has been validated by track monitoring conducted in the Netherlands.

Further analyses of the FE simulations show that squats mainly excite the vehicle–track system at two frequencies. The vibration component with the lower frequency can transfer down to the ballast layer, especially at the support close to the squat. The high frequency vibration component has similar magnitude at several fastenings near the squat and is negligible at the ballast layer. For the investigated rolling speed range between 40

and 140 km/h, both two vibration components increase in magnitude and wavelength with the rolling speed. The vibration component at the higher frequency can be absent when the rolling speed is sufficiently low, e.g. at 40 km/h for the simulated system.

By evaluating the stress under rolling contact and comparing it with material strength, it is derived that an initial rail surface defect such as an indentation can only grow into a mature squat when it is over a critical size of 6–8 mm under the typical Dutch railway condition. This critical size has also been verified by monitoring tests.

The work of this dissertation formed the basis for a 'Guideline to Best Practice of Squat Treatment', written upon invitation by the International Union of Railways.

Samenvatting

De Squats, een soort korte-golf spoorstaafdefect, is wereldwijd één van de grootste RCF problemen (Rolling Contact Fatigue, vermoeidheidscheuren in het spoorstaafoppervlak) geworden. Het doel van dit onderzoek is om het squating fenomeen beter te begrijpen, hierbij is het uiteindelijk de bedoeling om het voorkomen van squats te verminderen en ze zelfs te vermijden en zo de hiermee verbonden onderhoudskosten te drukken.

Hiervoor is een nieuwe modelleringsbenadering ontwikkeld in deze scriptie, een 3-D eindige elementen (Finite Element, FE) modellering van de voertuig-spoorweg interactie. Zowel het wiel als de spoorstaaf worden gesimuleerd als een 3-D continuüm. Een gedetailleerd model contact algorithme is geïntegreerd in het FE-model, om zo het rollend contact met wrijving tussen wiel en spoorstaaf op te lossen. Aandrijving en remmen worden voor verschillende situaties gesimuleerd. De gedetailleerde modellering van het wiel en spoor verzekert dat de hoogfrequente eigenmodes, gerelateerd aan squats, correct worden gesimuleerd. Andere structuren van het spoorweg-voertuig geheel worden op passende wijze gemodelleerd.

Een schatting van de contactspanningen, met voldoende nauwkeurigheid, is de basis van verdere dynamische, spannings-, en vermoedheidsanalyses van squats. Het FE-model is in dit onderzoek gevalideerd voor zowel de normale als tangentiële contactoplossingen door het te vergelijken met de oplossing van de wijdverspreide theorie van Hertz, en met de resultaten bekomen met het programma CONTACT van prof. Kalker. De theorie van Hertz en CONTACT zijn enkel toepasbaar op statische problemen, daarom is het steady-state rollend contact, tussen een glad wiel en een gladde spoorstaaf en met het contactpunt op het loopvlak van de spoorstaaf, gesimuleerd met een FE model ter validatie. De resultaten tonen aan dat het FE model betrouwbaar is om het rollend contact probleem op te lossen. Het FE model brengt daarenboven ook de eigenlijke contactgeometrie, de non-lineariteit van het materiaal en de transiënte verschijnselen in rekening en deze aspecten zijn verreist voor ingewikkelder gevallen zoals wiel/spoorstaaf rollend contact over een squat. Daarom is het nieuwe model een veelbelovende methode om het probleem van rollend contact met wrijving op te lossen.

Met het gevalideerde FE-model is de invloed van plastic vervorming op de oplossing van wiel-spoor rollend contact met wrijving verder onderzocht. Hiervoor wordt een bilineair elastisch-plastisch materiaal model aangewend. Er wordt geobserveerd dat de verandering van de contactgeometrie a.g.v. plastische deformatie zowel de normale als de tangentiele contactoplossing significant beïnvloedt.

De berekende golflengtes van dynamische kracht die ontstaat bij de passage van een wiel over een squat op de top van de rail komen goed overeen met de golflengtes geobserveerd op de spoorstaaf. Dit wil zeggen dat de trillingen afkomstig van de squats gevat kunnen worden in het FE model, hetgeen bewijst dat het model goed kan omgaan met de hoogfrequente trillingen van het rollende contact systeem. Op basis van simulaties en observaties is de hypothese naar voor geschoven dat squats groeien van licht naar zwaar. Deze hypothese is gevalideerd met spoorweg monitoring, uitgevoerd in Nederland.

De verdere analyses van de FE simulaties tonen aan dat de squats hoofdzakelijk twee frequenties opwekken in het voertuig-spoorweg systeem. De trilling met de lagere frequentie kan overgebracht worden aan het ballastbed, vooral bij de dwarsliggers dicht bij de squats. De trilling met de hogere frequentie heeft een gelijkaardige amplitude en heeft vooral invloed op de spoorstaafbevestiging, maar de invloed op het ballastbed is te verwaarlozen. Voor onderzochte rollende snelheidsinterval van 40-140 km/u, stijgen de twee trillingscomponenten zowel in omvang als golflengte met de rollende snelheid. De trillingscomponent bij de hogere frequentie kan afwezig zijn wanneer de rollende snelheid voldoende laag is, b.v. bij 40 km/u voor het gesimuleerde systeem.

Door de contactspanning te evalueren en het te vergelijken met de materiaalsterkte, wordt afgeleid dat een initieel defect van de spooroppervlakte, zoals een inkeping, slechts tot een zware squat kan uitgroeien, als het over de kritieke grootte van 6-8 mm, op de typische condities op het Nederlandse spoorwegnetwerk, bereikt. Het bestaan van deze kritische grootte is bevestigd door het monitoren van bestaande spoorlijnen.

Het werk van deze scriptie is de basis voor een Guideline to "Best Practice of Squat Treatment", geschreven op uitnodiging van de Internationale Unie van Spoorwegen.

Abbreviations

ALE	Arbitrary Lagrangian Eulerian
ABS	Anti-lock Brake System
BE	Boundary Element
COF	Coefficient Of Friction
ERRI	European Rail Research Institute
FE	Finite Element
FFT	Fast Fourier Transform
ICM	Intercitymaterieel (Inter city multiple unit)
MGT	Million Gross Tonnage
NS90	A sleeper type in the Netherlands
ProRail	Infrastructure manager of Dutch railway
RCF	Rolling Contact Fatigue
R260Mn	A rail material
S&C	Switches and Crossings
SNCF	French National railway
SRSD	Singular Rail Surface Defects
VIRM	Verlengd interregiomaterieel (A double-decker passenger train)
V-M	Von Mises
2D	Two Dimensional
3D	Three Dimensional
54E1	A rail profile

Table of contents

1   Introduction

In order to achieve a train operation with acceptable levels of vibration and noise, and with an allowable deterioration rate of the whole system, relatively high geometry quality of wheels and rails is necessary due to the high elastic modulus of steel material and the relatively high speeds of trains.

Rolling contact fatigue (RCF) may occur on initially smooth running surfaces of rails owing to the high contact stress. Once RCF exists, the contained cracks may propagate transversely and further result in a rail fracture, increasing the possibility of derailment. This research focuses on squats [1.1, 1.2], which are a type of RCF characterized by isolated occurrences on the running surfaces of rails.

1.1   The squat problem

A squat identifies a particular form of isolated rail surface damage that is located in the rail top, as shown in Figure 1.1. Field observations have shown that squats mainly occur on tangent tracks and shallow curves, and usually appear on only one rail at each location. Mature squats, which are normally accompanied by the widening of the rolling band, have a typical lung-like shape. One could also see that they look like permanent deformation caused by somebody sitting or squatting on the rail (See Figures 1.1(b) and 1.1(c)). According to [1.3], a network of cracks may be found beneath the dark surface of squats.

Squats were first reported in the 1950s in Japan where they were described as "black spots" [1.4–1.8]. In the 1970s they became known in the UK [1.9]. In other European countries they were reported later [1.10, 1.11]. Currently, squatting has become a major failure mode of rails worldwide. According to [1.12], 23.4% of defective rails in the French National Railway (SNCF) in 1999 were caused by squats; in Japan, squatting was recognized as a widespread phenomenon in both Shinkansen lines and narrow gauge lines [1.7, 1.8, 1.13].

1.1.1   Squats in the Netherlands and the Dutch classification

In the Netherlands, squats are mostly found on tangent tracks and shallow curves, on switches and crossings. The statistics given in Table 1.1 show that squats are widespread throughout the Dutch railway network [1.11]. Note that for the statistics, the whole track was divided into sections of 50 m long, and a section was considered squatted if one or more squats were present.
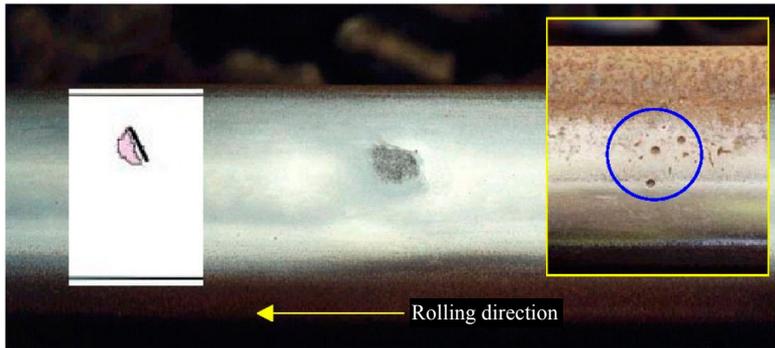
(a)
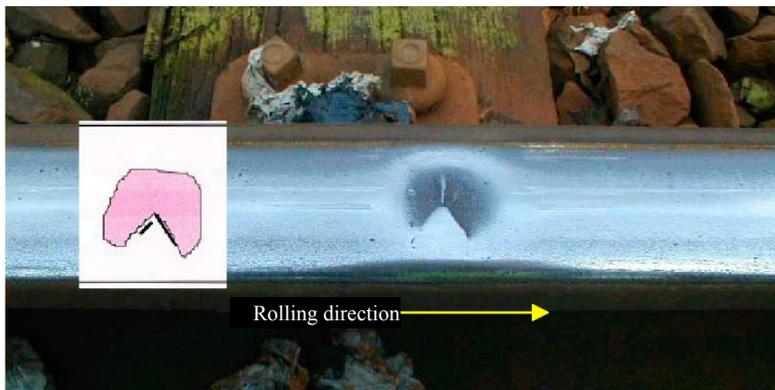
(b)
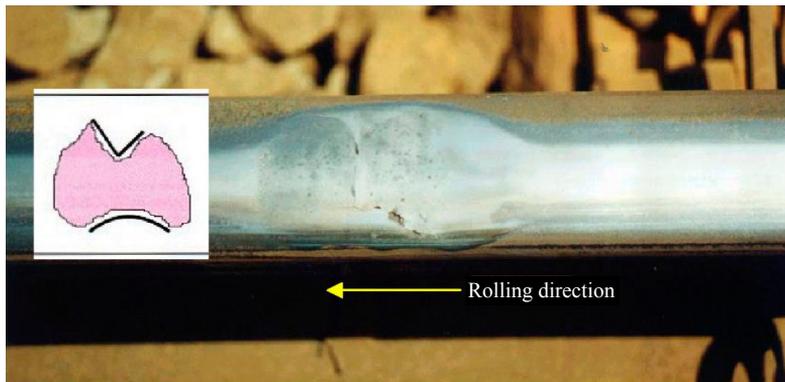
(c)

Figure 1.1. Squats at different stages. (a) Class A, (b) class B, and (c) class C (taken from [1.1]).

Table 1.1. Statistics of squats in the Netherlands in 2005 [1.14].

	Total	Squatted	Percentage of the squatted with respect to the total
Track length (km)	6500	1405	21.6%
Switch parts	8700	2321	26.7%

For track maintenance, squats in the Netherlands are classified into three categories, namely light (class A), moderate (class B), and severe (class C) [1.1], see Figure. 1.1. Schematic geometries of squats are also shown in the figure on the left side of each photograph. Light squats such as the one in Figure 1.1(a) do not normally have the typical two-lung shape mentioned above. They may be caused by indentations, wheel burns (also known as rail burns), welds, short pitch corrugation (hereinafter referred to simply as corrugation), etc. It is important to note that small black spots like the ones shown on the right side of Figure 1.1(a) should not be classified as class A squats because they will probably not grow up into mature squats due to their small sizes (more information is given in Chapter 5). Figures 1.2 and 1.3 give examples of squats in different states that were initiated from various origins.

(a)

(b)

Figure 1.2. Squats initiated from (a) an indentation at the center (Courtesy of René Heyder, Deutsche Bahn AG) [1.15] and (b) a wheel burn.

(a)

(b)

Figure 1.3. Squats initiated from (a) a thermite weld and (b) corrugation.

1.1.2   Squats versus head checks

Besides squats, head checks are another major type of rail RCF [1.16, 1.17]. An example of head checks is shown in Figure 1.4. The main differences between head checks and squats are listed in Table 1.2. Note that "high frequency" in this dissertation refers to a band from a few hundred Hz to above 2000 Hz (see Chapter 4).
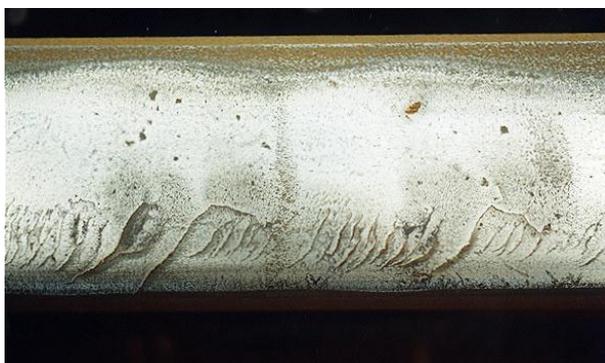
Figure 1.4. Severe head checks (taken from [1.17]).

Table 1.2. Differences between squats and head checks [1.17–1.20].

	Squats	Head checks
Occurrence on track sections	Mainly on straight tracks or shallow curves with radii larger than about 3000 m	Mainly on outer rails of curves with radii of ~ 500–3000 m
Susceptible parts of the rail	Rail top	Gauge corner and gauge shoulder
Appearance in the field	• Small depressions in the rail top • Two-lung shape for mature ones • V, U, Y, or circular shaped cracks often associated with mature ones • Often isolated	• An array of small, closely spaced, nearly parallel surface cracks • Continuously occur along a section of rail
Initiation mechanism	Visually related to rail top geometric defects Mechanism is not clear yet (to be investigated in this dissertation)	Centrifugal force in curves brings flange–gauge corner contact → Substantial lateral contact force and spin / (Relatively) conformal contact → Material at gauge corner or shoulder sustains higher stress level than at rail top, leading to head checks
Mechanical characteristic	Wheel–rail impact noise can be heard, suggesting a relation with (high-frequency) dynamic rolling contact	Attributed to (quasi-)static contact, resulting in continuous occurrence

An extensive discussion of the initiation mechanism and treatment of head checks is presented in [1.17] based on (quasi-) static rolling contact mechanics. In view of the distinctively different characteristics of squats and head checks, especially their relation with rail top geometric defects in initiation and growth, treatment of rolling contact in statics will not suffice for squats. A different approach should therefore be sought, as is presented in this dissertation.

1.2 Consequences of squats

From the viewpoint of railway infrastructure managers, the consequences of squats can mainly be seen from the following four aspects.

1.2.1 Accelerated rail deterioration

With the growth of a squat, the dynamic contact forces gradually increase due to the enlargement of the geometric deviation. This accordingly leads to an increase of the resulting stress. Thus, the growth rate of squats increases with time as shown in Figure 1.5 [1.21].
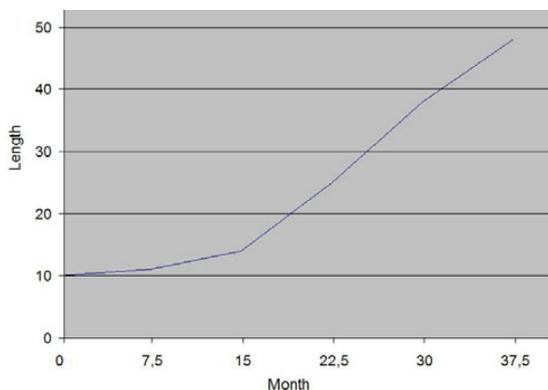
Figure 1.5. Length (in mm) of a squat vs. time (in months) [1.21].

It is widely believed that squats initiate in plastically deformed metal of the running surface, while are not associated with any metallurgical faults [1.3, 1.13, 1.16, 1.22]. Ultrasonic testing and eddy current testing performed in the Netherlands have shown that cracks are usually absent at the early stage of squats. Longitudinal–vertical sectioning of squatted rail specimens has shown that cracks contained in the squats first propagate at a shallow angle to the contact surface after initiation. These cracks may subsequently branch downwards into the rail head and form transverse cracks when they reach a depth of about 3 to 6 mm. In the final stage, these transverse cracks may further lead to a brittle fracture of the rail [1.3, 1.13, 1.18, 1.22].

As a result, the service life of the rail is greatly reduced by squats [1.9, 1.23, 1.24]. Squats may threaten rail safety, especially when multiple squats occur within a short distance (Figure 1.6).

1.2.2 Track deterioration owing to high frequency dynamic forces

Once a squat exists, the related geometric deviation can excite high dynamic force at the wheel–rail interface. The dynamic force, which is transmitted into the entire structure of the vehicle–track system, will accelerate the deterioration of ballast and fastenings and may cause hung sleepers, sleeper cracking, walking, and skewing [1.25–1.28]. Figure 1.6 shows an example of the influence of the dynamic force on sleepers and ballast, where the white stoning phenomenon can be seen.

(a)

(b)
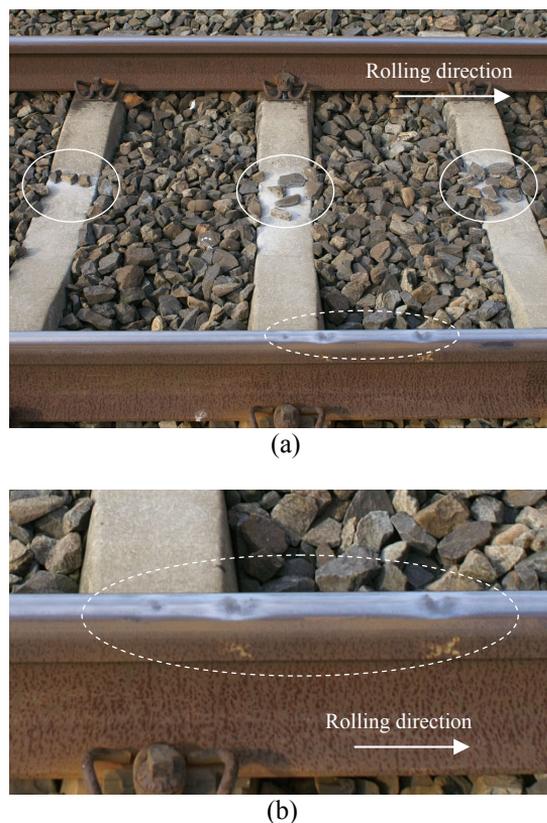
Figure 1.6. (a) White stoning caused by the relative movement between stones and sleepers in the presence of squats, (b) A zoom-in of (a).

1.2.3   Wheel–rail impact noise

Due to the vibrations excited by squats, noise emission is considerably increased in a squatted track section. This can be heard easily beside the track.

1.2.4   Increased track maintenance costs

Counter measures against squats must be employed to avoid broken rails, reduce the deterioration rate of the track, and minimize the related noise emission. According to [1.7, 1.8], an enormous amount of money is spent every year to inspect, renew, and grind rails against squats on the Japanese railway network. In the Netherlands, an annual budget of about 30 million Euros is currently allocated for squat treatment. Moreover, increased maintenance work can introduce extra disturbances to the train operation.

Rail grinding and milling can effectively remove shallow geometry deviations and material damage. However, for mature squats with deeply developed cracks, it becomes inefficient because the grinding or milling costs will be too high and the remaining rail head will be too thin. Therefore, expensive rail replacement is often inevitable if actions

are not taken in time. With each rail replacement, two new welds are introduced. This may increase the future maintenance work due to the susceptibility of welds to squatting and rail fracture.

The damages to rail pads, fastenings, sleepers and ballast lead to a further increase in the costs of maintenance.

References

[1.1]   Categorie indeling voor Squats aan de hand van voorbeelden, Prorail (infrastructure manager of Dutch railway), Version 2, March, 2004.

[1.2]   UIC Catalogue of Rail Defects, 4th edition, UIC Code 712R, February, 2000.

[1.3]   P. Clayton, M.B.P. Allery, P.J. Bolton, Surface damage phenomena in rails, in J. Kalousek, R.V. Dukkipati, G.M.L. Gladwell (Ed), Proceedings of the conference on Contact Mechanics and Wear of Rail/wheel Systems, Vancouver, British Columbia, University of Waterloo Press, July 6-9, 1982, 419–443.

[1.4]   P. Clayton, Tribological aspects of wheel–rail contact, a review of recent experimental research, Wear 191 (1996), 170–183.

[1.5]   R. Nakumura, S. Owaku and N. Enomoto, The rail shelly cracks in Japan, Rep. Railway Tech. Res. Inst. Jpn. National Railways 6 (1965), 34–44.

[1.6]   A. Ito and R. Kurihara, Shelling of rails experienced in Japanese railways, Bull. Permanent Way Sot. Jpn. 13 (1965), 17–32.

[1.7]   Y. Satoh, M. Tatsumi, K. Kasiwaya and M. Ueda, Development of anti-darkspot bainitic steel rail, QR of RTRI 40 (1990), 86–91.

[1.8]   M. Ishida, N. Abe, T. Moto, The effect of preventive grinding on rail surface shellings, QR of RTRI 39 (1998), 136–141.

[1.9]   P. Clayton and M.B.P. Allery, Metallurgical aspects of surface damage problems in Rails, Can. Metall. Q. 21 (1982), 31–46.

[1.10]   R. Deroche, J.P. Bettembourg, B. Prasil, J.P. Bertrand and C. Juckum: Rail quality and maintenance for modem railway operation, Kluwer, Dordrecht, 1993, 435–448.

[1.11]   J. Smulders, Management and research tackle rolling contact fatigue, Railway Gazette International (June) (2003), 439–442.

[1.12]   S. Kumar, Study of rail breaks: associated risks and maintenance strategies, Technical report, Lulea University of Technology, 2006: 07.

[1.13]   K. Kondo, K. Yoroizaka, Y. Sato, Cause, increase, diagnosis, counter measures and elimination of Shinkansen shelling, Wear 191 (1996), 199–203.

[1.14]   R. Dollevoet, ProRail rail defect research and management, a presentation in TU Delft 2008.

[1.15]   Z. Li, R. Dollevoet, M. Molodova and X. Zhao, Squat growth–some observations and the validation of numerical predictions, Wear 271(2011), 148–157, doi:10.1016/j.wear.2010.10.051.

[1.16]   D.F. Cannon, H. Pradier: Rail rolling contact fatigue – Research by the European Rail Research Institute, Wear 191 (1996), 1–13.

[1.17]   R. Dollevoet, Design of an anti head check profile based on stress relief, PhD Thesis, University of Twente, The Netherlands, 2010.

[1.18]   D.F. Cannon, An international cross reference of rail defects (2nd Edition), June 2003.

[1.19]   Rail Damages, the Blue Book of RailTrack, February, 2001, UK.

[1.20]   D4.1.4 Rail degradation algorithm, INNOTRACK (project no. TIP5-CT-2006-031415O) report, 2009.

[1.21]   Z. Li, M. Molodova, X. Zhao and R. Rollevoet, Squat treatment by way of minimum action based on early detection to reduce life cycle costs, Proceedings of the 2010 Joint Rail Conference, April 27-29, 2010, Urbana-Champaign, Illinois, USA.

[1.22]   K. Dang Van and M.H. Maitournam: Rolling contact in railways: modelling, simulation and damage prediction, Fatigue Fract Engng Mater Struct 26 (2003), 939–948.

[1.23]   S.L. Grassie, Rolling contact fatigue on the British railway system: treatment, Wear 258 (2005), 1310–1318.

[1.24]   D.F. Cannon, K.O. Edel, S.L. Grassie and K. Sawley, Rail defects: an overview. Fatigue Fract Engng Master Struct 26 (2003): 865–887.

[1.25]   H.H. Jenkins, J.E. Stephenson, G.A. Clayton, G.W. Morland and D. lyon, The effect of track and vehicle parameters on wheel/rail vertical dynamic forces. Railway Engineering Journal, January, (1974), 2–16.

[1.26]   K.L. Knothe and S.L. Grassie. Modelling of railway Track and Vehicle/track interaction at high frequencies. Vehicle System Dynamics 22 (1993), 209–262.

[1.27]   S.L. Grassie. Models of railway track and vehicle–track interaction at high frequencies: results of benchmark test. Vehicle system dynamics, 25 (suppl.) (1996), 243–262.

[1.28]   J.C. Zeman, J.R. Edwards, C.P.L. Barkan and D.A. Lange, Failure mode and effect analysis of concrete ties in North America. Proceedings of 9th International Heavy Haul Conference, Shanghai, China, 2009, 270–277.

2 Literature review and outline of this dissertation

In Section 2.1 a survey of existing work on squats is first presented. Then, three fields related to numerical investigations of squats, namely rolling contact mechanics, dynamics of the vehicle–track system, and fatigue and material behavior, are reviewed separately. With the three dimensional FE modeling developed in this dissertation, the three fields mentioned above are integrated. In Section 2.2, an outline of this dissertation is given.

2.1   Literature review

2.1.1   Existing investigations into squats

So far, there have been three main categories of research on squats: metallurgical research, stress and crack growth research, and practical treatment.

2.1.1.1 Metallurgical research

Clayton et al. [2.1] reviewed a metallurgical research program at British Rail on surface initiated rail problems; squats were found to be a problem of great concern on some routes in the UK. Longitudinal–vertical sections of rail specimens showed surface initiated cracks; such cracks can branch downwards, and when they reach a critical size, brittle fracture can result in a broken rail. It was found that fatigue life was reduced as contact pressure was increased, and the creep force played an important role. A difference of a factor of ten was observed in the rolling contact fatigue resistance in two tested materials. More of Clayton's squats related work can be found in [2.2, 2.3].

Marich [2.4] presented the Australian experience with squats. He linked squats with the hard and brittle 'white etching' layer which is most commonly found on infrequently ground rail. Such a layer can form due to adiabatic shear between the wheel and rail surfaces caused by micro-slip under traction. He concluded therefore that the development of squats was very similar in nature, but not in degree, to the development of wheel burns. Nevertheless, he found that this was not sufficient to explain why squats usually develop only on one rail or what particular operating/track conditions enhance their development.

2.1.1.2 Stress and crack growth research

Bold et al. [2.5] studied the growth rate of squat type shallow-angled surface-initiated rolling contact fatigue cracks. In the 1990´s, the European Rail Research Institute (ERRI) carried out an RCF research program (Cannon and Pradier [2.6]); squats were modeled by Bogdański, leading to a series of publications, e.g. [2.7, 2.8]. Bogdański et al. modeled a squat as a plane oblique semi-elliptical crack. The state of stress in the vicinity of the crack front was determined, and the values and ranges of the stress intensity factors at the crack front were calculated. By combining crack front loading histories with mixed-mode fatigue crack growth rate data, Bogdański and Brown [2.9] further analyzed the growth of squat type cracks. In [2.8], the effect of entrapped liquid was modeled.

Dang Van and Maitournam [2.10] presented, for the case of squats, calculations of stresses and strains in a rail subjected to repeated moving contacts. Stationary methods were employed. Busquet et al. [2.11] computed plastic flows in the near-surface layer as a function of traction coefficient; surface contact load distribution was based on the solution of Kalker [2.12]. More work on squats can be found in [2.13, 2.14].

2.1.1.3 Practical treatment

Practical treatments of squats have been studied in [2.15–2.17]. A twin-disc set-up was employed in [2.17] to study the influence of grinding frequency and grinding depth on squatting in Japanese high speed lines, based on which a preventive grinding policy against squats was developed.

2.1.2 Missing aspects in the existing investigations into squats

It is noticed that in the existing investigations, the dynamic effects of the contact geometry deviation on the contact force, stress, and strain are not included. Furthermore, the influences of the track system on squat initiation and growth are not considered.

A correlation analysis of the Dutch railway network [2.18] was performed in an attempt to relate squat occurrence to certain parameters of the vehicle–track interaction system and to observations of phenomena in the tracks around squats. It was identified that squats were correlated with some short wave irregularities. For example, if the analysed rail were divided into two parts, as shown in Figure 2.1, about 74% of the squats were found to occur on part I and the rest on part II. In other words, 74% of squats were found on the half rail above the sleepers. This may suggest that the stiffness and damping characteristics of the track, particularly those of the rail and the rail pad, may have played a role. Furthermore, it was found that squats occurred frequently at places of sudden stiffness changes in switches and crossings and occasionally at the ends of connecting bars of insulated rail joints (Figure 2.2).
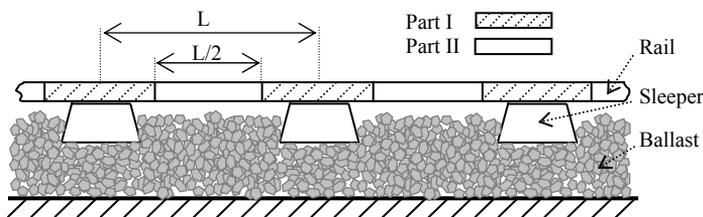
Figure 2.1. Definition of parts I and II of a rail during correlation analysis. L is a sleeper span.

| (a) | (b) |
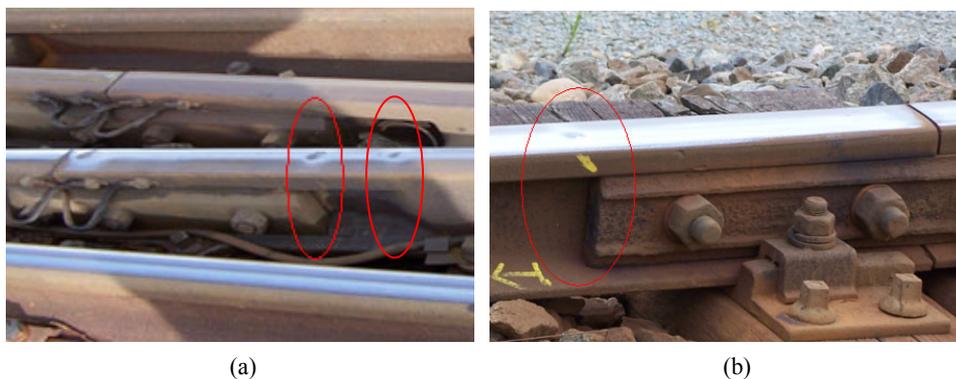
Figure 2.2 Occurrence of squats at locations where stiffness suddenly changes: (a) at switches and crossings and (b) at the end of insulated rail joints.

It is also noticed that the existing studies all focus on mature squats and do not deal with their initiation. If the initiation and growth mechanisms of squats are known, the problem may then be tackled at the root causes. This may finally lead to predictive and preventive maintenance actions based on measurement, detection, and prediction.

This dissertation therefore concentrates on squats at their initiation stage, i.e. light squats. The approach will mainly be numerical, supported by measurements and observations from field monitoring. A numerical approach is preferred because there is no existing test facility which can simulate the loading conditions of squats, namely the combined effect of rail surface geometry defects and vehicle–track structure, and before the initiation mechanism is understood, such a test rig cannot be properly designed.

On the other hand, numerical analyses can be employed to quantify relationships between the influential parameters identified in the correlation analysis and the dynamic rolling contact force, stress, and strain. A few of the most influential parameters can subsequently be further isolated. Track monitoring provides realistic inputs to the analyses and provides data for validation.

In the next three sub-sections, the three most important areas for accurate numerical simulation of vehicle–track interaction at squats are reviewed, namely contact mechanics, vehicle–track interaction, and fatigue and material behavior.

2.1.3   Contact mechanics

Contact exists in every corner of the world. Nobody can even walk without frictional contact. To tackle this extensively existent phenomenon and its related problems, contact mechanics has been developed as a branch of mechanics on the basis of continuum mechanics. The railway, characterized by the wheel–rail contact, has significantly motivated the development of contact mechanics in the past decades. Other applications of contact mechanics include the design of gears and bearings, metal forming or cutting processes, crash analysis, rolling contact of tires, and so on [2.19]. This review of contact mechanics focuses on the topics related to the wheel–rail rolling contact.

2.1.3.1   Normal and tangential contact

Traditionally, the contact problem has been divided into the normal and the tangential problems for solution. This is justified because the friction force transmitted between the elastic contact bodies of similar materials usually has a negligible influence on the shape of the contact patch and the distribution of the pressure [2.20]. The division is necessary because of the limitation of analytical methods on the one hand, and for the sake of simplicity of the solution on the other hand. In the normal problem, the contact patch and the pressure are solved for. The tangential problem is associated with friction, and solutions are sought for the distributions of surface shear stress[*] and micro-slip, as well as the related areas of adhesion and slip in the contact patch. Here, micro-slip is referred to as the relative velocity of two material particles in contact where no gross sliding occurs:

---

[*] In some of the literature, the pressure and surface shear stress are referred to as the normal and tangential tractions, respectively.

one is in the rail surface and the other in the wheel surface. Slip area is the part of the contact patch where micro-slip occurs.

The contact patch is not known *a priori* for contact problems like the wheel–rail contact, leading to a nonlinear boundary value problem. The boundary conditions, which are in the form of inequalities as shown in Equations (2.1) and (2.2) [2.12], need to be met in the solution.

For the normal problem:

$$g \geq 0, \quad p \geq 0, \quad p \cdot g = 0 \tag{2.1}$$

For the tangential problem:

$$|\boldsymbol{\tau}| \leq fp, \quad \boldsymbol{\tau} \cdot \mathbf{v} \leq 0 \tag{2.2}$$

where, g is the gap between the contact surfaces; p, $\boldsymbol{\tau}$ and $\mathbf{v}$ are the pressure, the surface shear stress and the micro-slip, respectively; and f is the coefficient of friction. Clearly, the no-penetration condition is included in Equation (2.1). Coulomb's law of friction is usually employed in Equation (2.2).

2.1.3.2   Some representative early solutions of contact problems

In the work of Hertz in 1882 [2.21], the theory of elasticity was applied to solve the normal problem of contact between bodies of quadratic surfaces with half-space assumption. Later, Mindlin [2.22] based on the Hertz normal solution analytically treated the shift problem in which contact bodies were pressed together and then shifted with respect to each other by a tangential force which was less than its friction limit.

Carter [2.23] was the first to treat wheel–rail rolling contact. He modeled the wheel and rail as two cylinders with parallel axes, in which creepage only occurs in the direction of rolling—the longitudinal direction. Vermeulen and Johnson [2.24] extended Carter's 2D theory to pure creepage without spin for half space.

2.1.3.3   Kalker's theories and derivatives

Using a boundary element (BE) approach, Kalker [2.12, 2.25] developed several numerical computer programs to solve the 3D rolling contact problem with arbitrary longitudinal and lateral creepages, as well as spin. The Boussinesq–Cerruti solution [2.26] of elastic half space is employed for the fundamental solution (referred to as the influence function by Kalker). This approach has been implemented in the computer code CONTACT, with which problems of arbitrary planar contact patch can be treated. Currently, CONTACT is extensively employed to obtain detailed solutions for wheel–rail contact when the tangential problem has to be treated, with the assumptions of half space, linear elasticity, and steady state rolling.

Kalker also developed two derivatives from this "exact" solution [2.12]. The first is the linear theory [2.27], which is suitable for railway vehicle dynamics with small creepages. The second is the simplified theory [2.12] developed on the basis of the linear theory and on the Winkler foundation, mostly known by its implementation in the

computer program FASTSIM [2.28]. It is now used extensively for vehicle system dynamics, corrugation studies etc. where large creepages and moderate spin [2.29], varying friction coefficient [2.30], or transient rolling due to varying creepages and normal force [2.31] need to be taken into account. Another derivative was made by Shen, Hedrick, and Elkins [2.32], which is suitable for vehicle dynamics with larger creepages where the linear theory is not applicable.

2.1.3.4 Some other boundary element method solutions

With the fundamental solution obtained from finite element (FE) calculation, Li [2.33] extended Kalker's work from half space to quasi-quarter spaces and from constant geometrical spin to varying geometrical spin. Curved contact geometry like the conformal contact between wheel flange root and rail gage corner, especially when they are worn, can be dealt with. This solution has been implemented in the computer program called WEAR. In order to solve continuously tens of thousands of contact problems in a wear simulation, automatic and adaptive meshing, as well as robust convergence, is achieved. This solution has been successfully applied to analysis of head check initiation and the design of an anti-head-check rail profile [2.34].

Based on Kalker's formulation [2.12], González and Abascal [2.35] proposed a solution method which couples BE and FE to treat 3D frictional rolling in steady state. In their numerical example of wheel–rail rolling, the rail is considered as being rigid and the condition that slip is opposite to tangential traction is not well satisfied. In general, in their numerical examples slip is present at the leading edge where no slip is expected.

BE methods have been proven to be well suited for linear elastic problems. With such a method, contact bodies need only be discretized on the boundaries for calculation. This leads to their main advantage over the FE method: much lower computation costs. Nevertheless, due to the lack of mesh in the interior of contact bodies, the extension of the BE method to non-linear problems, including material and geometric non-linearities, is not well-developed [2.36].

2.1.3.5 FE solutions

With the FE method, contact geometry can be simulated more realistically than with the BE method. Furthermore, material non-linearity can be included in an FE model by employing a proper constitutive equation; dynamic behavior related to the contact can also be taken into account, for instance, by performing a transient analysis. Solutions of normal contact problems in statics can be found in [2.37–2.41]. In [2.39], convective dissipation due to friction is also discussed for the case of gross sliding.

Rolling contact needs special formulation and a special algorithmic approach. Padovan [2.42] and Nakajima and Padovan [2.43] proposed the moving Lagrange observer approach and applied it to impact contact between a deformable tire and a rigid obstacle in two dimensions. Disadvantages of this approach arise from the additive decomposition of motion with restrictions to kinematically linear problems etc.

A more general kinematic formulation of rolling contact, namely the Arbitrary Lagrangian Eulerian (ALE) method, has been developed by Oden and Lin [2.44], Bass

[2.45], Faria et al. [2.46], Nackenhorst [2.47], Ziefle and Nackenhorst [2.48], and Hu and Wriggers [2.49]. Currently, ALE is well established for tire–road contact where the tire is deformable and the road could be considered as being rigid. It offers a time independent formulation and is advantageous for stationary rolling [2.47].

For wheel–rail contact, the deformability of both the wheel and the rail must be considered simultaneously due to their similar modulus. Furthermore, the contact surfaces may not be simulated as planes, especially for the case of squats. The ALE approach was applied to wheel–rail rolling in [2.50, 2.51]. However, remarkable difference still exists compared to well established results. One problem is related to the regularization of stick-slip behavior. Regularization is employed to avoid non-differentiability of Coulomb's law at the onset of slip. The regularized constitutive interface laws physically have the drawback that they only describe the stick-slip motion in an approximate fashion. For a value of the regularization parameter that is too large, such a model may not be able to predict real stick-slip motions [2.19]. Future work on a reliable, stable, and efficient numerical algorithm is needed for the solution of wheel–rail frictional rolling by ALE [2.51, 2.52].

Discussions on frictional wheel–rail rolling in two dimensions in statics can be found in [2.53, 2.54].

2.1.3.6   Measurements of contact loads and stresses

Currently, wheel–rail contact forces can be measured by two approaches, i.e. on-board measurement [2.55–2.57] and trackside measurement [2.58, 2.59]. The measuring principle is to deduce the contact forces from the recorded deformation of a measurement wheel or rail instrumented with strain gauges or other equipment. Calibration with specially designed tests is needed [2.55–2.59].

No experimental technique is currently available for measuring the size and shape of the contact patch and the stress distributions in it, especially for high frequency dynamic contact. In some attempts to measure the static contact stresses [2.60–2.63], in which only the pressure could be assessed, analytical or numerical contact solutions were employed for calibration. In contrast, measurement of contact stresses in the tire–road interface is successful due to the larger contact patches and much lower stress levels[*] [2.64].

For these reasons, methods for the solution of the wheel–rail contact problem become more important.

---

[*] For the tire–road contact patches measured in [2.64], their typical area is over $70,000 \text{ mm}^2$ and the maximum pressure is less than 2 MPa. A typical wheel–rail contact patch is about 150 mm^2 in area with the maximum pressure over 1 GPa (see the results shown later in Chapter 3).

2.1.4   Dynamics of the vehicle–track system

As a vehicle runs along a track dynamic forces arise. Until recently, the dynamic interaction between the vehicle and the track has mainly been investigated with multi-body models for the vehicles and beam models for tracks, coupled through simplified wheel–rail contact, see Sections 2.1.4.1–2.1.4.4. The recent trend of modeling wheels and rails as continua with finite elements is reviewed in Section 2.1.4.5.

2.1.4.1   Vehicle modeling

Most rail vehicles consist of a car body and two bogies with two levels of suspension. The suspensions are commonly designed to ensure that rigid body modes of the bogie frame and the car body occur below about 10 Hz [2.65]. For example, the bounce frequency of vehicle bodies is reported to be 0.9–1.4 Hz in [2.66, 2.67]. This means that the sprung masses of vehicles are well isolated from the high frequency vibrations of wheelsets and can be simplified as lumped masses when modeling the high frequency vehicle–track interaction.

In many dynamics models, wheels are approximated with lumped masses [2.68–2.77]. However, it is anticipated that the flexibility of wheels should considerably influence the high frequency vehicle–track interaction. This has been confirmed in [2.78–2.80], in which the structural flexibility of a wheel was introduced by considering its eigenmodes derived from FE solutions. In [2.80], it is even found that the rotation of the wheelset may also play an important role in the high frequency vehicle–track dynamics under certain conditions. The lowest eigen-frequency of a wheel is typically above 200 Hz [2.66].

2.1.4.2   Track modeling

Ballasted track consists of rails, railpads and fastening, sleepers, ballast and subgrade. Depending on the objectives of investigations, the support of rails has been modeled as being continuous or discrete and of one-layer or multi-layer structure [2.65] (see Figure 2.3). For investigations of high frequency vehicle–track interaction, for corrugation studies for instance, multi-layer track models with discretely supported rails are usually employed [2.69, 2.73, 2.75, 2.77]. Three vertical track resonances have mainly been considered: track vibration on ballast bed, rail bouncing on railpads, and the so-called pinned-pinned vibration, in ascending order of frequency. According to [2.66], the frequencies of these resonances generally cover a band between several tens of Hz and about 1000 Hz. The last two resonances are close in frequency to the vibrations related to corrugation.
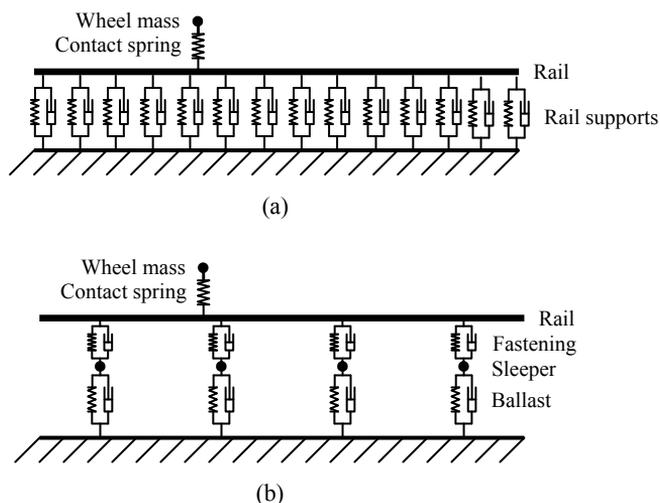
Figure 2.3. The widely employed vehicle–track interaction model: (a) one-layer track model with continuous support; (b) Two-layer track model with discrete support. A Euler–Bernoulli beam or Timoshenko beam is used to represent the rail [2.65].

The rail is often modelled as an Euler–Bernoulli or Timoshenko beam [2.69, 2.81, 2.75]. More advanced models allow relative movement between rail head, web, and foot [2.82, 2.83]. Their application is limited up to 1500 Hz [2.82, 2.84, 2.85]. When vibration modes are calculated from real cross-sections of the rail [2.86, 2.87], the models can fit track dynamic measurements up to 1500 Hz as well. Between 1500 and 3000 Hz, the measured frequency distribution could not be properly reproduced except for the so-called second order pin-pin anti-resonance (2700–3000 Hz) [2.86].

For discrete support, railpads and fastening are usually modelled with linear springs and viscous dampers in parallel [2.75, 2.88]. When the rail is continuously supported, the support is actually represented as a viscoelastic layer [2.69, 2.82]. Modelling of sleepers by rigid masses or beams is sufficient for the biblock sleeper [2.86, 2.75]. For the monoblock sleeper, improvement is still needed if vibrations between 450 and 1000 Hz are important [2.84, 2.89]. Modelling ballast as half-space or as a viscoelastic foundation has a significant influence on vibrations below 250 Hz [2.90]. The importance of high frequency vibration on the ballast bed can be seen from the increased need for ballast maintenance at corrugation sites, although ballast deterioration is not wholly a problem of high frequency loading [2.65].

2.1.4.3 Modeling of the wheel–rail contact

For vehicle–track interaction, simplified contact models are usually used to limit the computational costs. A Hertzian spring is often employed for the normal contact [2.65, 2.68–2.79]. For tangential force, Kalker's linear theory, the Shen–Hedrick–Elkins theory, and FASTSIM are frequently employed [2.91, 2.66]. A problem with this approach is that the point contact of the spring cannot include the effect of the finite size of an actual contact patch. This becomes important when the effects of contact geometry and contact patch on stress, strain, and wear are not negligible, such as at corrugation. To replace the

single spring, Winkler bedding was introduced in [2.92]. In more advanced models, the three dimensional solution of Kalker [2.12] has recently been incorporated in corrugation calculations [2.93]. Viscosity and frequency dependent creep coefficients are considered in [2.77].

2.1.4.4   Models of high frequency vehicle–track interaction

Combining the sub-models of vehicle, track and contact, a variety of models of high frequency vehicle–track interaction have been developed for investigations of rolling contact noise, corrugation, out-of-roundness of wheels, and impact at fish-plated /insulated rail joints and turnout frogs, e.g. [2.29, 2.71, 2.75, 2.77, 2.94–2.97].

According to [2.96–2.97], the calculated dynamic forces vary significantly across the models employed. This may be due to different simplifications and limitations of the sub-models.

2.1.4.5   FE modeling with a mesh of solid elements

3D FE models of solid elements with explicit time integration schemes have been employed in the literature to treat the wheel–rail interaction in the normal direction at joints or crossings [2.98–2.100]. The track was simulated by a one-layer model with discretely supported rails in [2.98, 2.99], while rigid support of rails was assumed in [2.100].

2.1.5   Fatigue and material behavior

2.1.5.1   RCF and fatigue

Fatigue is defined in [2.101] as the degradation of mechanical properties leading to the failure of a material or a component under cyclic loading. If the degradation is initiated at locations where the load is applied through rolling contact, or in the vicinity thereof, it is referred to as rolling contact fatigue (RCF).

Depending on the location of its initiation, RCF can roughly be classified into two categories: subsurface and surface initiated RCF. The occurrence of subsurface initiated RCF has been greatly reduced in the past decades due to cleaner steel made by improved manufacturing techniques, and nowadays the surface initiated RCF (referred to as RCF for short in this dissertation) has instead become preponderant [2.102–2.104]. For rails, head checks and squats are the two main types of RCF, as specified in Chapter 1.

2.1.5.2   Material behavior under cyclic loading

Ignoring wear, the material behavior under cyclic loading can be classified into four categories depending on the level of the applied load, namely elastic, elastic shakedown, plastic shakedown (or cyclic plasticity), and ratcheting [2.105]. When the load level is lower than the elastic limit, purely elastic deformation will take place at all times. If the elastic limit is exceeded but the load is within the elastic shakedown limit, yield may occur in the first loading cycles, and deformation will soon become purely elastic once a

steady state is achieved. A design normally ensures that the load level is lower than the elastic shakedown limit to maximize the service life of a material. When cyclic plasticity or ratcheting occurs, plastic deformation will be present in each loading cycle. This can easily lead to material failure through the damage mechanism of low cycle fatigue or ratcheting [2.106]. The essence of the shakedown phenomenon is the residual stresses and work hardening accompanied by plastic deformation [2.105].

2.1.5.3   Shakedown map

The shakedown map is widely used to predict rail RCF [2.10, 2.106–2.111]. It covers all the material behavior mentioned in Section 2.1.4.2 and can consider [2.105] or ignore [2.26] work hardening. In [2.112], the shakedown map was further extended to include the influence of thermal stresses.

Based on the shakedown map, RCF occurrence can be judged by the levels of the contact loads (both normal and tangential) under the assumption of Hertzian contact. However, detailed information such as the level of plastic deformation corresponding to the crack initiation and the residual life of a material cannot be obtained. Thus, the shakedown map may only be used for qualitative analyses. Its application to squats is further limited by the assumption of statics behind it.

One important phenomenon that can be seen in the shakedown map is that first material yield (corresponding to the state where only one particle yields under contact loads) gradually moves from subsurface to surface with increasing tangential force. Under full sliding conditions, the critical state occurs as the friction coefficient approximately equals 0.3 [2.26]. As a consequence, the shakedown limit (in terms of pressure level) considerably reduces with the increase of the tangential force.

In order to get a better insight into squatting phenomena, different stages of a fatigue fracture, i.e. micro-crack initiation, macro-crack propagation, and final failure, should be taken into consideration. In the next section, fatigue criteria for crack initiation will be reviewed, considering the importance of the early stages of squatting in efficient track maintenance. Investigations into the propagation of macro-cracks contained in squats have been reviewed in Section 2.1.1.

2.1.5.4   Fatigue criteria and stress analyses

According to [2.106], field observations and laboratory investigations of cracks initiated in rails have shown that high-cycle fatigue is not the dominant fatigue regime. Low-cycle fatigue and ratcheting, which are considered to be independent of each other, compete to determine the occurrence of cracks in pearlitic rail steels [2.106]. In contrast, ratcheting and fatigue criteria were combined in [2.113] to predict the initiation of fatigue cracks for Hadfield steel and Bainitic alloy[*].

Considering the different material behavior mentioned in Section 2.1.4.2, various damage models have been developed to predict the initiation of fatigue cracks. In [2.106],

---

[*] Hadfield steel is used in diamond crossings and frogs in the railway industry, while Bainitic alloy is a candidate for such applications [2.113].

two different criteria for low-cycle fatigue, namely the Coffin–Manson relation and the Smith–Watson–Topper relation, were introduced for rails in plastic shakedown state, as well as a criterion for ratcheting failure. Similar fatigue criteria were also employed in other publications such as [2.114, 2.10]. Note that head checks and squats are not distinguished from each other in these damage models. A broader and more general review of damage models on contact fatigue initiation developed before the 1990s can be found in [2.115].

For the above-mentioned damage models, their common basis is a stress and/or strain estimation. Efforts have been made in the literature to calculate the stresses/strains of rails. For example, the cyclic plastic deformation of continua was studied in two dimensions [2.116–2.118] and in three dimensions [2.119]. Residual stresses, as a direct consequence of plastic deformation, are normally presented in these publications. However, the accuracy of these calculations is greatly limited by the fact that the phenomenon of contact is omitted for simplicity. The rolling process, if included in the calculation, is usually simulated with the translational moving of contact load (such as Hertzian pressure distribution).

To obtain more accurate stress estimates, the wheel–rail contact, as the key to stress analysis of squats, has to be considered. According to the state-of-the-art of contact mechanics given in Section 2.1.3, however, the tangential contact problem between the wheel and the rail cannot be solved by the existing approaches. This is one of the intentions of the 3D transient FE model developed in this dissertation, by which the frictional wheel–rail rolling contact at rail surface defects is solved in the time domain. The improved accuracy in stress calculation will build a basis for further investigation of squat initiation.

2.2   Outline of the work of this dissertation

As discussed, this dissertation will concentrate on squats at their initiation stage, i.e. light squats. The approach will mainly be numerical, supported by measurements and observations from field monitoring.

The literature review and correlation analysis have revealed that rail surface geometry defects, high frequency wheel–rail interaction, plasticity, friction, and tangential contact stress are important factors. It is discussed that an FE based approach is desirable for the complex contact geometry and for the high frequency vibrations which may involve the continua of the wheel and rail, as well as the structure of the vehicle–track system. It has also been shown that a proper FE solution for such needs is not yet available.

The dissertation is therefore composed of the following chapters, as shown in Figure 2.4.

In Chapter 3, an FE solution procedure is developed using an explicit integration scheme to solve the frictional wheel–rail rolling contact. It is validated for contact between smooth surfaces in elasticity and in statics where Hertzian and Kalker's solutions are available. Afterwards, rolling in elasto-plasticity is discussed.

In Chapter 4, the FE solution of rolling contact is integrated into the vehicle–track interaction system. Rail surface geometry defects are introduced to examine the dynamic contact force, stress and strain, and parameters of the system are varied to study their effects. The explicit time integration scheme employed renders it intrinsically suitable for analyses of high frequency dynamic problems. Validations of the findings are also presented. The model is also employed in Chapter 4 to calculate the dynamic forces at the discrete supports of a rail in the presence of squats.

In Chapter 5, a critical size for defects to grow into squats is determined on the basis of dynamic stresses. This is useful for separating trivial defects from those growing into squats.

In Chapter 6, conclusions are drawn and recommendations are made for further research.

It is assumed, based on correlation between squats and indentations, corrugations etc., that squats grow from small rail top defects – light squats, and there are no cracks present in light squats. In the numerical models, therefore, no cracks are considered. Material is assumed to be homogeneous and isotropic. Linear elastic and bi-linear elasto-plastic material models are employed, where appropriate. Furthermore, no cyclic loading is considered. The effect of pre-existing residual stress is not included in the modeling.
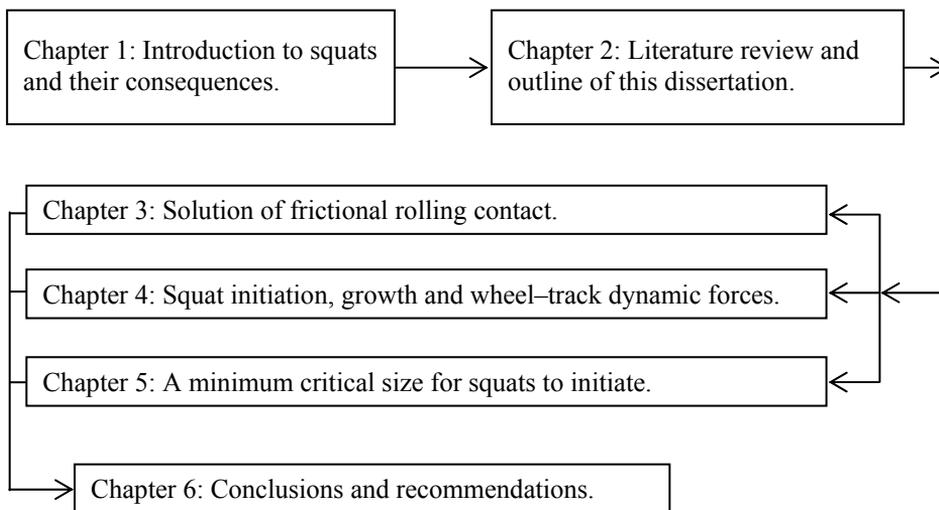
Figure 2.4. The structure of this dissertation.

References

[2.1]   P. Clayton, M.B.P. Allery, and P.J. Bolton, Surface damage phenomena in rails, in J. Kalousek R.V. Dukkipati and G.M.L. Gladwell Ed. Proceedings of the conference on Contact Mechanics and Wear of Rail/wheel Systems, Vancouver, British Columbia, July 6-9, 1982, 419–443.

[2.2]   P. Clayton and M.B.P. Allery, Metallurgical aspects of surface damage problems in Rails. Can. Metall. Q. 21 (1982), 31–46.

[2.3]   P. Clayton and D.N. Hill (1987), Rolling contact fatigue of a rail steel, G.M.L. Gladwell, H. Ghonem and J. Kalousek Ed. Proceedings of the conference on Contact Mechanics and Wear of Rail/wheel Systems II, University of Rhode Island, Kingston, R.I., July 8 - 11, 1986, 361–378.

[2.4]   S. Marich, Practical/realistic implementation of wheel/rail contact technologies – the Australian experience. Proceedings of the 7th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems (CM2006), Brisbane, Australia, September 24-26, 2006, 3–21.

[2.5]   P.E. Bold and M.W. Brown, Shear mode crack growth and rolling contact fatigue, Wear, 144 (1991), 307–317.

[2.6]   D.F. Cannon and H. Pradier, Rail rolling contact fatigue – Research by the European Rail Research Institute, Wear 191 (1996), 1–13.

[2.7]   S. Bogdanski, M. Olzak, J. Stupnicki, Numerical stress analysis of rail rolling contact fatigue cracks, Wear, 191 (1996), 14–24.

[2.8]   S. Bogdański and R. Lewicki, 3D Model of Entrapped Liquid in Rolling Contact Fatigue Cracks in Rails, Wear 265 (2008), 1356–1362.

[2.9]   S. Bogdanski and M.W. Brown, Modeling the three-dimensional behavior of shallow rolling contact fatigue cracks in rails, Wear 253 (2002), 17–25.

[2.10]   K. Dang Van, M.H. Maitournam, On some recent trends in modeling of contact fatigue and wear in rail. Wear 253 (2002), 219–227.

[2.11]   M. Busquet, L. Baillet, C. Bordreuil, Y. Berthier, 3D finite element investigation on the plastic flows of rolling contacts—correlation with railhead microstructure observations, Wear 258 (2005), 1071–1080.

[2.12]   J.J. Kalker, Three-dimensional elastic bodies in rolling contact, Kluwer Academic Publishers, Dordrecht, 1990, 59–78, 100–117, 137–184.

[2.13]   R. Smallwood, J. C. Sinclair and K. J. Sawley, An optimization technique to minimize rail contact stresses, Wear 144 (1991), 373–384.

[2.14]   K. Dang Van and M.H. Maitournam: Rolling contact in railways: modelling, simulation and damage prediction, Fatigue Fract Engng Mater Struct 26 (2003), 939–948.

[2.15]   K. Kondo, K. Yoroizaka and Y. Sato, Cause, increase, diagnosis, countermeasures and elimination of Shinkansen shelling Wear 191 (1996), 199–203.

[2.16]   M. Ishida, M. Akama, K. Kashiwaya and A. Kapoor: The current status of theory and practice on rail integrity in Japanese railways-rolling contact fatigue and corrugations, Fatigue Fract Engng Mater Struct 26 (2003), 909–919.

[2.17]   M. Ishida, N. Abe, and T. Moto, The effect of preventive grinding on rail surface shellings, QR of RTRI 39 (3) (1998), 136–141,.

[2.18]   Z.L. Li, Correlation Analysis of Squats, Internal Project Report, Railway Engineering Group, TU Delft, March 5, 2006.

[2.19]   P. Wriggers, Computational contact mechanics (2nd Ed.), Springer, Heidelberg, 2006.

[2.20]   H. Poritsky, N.Y. Schenectady, Stress and deflections of cylindrical bodies in contact with application to contact of gears and of locomotive wheels, Journal of Applied Mechanics 17 (1950), 191–201.

[2.21]   H. Hertz, Über die berührung fester elastischer körper, J.reine und angewandte Mathematik 92 (1882), 156–171.

[2.22]   R.D. Mindlin, Compliance of elastic bodies in contact, Journal of applied Mechanics 16(1949), 259–268.

[2.23]   F.W. Carter, On the action of a locomotive driving wheel, Pro. R. Soc., London, A. 112 (1926), 151–157.

[2.24]   P.J. Vermeulen, K.L. Johnson, Contact of non-spherical elastic bodies transmitting tangential forces, J. Appl. Mech. 31 (1964), 338–340.

[2.25]   J.J. Kalker, wheel–rail rolling contact theory, Wear (1991), 243–261.

[2.26]   K.L. Johnson, Contact Mechanics, Cambridge University Press, London, 1985.

[2.27]   J.J. Kalker, On the rolling contact of two elastic bodies in the presence of dry friction. PhD dissertation, TU Delft, 1967.

[2.28]   J.J. Kalker, A fast algorithm for the simplified theory of rolling contact, Vehicle system dynamics 11 (1982), 1–13.

[2.29]   O. Polach, Creep forces in simulations of traction vehicles running on adhesion limit, Wear 258 (2005), 992–1000.

[2.30]   O. Arias-Cuevas, Z. Li, R.I. Popovici and D.J. Schipper Simulation of curving behaviour under high traction in lubricated wheel–rail contacts, Vehicle System Dynamics 48 (Suppl.) (2010), 299–316.

[2.31]   Z. Shen and Z. Li (1996), A fast non-steady state creep force model based on the simplified theory, Wear 191(1996), 242–244.

[2.32]   Z.Y. Shen, J.K. Hedrick and J.A. Elkins, A comparison of alternative creep force models for rail vehicle dynamic analysis, Vehicle System Dynamics 12 (1983), 79–83.

[2.33]   Z. Li, Wheel–rail rolling contact and its application to wear simulation, PhD dissertation, TU Delft, the Netherlands, 2002.

[2.34]   R. Dollevoet, Design of an anti head check profile based on stress relief, PhD Thesis, Z. Li as co-promotor, University of Twente, The Netherlands, 2010.

[2.35]   J. A. González, R. Abascal, A 3D FEM_BEM rolling contact formulation for unstructured meshes, International Journal of Solids and Structures 47 (2010), 330–353.

[2.36]   H. Gun, Boundary element method for prediction of hardness of dentin from punch-nano-indentation test, Computer method and programs in biomedicine 86 (2007), 227–231.

[2.37]   M. Wiest, E. Kassa, W. Daves, J.C.O. Nielsen, and H. Ossberger, Assessment of methods for calculating contact pressure in wheel–rail/switch contact. Wear 265 (2008), 1439–1445.

[2.38]   M. Pau, F. Aymerich, and F. Ginesu, Distribution of contact pressure in wheel–rail contact area, Wear 253 (2002), 265–274.

[2.39]   W. Yan, F.D. Fischer, Applicability of the Hertz contact theory to rail-wheel contact problem, Archive of applied Mechanics 70 (2000), 255–268.

[2.40]   A. Sladkowski, M. Sitarz, Analysis of wheel–rail interaction using FE software, Wear 258 (2005), 1217–1223.

[2.41]   T. Telliskivi, U. Olofsson, Contact mechanics analysis of measured wheel–rail profiles using the finite element method. Proc Instn Mech Engrs, Part F, 215 (2001) 65–72.

[2.42]   J. Padovan, Finite element analysis of steady and transiently moving/rolling nonlinear viscoelastic structure - I. theory. Computers & structures 27(2) (1987) 249–257.

[2.43]   Y. Nakajima and J. Padovan, Finite element analysis of steady and transiently moving/rolling nonlinear viscoelastic structure - III. impact/contact simulation. Computers & structures, 27(2) (1987), 275–286.

[2.44]   J.T. Oden, T.L. Lin, On the general rolling contact problem for finite deformations of viscoelastic cylinder, Computer methods in applied mechanics and engineering 57 (1986), 297–376.

[2.45]   J.M. Bass, Three-dimensional finite deformation, rolling contact of a hyperelastic cylinder formulation of the problem and computational results, Computers & Structures 26 (1987), 991–1004.

[2.46]   L.O. Faria, J.T. Oden, B. Yavari, W.W. Tworzydlo, J.M. Bass and E.B. Becker, Tire modeling by finite elements, Tire science and technology, 20(1) (1992), 33–56.

[2.47]   Nackenhorst, On the finite element analysis of steady state rolling contact. In M.H. Aliabadi and C.A. Brebbia (editors), Contact Mechanics - Computational Techniques, Computational Mechanics Publication, Southampton, Boston, 1993, 53–60.

[2.48]   M. Ziefle, U. Nackenhorst, Numerical techniques for rolling rubber wheels: treatment of inelastic material properties and frictional contact, Comput. Mech 42 (2008), 337–356.

[2.49]   G. Hu and P. Wriggers, On the adaptive finite element methods of steady-state rolling contact for hyperelasticity in finite deformations, Comput. Methods Appl. Mech. Engrg. 191 (2002), 1333 – 1348.

[2.50]   S. Damme, U. Nackenhorst, A. Wetzel, and B. Zastrau, On the numerical analysis of the wheel–rail system in rolling contact, in: K. Popp, W. Schielen (Eds.), System Dynamics and Long-term Behavior of Railway Vehicles, Track and Subgrade, Springer, Berlin, Heidelberg, 2003, 155–174.

[2.51]   U. Nackenhorst, The ALE-formulation of bodies in rolling contact – theoretical foundations and finite element approach, Comput. Methods Appl. Mech. Engrg, 193 (2004), 4299–4322.

[2.52]   M. Brinkmeier, U. Nackenhorst and M. Ziefle, Modern approaches on rolling contact, in P. Wriggers and T. A. Laursen (Eds.), Computational contact mechanics, CISM courses and lectures, vol. 498, Springer, 2007.

[2.53]   M. Bijak-zochowski, P. Marek, Residual stress in some elasto-plastic problems of rolling contact with friction, J. Mech. Sci. 39 (1997) 15–32.

[2.54]   Y.C. Chen, L. Chen. Effects of insulated rail joint on the wheel/rail contact stresses under the condition of partical slip, Wear 260 (2006), 1267–1273.

[2.55]   A. Matsumoto, Y. Sato, H. Ohno, M. Tomeoka, K. Matsumoto, J. Kurihara, T. Ogina, M. Tanimoto, Y. Kishimoto, Y. Sato, T. Nakai, A new measuring method of wheel–rail contact forces and related considerations, Wear, 265 (2008), 1518–1525.

[2.56]   E. Kassa, J.C.O. Nielsen, Dynamic interaction between train and railway turnout: full-scale field test and validation of simulation models, Vehicle System Dynamics, 46(suppl.) (2008), 521–534.

[2.57]   P. Gullers, L. Andersson and R. Lunden, High-frequency vertical wheel–rail contact forces – field measurements and influence of track irregularities, Wear 265 (2008), 1472–1478.

[2.58]   A. Johansson, J.C.O. Nielsen, Out-of-round railway wheels – wheel–rail contact forces and track response derived from field tests and numerical simulations, J. Rail and Rapid Transit 217(2003), 135–146.

[2.59]   T. Roeleveld, Measurement of wheel–rail contact forces, Symposium of Advances in Contact Mechanics: a tribute to Prof. J.J. Kalker, 22–24 October 2008, Delft, The Netherlands, 45–46.

[2.60]   Pau, M., Aymerich, F., and Ginesu, F., Ultrasonic Measurement of Nominal Contact Area and Contact Pressure in a Wheel/Rail System, Proc. Inst. Mech. Eng., F J. Rail Rapid Transit 214 (2000), 231–243.

[2.61]   R.S. Dwyer-Joyce, B.W. Drinkwater, In situ measurement of contact area and pressure distribution in machine elements, Triboloby Letters 14(2003), 41–52.

[2.62]   M.B. Marshall, R. Lewis, R.S. Dwyer-Joyce, Experimental Characterization of wheel rail contact patch evolution, Journal of Tribology 128 (2006), 493–504.

[2.63]   R.S. Dwyer-Joyce, R. Lewis, A comparison of three methods for the measurement of wheel rail contact, Proceedings of IJTC2008, October 20–22, 2008, Miami, Florida, USA.

[2.64]   M. de Beer, C. Fisher and F.J. Jooste, Determination of pneumatic tyre/pavement interface contact stresses under moving loads and some effects on pavements with thin asphalt surfacing layers, 8th International Conference on Asphalt Pavements, Washington, 10–14 August, 1997, 179–227.

[2.65]   K.L. Knothe, S.L. Grassie, Modelling of railway Track and Vehicle/track interaction at high frequencies, Vehicle System Dynamics 22 (1993), 209–262.

[2.66]   Handbook of railway vehicle dynamics, edited by S. Iwnicki, CRC press Taylor & Francis, Boca Raton, FL, USA, 2006, 73–74, 96–107, 147–149, 285–290.

[2.67]   N. Shieh, C. Lin, Y. Lin, and K. Liang, Optimal design for passive suspension of a light rail vehicle using constrained multi-objective evolutionary search, Journal of Sound and Vibration 285 (2005), 407–424.

[2.68]   H. Jenkins, J. Stephenson, G. Clayton, G. Morland and D. lyon, The effect of track and vehicle parameters on wheel/rail vertical dynamic forces, Railway Engineering Journal (January) (1974), 2–16.

[2.69]   S.L. Grassie, R.W. Gregory, D. Harrison, and K.L. Johnson, The dynamic response of railway track to high frequency vertical excitation, Journal Mechanical Engineering Science 24 (1982), 77–90.

[2.70]   C. Andersson, T Dahlberg, Wheel/rail impacts at railway turnout crossing, Journal of Rail and Rapid Transit 212 (1998), 123–134.

[2.71]   T.X. Wu, D.J. Thompson, The effects of track non-linearity on wheel/rail impact, Journal of Rail and Rapid Transit 218 (2004), 1–15.

[2.72]   M.X.D. Li, E.G. Berggren, and M. Berg, Assessment of vertical track geometry quality based on simulations of dynamic track-vehicle interaction, Proc. IMechE Part F: J. Rail and Rapid Transit 223 (2009), 131–139.

[2.73]   G. Xie, S.D. Iwnicki, Simulation of wear on a rough rail using a time-domain wheel–track interaction model. Wear 265 (2008), 1572–1583.

[2.74]   M.J.M.M. Steenbergen, Modeling of wheels and rail discontinuities in dynamic wheel–rail contact analysis, Vehicle System Dynamics 44 (2006), 763–787.

[2.75]   M. Hiensch, J.C.O. Nielsen, and E. Verheijen, Rail corrugation in The Netherlands— measurements and simulations, Wear, 253 (2002), 140–149.

[2.76]   S. Iwnicki, Y. Bezin, G. Xie and E. Kassa, advances in vehicle–track interaction tools, Railway Gazette International 165 (2009) (September), 47–52.

[2.77]   K.L. Knothe, A Groß-thebing, Short wavelength rail corrugation and non-steady-state contact mechanics, Vehicle system dynamics 46 (2008), 49–66.

[2.78]   N. Chaar, M. Berg, Simulation of vehicle–track interaction with flexible wheelsets, moving track models and field tests, Vehicle System Dynamics 44 (2006), 921–931.

[2.79]   B. Ripke, K Knothe, Simulation of high frequency vehicle–track interactions, Vehicle System Dynamics 24 (1995), 72–85.

[2.80]   L. Baeza, J. Fayos, A. Roda, R. Insa, High frequency railway vehicle–track dynamics through flexible rotating wheelsets, Vehicle System Dynamics 46 (2008), 647–662.

[2.81]   M. Fermer, J.C.O. Nielsen, Vertical interaction between train and track with soft and stiff railpads - full-scale experiments and theory, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 209 (1) (1995), 39–47.

[2.82]   D.J. Thompson, Wheel–rail noise generation, part iii: Rail vibration, Journal of Sound and Vibration 161 (3) (1993), 421–446.

[2.83]   T.X. Wu, D.J. Thompson, A double Timoshenko beam model for vertical vibration analysis of railway track at high frequencies, Journal of Sound and Vibration 224 (2) (1999), 329–348.

[2.84]   N. Vincent, D.J. Thompson, Track dynamic behaviour at high frequencies. part 2: experimental results and comparisons with theory, Vehicle System Dynamics 24 (Suppl) (1995), 100–114.

[2.85] A. Bhaskar, Wheel–rail dynamics with closely conformal contact part 1 : Dynamic modelling and stability analysis, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 211 (1) (1997), 11–24.

[2.86] L. Gry, Dynamic modelling of railway track based on wave propagation, Journal of Sound and Vibration 195 (3) (1996), 477–505.

[2.87] K. Knothe, Z. Strzyzakowski, and K. Willner, Rail vibrations in the high frequency range, Journal of Sound and Vibration 169 (1) (1994), 111–123.

[2.88] A. de Man, Dynatrack: a survey of dynamic railway track properties and their quality, PhD thesis, TU Delft (2002).

[2.89] S. Hammarlund, A. Jahlenius, and T. Dahlberg, Goose hill measurements confirm x2000's low dynamic forces, Railway Gazette International 150 (7) (1994), 439.

[2.90] K. Knothe, Y. Wu, Receptance behaviour of railway track and subgrade, Archive of Applied Mechanics 68 (7-8) (1998), 457–470.

[2.91] K. Knothe, R. Wille and B. W. Zastrau. Advanced contact mechanics - Road and Rail, Vehicle system dynamics 35 (2001), 361–407.

[2.92] A. Pieringer, W. Kropp and D.J. Thompson, Investigation of the dynamic contact filter effect in vertical wheel/rail interaction using a 2D and a 3D non-Hertzian contact model, Wear 271 (2011), 328 – 338.

[2.93] G. Xie and S.D. Iwnicki, Calculation of wear on a corrugated rail using a three-dimensional contact model, Wear 265 (2008), 1238 – 1248.

[2.94] D.J. Thompson, The Influence of the Contact Zone on the Excitation of Wheel/Rail noise, Journal of Sound and Vibration 267 (2003), 523–535.

[2.95] J.C.O. Nielsen, R. Lundén, A. Johansson and T. Vernersson, Train–Track Interaction and Mechanisms of Irregular Wear on Wheel and Rail Surfaces, Vehicle system dynamics 40 (2003) (1–3), 3–54

[2.96] S.L. Grassie, Models of railway track and vehicle–track interaction at high frequencies: results of benchmark test, Vehicle system dynamics 25 (suppl.) (1996), 243–262.

[2.97] J. Leong, M. Murray and D. Steffes, Examination of railway track dynamic models capabilities against measured field data. IHHA2007, Kiruna, June 11–13 2007, 257–267.

[2.98] Z. Wen, X. Jin, and W. Zhang, Contact-impact stress analysis of rail joint region using the dynamic finite element method, Wear 258 (2005), 1301–1309.

[2.99] T. Pang, M. Dhanasekar, Dynamic finite element analysis of the wheel–rail interaction adjacent to the insulated joints, Proceedings of 7th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems (CM2006), Brisbane, Australia, September 24–26, 2006: 509–516.

[2.100] M. Pletz, W. Daves, F.D. Fischer and H. Ossberger. A dynamic wheel set-crossing model regarding impact, sliding and deformation. Proceedings of 8th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems (CM 2009), 801–808.

[2.101] M. A. Meyers and K.K. Chawla, Mechanical behavior of materials, Prentice Hall, Upper Saddle River, New Jersey, 1999.

[2.102] C.O. Frederick, Future rail requirements, Rail quality and maintenance for modern railway operation, Delft, 1992, 3–14.

[2.103] D.F. Cannon, K.O. Edel, S.L. Grassie and K. Sawley, Rail defects: an overview. Fatigue Fract Engng Master Struct, 26 (2003): 865–887.

[2.104] A.V. Olver, The mechanism of rolling contact fatigue: an update, Proc. IMechE, Part J: J. Engineering Tribology, 219 (2005), 313-330.

[2.105] K.L. Johnson, Plastic flow, residual stresses and shakedown in rolling contact, proceedings of the international symposium on contact mechanics and wear of rail/wheel systems. Kingston, R.I. July 8–11, 1986, 83–98.

[2.106] J.W. Ringsberg, M. Loo-Morrey, B.L. Josefson, A. Kapoor, J.H. Beynon, Prediction of fatigue crack initiation for rolling contact fatigue, International Journal of Fatigue 22 (2000) 205–215.

[2.107] G. Donzella, A. Mazzù and C. Petrogalli, Competition between wear and rolling contact fatigue at the wheel–rail interface: some experimental evidence of rail steel, Proc. IMechE, Part F: J. rail and Rapid Transit, 223 (2009), 31–44.

[2.108] A. Ekberg, E. Kabo and H. Andersson, An engineering model for prediction of RCF of railway wheels, Fatigue Fract Engng Mater Struct 25 (2002), 899–909.

[2.109] G.N. Haidemenopoulos, A.D. Zervaki, P. Terezakies J. Tzanis, A.E. Giannakopoulos, and M. K. Kotouzas, Investigation of rolling contact fatigue cracks in a grade 900A rail steel of a metro track, Fatigue Fract Engng Mater Struct 29 (2006), 887–900.

[2.110] S. Stichel, H. Mohr, J. Ågren and R. Enblom, Investigation of the risk for rolling contact fatigue on wheels of different passenger trains. Vehicle System Dynamics, 46 (Suppl.) (2008), 317–327.

[2.111] D.T. Eadie, D. Elvidge, K. Oldknow, R. Stock, P. Pointner, J. Kalousek, and P. Klauser, The effects of top of rail friction modifier on wear and rolling contact fatigue: Full-scale rail–wheel test rig evaluation, analysis and modeling, Wear 265 (2008), 1222–1230.

[2.112] A. Böhmer, M. Eetz and K. Knothe, Shakedown limit of rail surfaces including material hardening and thermal stresses, Fatigue Fract Engng Mater Struct 26 (2003), 985–998.

[2.113] Y. Jiang, H. Sehitoglu, A model for rolling contact failure, Wear 224 (1999), 38–49.

[2.114] J Sandström, A Ekberg, Numerical study of the mechanical deterioration of insulated rail joints, Proc. IMechE Part F: J. Rail and Rapid Transit 223 (2009), 265–273.

[2.115] T.E. Tallian, Simplified contact fatigue life prediction model: Part I. Review of published models, Part II. New model, ASME Journal of Tribology 114 (1992), 207–220.

[2.116] V. Bhargava, G.T. Hahn, and C.A. Rubin, An elastic-plastic finite element model of rolling contact, Part I and Part II, Journal of Applied Mechanics 52 (1985), 67–82.

[2.117] A.F. Bower, Cyclic hardening properties of hard-drawn copper and rail steel; A.F. Bower, KL. Johnson, The influence of strain hardening on cumulative plastic deformation in rolling and sliding line contact. J. Mech. Phys. Solids 37 (4) (1989), 455–493.

[2.118] Z. Wen, X. Jin and Y. Jiang, Elastic-plastic finite element analysis of non-steady state partial slip wheel–rail rolling contact, Journal of Tribology 127 (2005), 713–721.

[2.119] Y. Jiang, B. Xu, and H. Sehitoglu, Three-dimensional elastic-plastic stress analysis of rolling contact, ASME J. Tribol. 124 (2002), 699–708.

3. Solution of frictional rolling contact

Solution to the wheel–rail rolling contact is the key to the calculation of vehicle–track interaction, and to a stress analysis of rail. In this chapter, an FE model is developed employing an explicit integration scheme to solve the frictional wheel–rail rolling contact. In section 3.1, the model is validated in elasticity and in statics for the normal and tangential solutions. In Section 3.2, the influence of plastic deformation on normal and tangential solutions of rolling contact is investigated with the validated FE model by introducing a bi-linear material model.

Chapter 1: Introduction to squats and their consequences.

Chapter 2: Literature review and outline of this dissertation.

Chapter 3: Solution of frictional rolling contact.

Chapter 4: Squat initiation, growth and wheel–track dynamic forces.

Chapter 5: A minimum critical size for squats to initiate.

Chapter 6: Conclusions and recommendations.
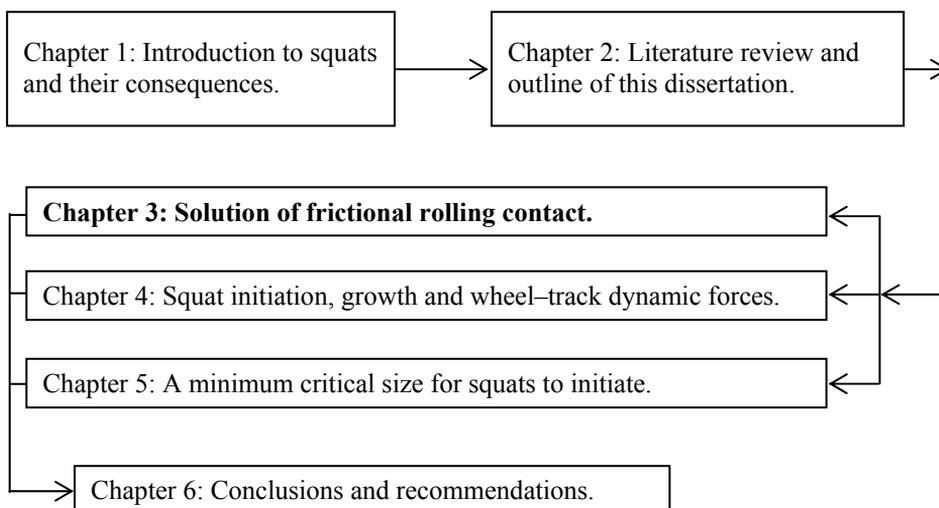
3.1 The solution of frictional wheel–rail rolling contact with a 3D transient finite element model: Validation and error analysis[1]

ABSTRACT

Wheel–rail frictional rolling contact has been solved in three dimensions for linear elasticity and statics, and with the half space assumption. In recent years, the normal contact problem has also been treated in elasticity and elasto-plasticity with the finite element (FE) method; the half-space assumption is dropped. In this paper, a 3D transient FE model is presented to solve both the normal and the tangential contact problems simultaneously for arbitrary geometry. The explicit time integration scheme employed renders it intrinsically suitable for transient analysis of high frequency dynamic processes. The model is first validated with Hertz theory and Kalker's computer program CONTACT for the normal solution and with CONTACT for the tangential solution, for the case of wheel tread–rail top contact. Subsequently, the influence of element size on solution accuracy is examined by comparing the results of different element sizes. The analyses show that the presented model is valid for the solution of frictional rolling.

3.1.1   Introduction

The rolling contact of a wheel on a rail is the key to many wheel–rail related problems such as wear, plastic deformation, rolling contact fatigue (RCF), vehicle dynamics, and vehicle–track interaction. Recently, it has attracted more and more research attention with the continuous increase in running speed, axle load, and traffic density on many railway networks. However, there are no established experimental techniques to determine the contact stress, especially under dynamic conditions. Therefore, emphasis has been placed on analytical and computer based numerical approaches.

Traditionally the wheel–rail contact has been divided into the normal and tangential problems for solution. This is justified because the friction force transmitted between the contact bodies usually has negligible influence on the shape of the contact patch and the distribution of pressure under the elastic condition [3.1.1]. The division was necessary because of the limitations of analytical methods and for the sake of simplicity of the solution. In the normal problem the contact patch and pressure are solved for. The tangential contact problem is mainly associated with friction, in which a solution is sought for the distributions of surface shear stress (tangential traction) and micro-slip, and the related areas of adhesion and slip in the contact patch.

Micro-slip is the relative velocity of two contacting material particles: one is in the rail surface and the other in the wheel surface. In the literature on rolling contact, 'micro-slip' is often called 'slip' for short [3.1.2]. Since 'slip' may also mean creepage in Railway Engineering and in Tribology [3.1.3], 'micro-slip' is used in this paper to avoid confusion. 'Slip area' is, however, used when an area of micro-slip is referred to.

The normal problem of contact between bodies of quadratic surfaces was solved by Hertz in 1882 [3.1.4]. Frictional rolling contact was first treated by Carter in two dimensions in 1926 [3.1.5], in which the surface shear stress in the contact patch satisfied the Coulomb law everywhere. By assuming an elliptical adhesion area, Vermeulen and Johnson obtained a 3D solution for rolling contact between bodies of quadratic surfaces without spin [3.1.6]. The 3D rolling contact problem with arbitrary longitudinal and lateral creepages and spin was solved by Kalker [3.1.2] with several numerical methods, with the latest development being implemented in the computer code CONTACT. Nevertheless, CONTACT is still based on linear-elasticity, statics, and the half-space assumption, like all the other solutions mentioned above.

With the finite element (FE) method the half-space assumption can easily be dropped, and material properties that are more complex than linear elasticity can usually be dealt with. Furthermore, the dynamic behavior can also be considered by performing transient FE simulations with explicit time integration scheme. Using the FE method Bogdansky et al. solved the normal problem of wheel–rail contact in three dimensions for elasticity and statics [3.1.7]. More recently, the normal problem of contact of elastic-perfectly plastic material was investigated by Wiest et al. [3.1.8] with a 3D static FE model. Other solutions of the normal problem with FE can be found in [3.1.9–3.1.17]. However, treatment of the tangential problem of wheel–rail rolling contact with FE is still

rare. Only some work in 2D explicit solution [3.1.13] or solutions in statics [3.1.7, 3.1.14–3.1.17] have been reported.

Li et al. [3.1.18–3.1.20] treated wheel–rail rolling contact in the vehicle–track interaction system with an explicit FE method. The reason to opt for an explicit scheme was that it allows for the investigation of transient rolling when a high frequency dynamic effect has to be considered, for example, contact at track short wavelength defects such as squats, poor welds, and fish-plated joints. Details of the solution of the contact problem were, however, not presented there.

In this paper, the 3D transient FE model developed in previous work [3.1.18–3.1.20] is further improved to obtain a detailed solution of frictional rolling contact. The focus is set on the tangential problem. By applying a fine mesh, the contact is solved under the elastic condition for validation against the Hertz theory and CONTACT. Subsequently, the influence of element size on the solution is examined with coarser mesh. Finally, conclusions about the FE model are drawn.

3.1.2 The FE model

A schematic diagram of the model is shown in Figure 3.1.1. The contact surface of the wheel is cylindrical. The rail profile is 54E1. This means that the contact occurs in the lateral centre of the rail top against the wheel tread. The sprung mass, which together with the unsprung mass (the wheel) forms the wheel load, is lumped and supported by a group of springs and dampers of the primary suspension. The second suspension of a vehicle, if any, is not considered because the vibration of the sprung mass has negligible effect on the transient state of the wheel–rail contact due to the fact that the wavelength of the vibration of the sprung mass, in the order of metres, is much longer than the dimension of the contact patch between wheel and rail, which is in the order of 10mm. Note that the contact surfaces of both the wheel and the rail are smooth.
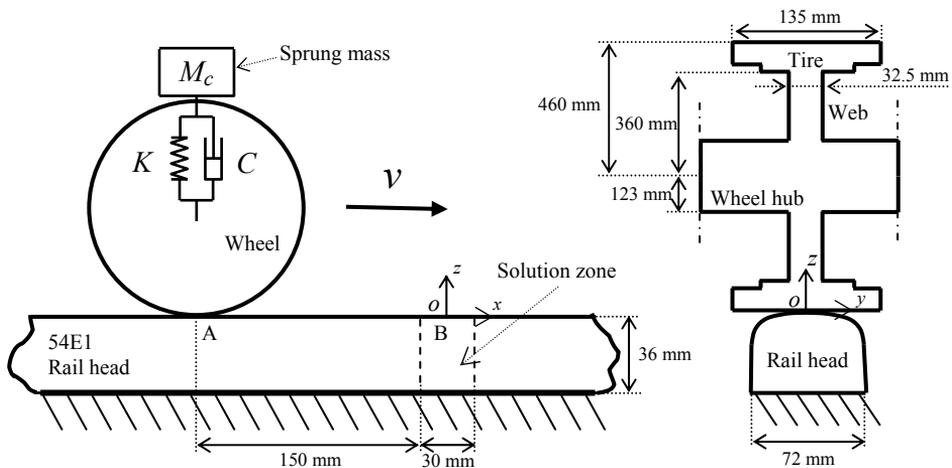
Figure 3.1.1. Schematic diagram of the transient FE model. The contact point when the wheel and rail are considered rigid is in the lateral centre of the rail top ($y = 0$).

A central difference method is employed for the explicit integration scheme. The critical time step is determined by the Courant stability condition [3.1.21]. This condition requires that the time step is small enough to ensure that a sound wave may not cross the smallest element during one time step. As long as the time step is smaller than the critical time step, convergence is guaranteed. This also means that the FE model will automatically include in its solution all the vibration modes which belong to both the physical model and the FE mesh. Therefore, the vibration of the structure and continua excited by the rolling of an elastic wheel on an elastic rail, and the associated wave propagation, will all be included in the FE solution. In contrast, the Hertz theory and CONTACT are both based on statics. To ensure that the results of the FE model are comparable to those of the Hertz theory and CONTACT for the purpose of validation, such vibration and wave propagation should be minimized. This can be achieved by supporting the rail head continuously on a rigid foundation (see Figure 3.1.1) and by relaxing the dynamic process of the wheel-on-rail rolling until a quasi-static state is achieved.

The solution process of the rolling contact is as follows. The wheel is set to roll along the rail towards the zone where the solution is sought, which is 0.15 m away from the initial position of the wheel (see Figure 3.1.1). This distance is needed for the rolling contact to reach the required quasi-static state. As illustrated in Figure 3.1.1, a 3D right-handed Cartesian coordinate system (*Oxyz*) is defined, of which the origin *O* is located at the centre of the solution zone (position B), in the middle point of the 54E1 profile. During the rolling of the wheel, a solution is sought at the instant when the centre of the contact patch is at *O*.
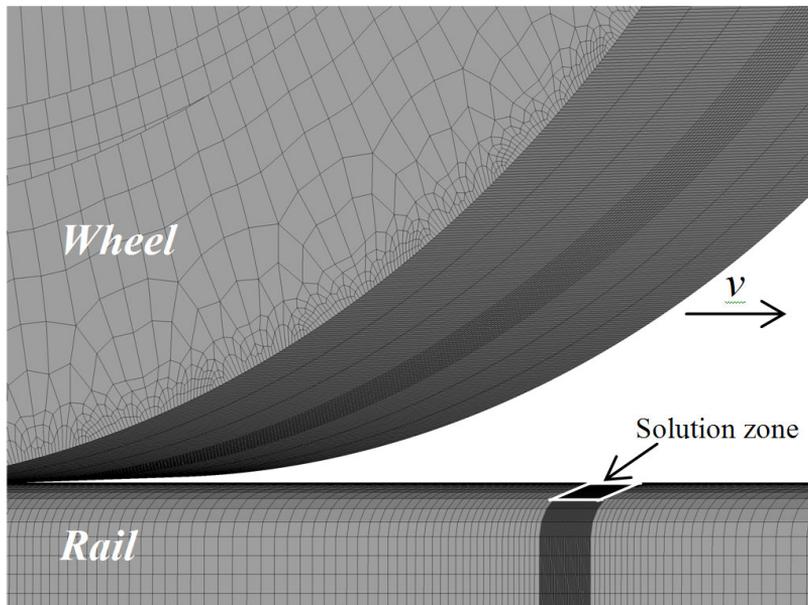
Figure 3.1.2. Meshing of the 3D transient FE model.

The wheel and the rail head are meshed with 8-node solid elements. In order to achieve a high accuracy of the solution with a reasonable model size, non-uniform meshing is used, with finer elements in the contact surfaces, and the finest element in the solution zone, as shown in Figure 3.1.2. The model is discretized into three meshes of different finenesses, with the dimensions of the faces of the finest elements in the contact surfaces being 0.33×0.33, 0.63×0.63, and 1.3×1.3mm, respectively. They are hereinafter called the 0.33×0.33, 0.63×0.63, and 1.3×1.3 meshes, respectively. The wheel–rail interaction is calculated using a surface-to-surface contact algorithm [3.1.22]. In this work, the model is first validated in Section 3.1.3 with the 0.33×0.33 mesh. Subsequently the influence of the element size is studied in Section 3.1.4 by comparing the results of the three meshes.

The model parameters are listed in Table 3.1.1. Different friction exploitation levels are simulated by applying the corresponding traction force (F_L) at the contact, with the traction coefficient (μ) being

$$\mu = F_L / F_N \leq f = 0.5 \tag{3.1.1}$$

where F_N is the normal contact force and f is the friction coefficient. Hence, a rolling in traction with prescribed longitudinal creep force is simulated in the present work. It should be noted that no lateral creepage or kinematic spin is involved. Spin due to contact geometry is automatically taken into account by the solution method. Here, the friction coefficient of 0.5 is chosen, while the maximum traction coefficient used later in this paper is 0.3. This choice is made to ensure that when later the traction coefficient takes a value of 0.3 or less, the division of the contact patch into the areas of adhesion and slip is the most convenient for clear comparison of results by the FE method and by CONTACT. Coulomb's law of friction is used in this paper for the FE solution because it is employed in CONTACT.

Table 3.1.1 Values of parameters used.

Parameters	Values
Lumped sprung mass, M_c	13.4 t
Wheel diameter, ϕ	0.92 m
Wheel mass, M_w	900 kg
Friction coefficient, f	0.5
Rolling speed, v	140 km/h
Stiffness of primary suspension, K	1150 KN/m
Damping of primary suspension, C	2500 Ns/m
Young's modulus of material, E	210 GPa
Poisson's ratio of material, υ	0.3
Density of material, ρ	7800 kg/m^3

3.1.3 Validation of the FE model

The correctness and accuracy of CONTACT have been proven for the elastic contact between wheel tread and rail top. It is therefore taken as reference to validate the transient FE model for a solution in linear elasticity. To this end, FE results are sought for a quasi static state of rolling contact as mentioned above.

Two cases of rolling contact are solved in this section with different longitudinal creep forces, that is, $\mu = 0.15$ and 0.3, respectively. The 0.33×0.33 mesh is employed. The values of parameters listed in Table 3.1.1 are used, where appropriate, for the Hertz solution, the CONTACT solution, and the FE solution. Note that the results shown below are all for the rail.

3.1.3.1 Solution of the normal problem

In the FE approach, the contact patch is determined with the normal nodal force as follows:

$$\text{A node is in contact if:} \qquad \left| F_{n\_N} \right| > 0 \qquad\qquad (3.1.2)$$

where F_{n_N} is the nodal force in the direction normal to the local surface.

It is known that the radius of the 54E1 rail profile in the z-y plane is 0.3 m in the central part. Together with the radius of the wheel (0.46 m), the solution of the normal contact can also be obtained using Hertz theory. Table 3.1.2 compares the solutions of the normal contact problem by Hertz theory, CONTACT, and the FE model. Note that the radii of curvature of the wheel contact surface in the z-y plane and of the rail contact surface in the z-x plane are both assumed to be infinite.

Due to the discretization of the system, the semi-axes of the contact patch obtained from the FE approach can only be a multiple of element length. For instance, as shown in Table 3.1.2, the semi-axis a is exactly 8.00mm in the FE solution, which is the length of 24 elements. However, the semi-axis is 7.73mmin the Hertz solution, corresponding to the length of 23.19 elements. This is why the relative difference of 3.5% is generated in semi-axis a between the two solutions. In terms of the area of the contact patch, which is proportional to $a \times b$, their difference is even larger (see Table 3.1.2). Nevertheless, the error in dimensions of the contact patch does not influence the pressure distribution significantly (see Figure 3.1.3) due to the fact that the pressure is almost vanishing on the border of the contact patch. Hence, the error in maximal pressure, the more important quantity in this case, is a maximum of 0.3% for the FE solution. Note that this difference, or error, caused by discretization, which is inherent to various numerical methods, also exists in the solution by CONTACT.

Table 3.1.2 Comparison of the solutions of the normal problem by the three methods.

Approach	Contact patch			Maximum Pressure (MPa)
	Semi-axis a (mm)	Semi-axis b (mm)	Area (mm^2)	
Hertz	7.73	5.86	142.31	1340
CONTACT	7.80	5.69	139.68	1347
FE	8.00	6.03	151.55	1344
Difference w.r.t. Hertz	3.5%	2.9%	6.5%	0.3%
Difference w.r.t. CONTACT	2.6%	6%	8.5%	–0.2%

Figure 3.1.3 gives the distributions of contact pressure along the longitudinal axis of the contact patch ($y = 0$, 'longitudinal axis' will be used for short hereinafter), which shows good agreement between the results of the three approaches. It should be noted that the small pressure fluctuation (<5% of the maximal Hertz pressure) of the FE results with respect to the Hertz solution is caused by the high frequency vibrations and wave propagation of the wheel and rail continua. Similar fluctuations can also be observed in later figures and will be discussed further in Section 3.1.5.

Based on the discussions above, it can be concluded that except for the discretization error, the normal solution of the FE method agrees well with those of the Hertz theory and CONTACT.
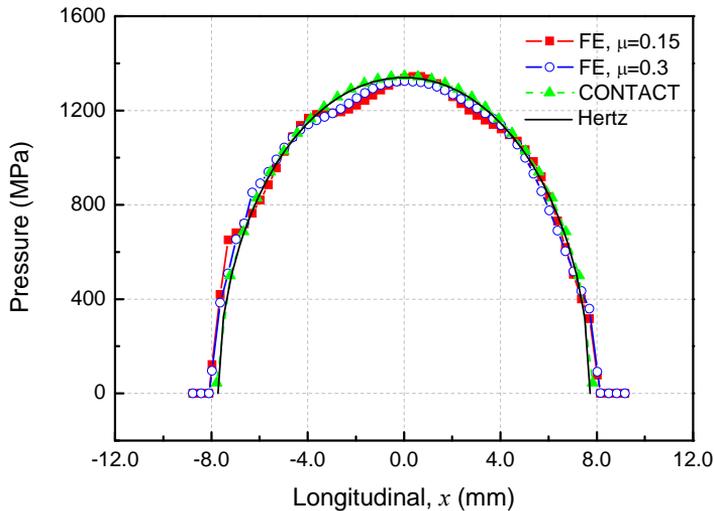
Figure 3.1.3. Distribution of contact pressure along the longitudinal axis of the contact patch ($y = 0$).

3.1.3.2   Solution of the tangential problem

a) Areas of slip and adhesion

The most characteristic feature of frictional rolling contact is the distinction between the adhesion area and slip area in the contact patch. The area of adhesion is determined as given in Equation (3.1.3), whereas the remaining part of the contact patch is in micro-slip.

$$\text{The node is in adhesion if:} \quad f\left|F_{n\_N}\right| - \left|F_{n\_T}\right| > \varepsilon_T \quad\quad\quad (3.1.3)$$

where F_{n_T} is the nodal forces in the tangential direction and ε_T is a tolerance. In the present work, ε_T is 0.3% of the maximal tangential nodal force in the contact patch. In other words, a node could be on the wrong side of adhesion-slip border if the difference between the actual tangential nodal force and its limit determined by the Coulomb's law of friction is less than 0.3% of the largest tangential nodal force in the contact area.

Figure 3.1.4 shows the adhesion-slip area distributions in the contact patches when the traction coefficients are 0.15 and 0.3 respectively. Every block or bullet in the plots represents a node in the FE model. As can be seen, the adhesion area is located at the leading part of the contact patch and decreases in size with the increase in the tangential load. Figure 3.1.5 shows the areas of slip and adhesion in the contact patch obtained with CONTACT under the same conditions; each block or bullet represents a surface element.

Comparing Figure 3.1.5 to Figure 3.1.4, it can be seen that the adhesion and slip areas obtained with the two methods are in good agreement with each other in their locations in the contact patches. Furthermore, the adhesion areas are 81% and 60% of the contact patch in the FE solutions for $\mu = 0.15$ and 0.3, respectively, which again agree well with the corresponding values of the CONTACT solutions, that is 83% and 60%. The largest relative difference is only 2%.

It is worth noting that in the FE solution there are isolated nodes in adhesion in the area of slip, i.e. the isolated green bullets in the blue area of blocks in Figure 3.1.4(b). On the other hand, in the CONTACT solution there are isolated nodes in micro-slip in the area of adhesion, i.e. the isolated blue blocks in the green area in Figure 3.1.5. They are generated due to certain numerical errors. Such errors happen mainly in an area near the border between the areas of slip and adhesion for the FE method, whereas they occur mainly in an area near the leading edge of the contact patch for CONTACT. Obviously, such errors are associated with the tolerances such as ε_T used for the determination of the areas of slip and adhesion. From the viewpoint of stress and strain analysis, such errors are tolerable because the employed tolerance is very small (0.3% is used in the FE solution as mentioned above).
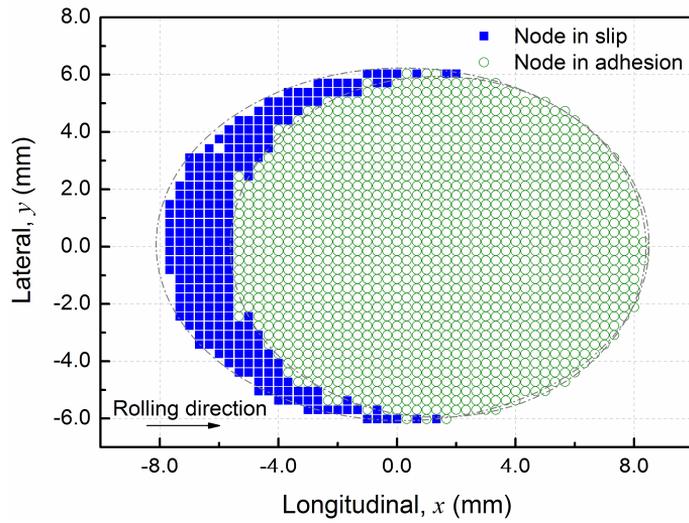
(a)
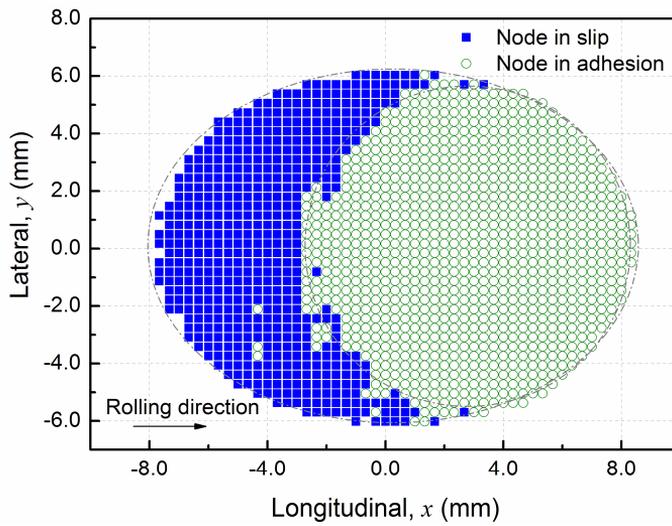
(b)

Figure 3.1.4. The adhesion-slip area distribution from the transient FE simulations. (a) $\mu = 0.15$ and (b) $\mu = 0.3$. The ellipses of dashed lines are approximations of the calculated contact patch and area of adhesion. These approximations are used in subsequent figures for convenience of visualization.
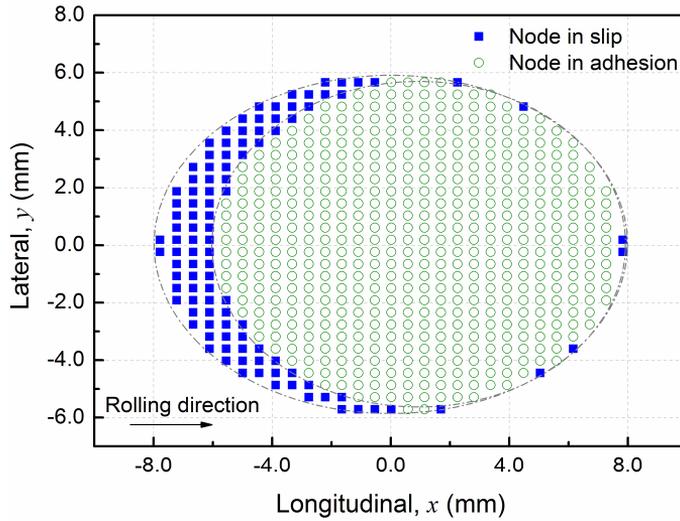
(a)
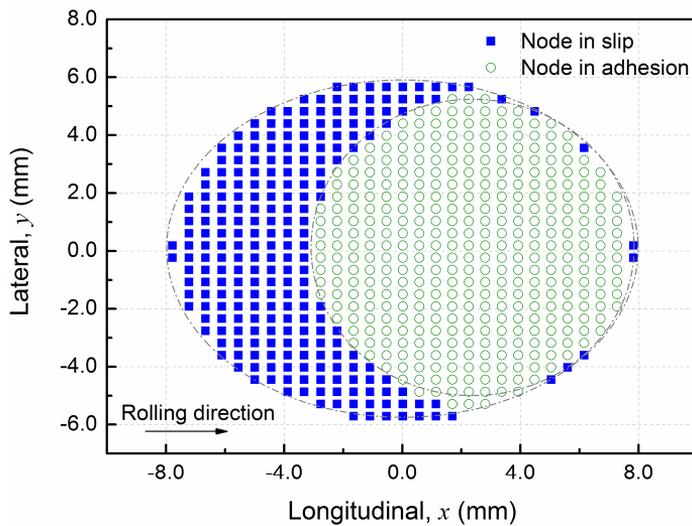
(b)

Figure 3.1.5**.** The adhesion-slip area distribution by CONTACT. (a) $\mu = 0.15$ and (b) $\mu = 0.3$.

b) Surface shear stress

Figure 3.1.6 shows the fields of surface shear stress in the contact patch when a traction coefficient of 0.3 is applied. The FE and CONTACT solutions are given in Figures 3.1.6(a) and 3.1.6(b), respectively. The arrows point in the direction of the stress and their length is proportional to the magnitude.
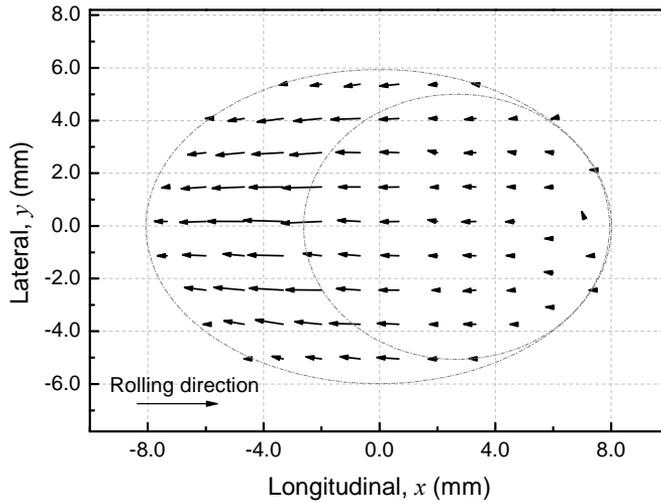
(a)
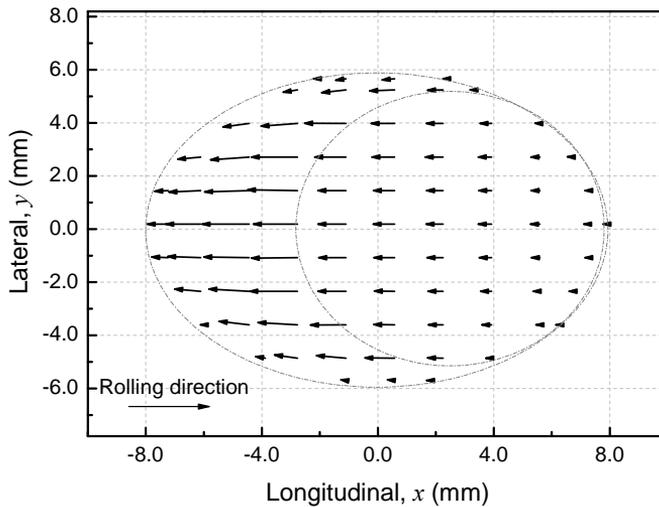
(b)

Figure 3.1.6. The vector graph of the surface shear stress when $\mu = 0.3$. (a) FE solution and (b) CONTACT solution.

As expected, the surface shear stress gradually increases in magnitude from the leading edge to the trailing edge of the adhesion area, whereas it follows the same trend as the pressure variation in the slip area. This can be seen more clearly in the 2D surface shear stress distribution along the longitudinal axis given in Figure 3.1.7. Furthermore, since the traction force is applied only in the rolling direction, the surface shear stress shown in Figure 3.1.6 is almost completely in the longitudinal direction in the entire contact patch. The small lateral component is due to some small geometrical spin. Such spin comes into existence because the contact patch is non-flat in the direction of the rolling axis of the wheel as a result of the non-flat rail head. The tangent to the contact

patch in the plane parallel to the *y-z* plane forms a varying angle α with the rolling axis of the wheel, so that local spin arises. $\alpha = 0$ at $y = 0$, and $|\alpha|$ increases when $|y|$ increases. Therefore, the lateral component of the surface shear stress is zero at $y = 0$, and increases with the increase in $|y|$. The spin is symmetrically distributed about the longitudinal axis since the rail profile is symmetrical about the *x-z* plane.

Comparing the FE and CONTACT solutions shown in Figure 3.1.6, it can be found that the direction of the surface shear stress agrees well in the whole contact patch. From Figure 3.1.7, it is further seen that their magnitudes are also in good agreement in the slip area, whereas in the adhesion area the difference is up to 20%.
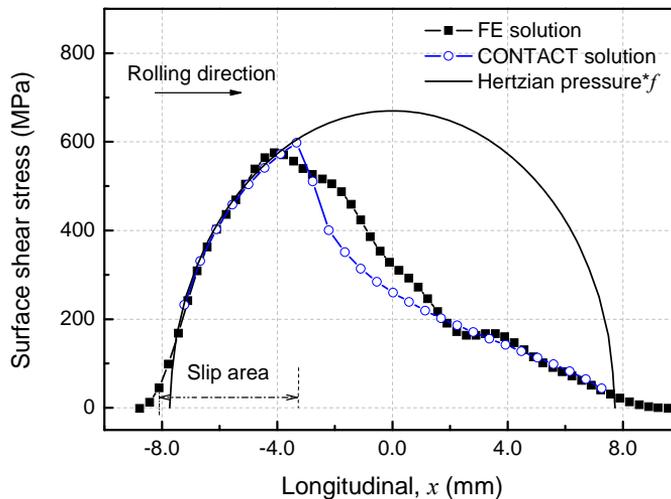
Figure 3.1.7. The surface shear stress distribution (absolute values) along the longitudinal axis when $\mu = 0.3$.

c) Micro-slip

The field of the micro-slip obtained from the FE simulation is shown in Figure 3.1.8(a) for $\mu = 0.3$. Figure 3.1.8(b) gives the corresponding result obtained by CONTACT. Clearly, the magnitude of the micro-slip is zero in the adhesion area, as should occur by definition. In the slip area, its magnitude progressively increases towards the trailing edge of the contact patch with the direction opposite to the shear stress shown in Figure 3.1.6. Again, the two graphs shown in Figure 3.1.8 are in agreement with each other in both the direction and the trend of the change in magnitude of the vectors.

Figure 3.1.9 further shows the magnitude distributions of the micro-slip along the longitudinal axis, from which the good agreement between the two approaches can clearly be observed. Only a slight difference occurs at the border of the areas of adhesion and slip, where the influence of numerical error is relatively large due to the small magnitude of the micro-slip.
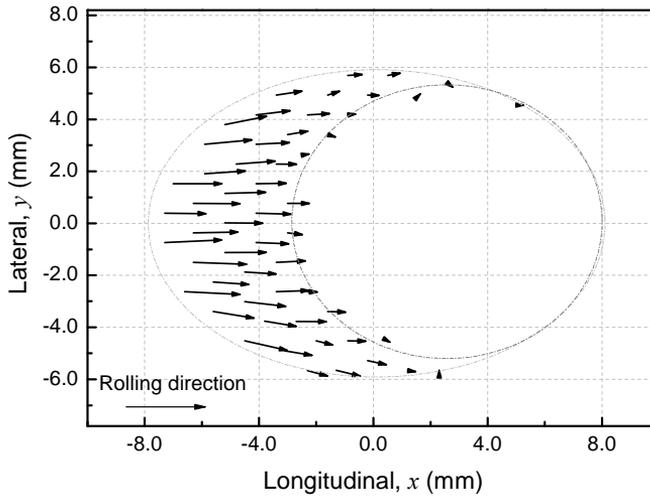
(a)
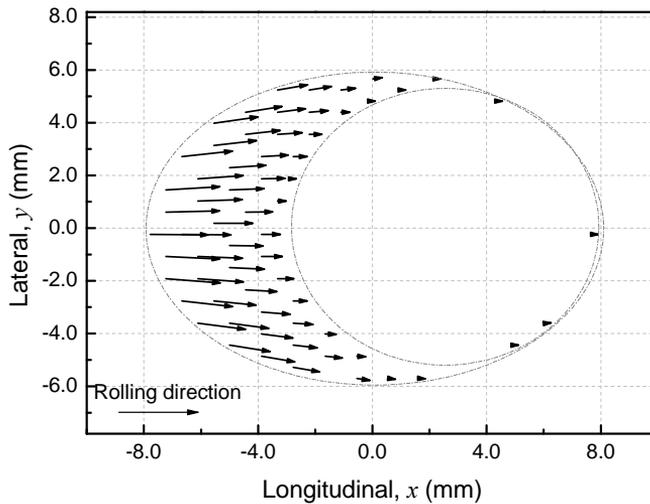
(b)

Figure 3.1.8. Vector diagram of the micro-slip in the contact patch when μ = 0.3. (a) FE solution and (b) CONTACT solution.

3.1.3.3   The influence of tangential load on pressure

With different traction coefficients of μ = 0.15 and 0.3, the contact patch from the FE analysis is almost entirely the same in size and shape, as shown in Figure 3.1.4. Moreover, the pressure distribution is also independent of the traction coefficients (see Figure 3.1.3). Therefore, the FE results confirm that the influence of the tangential load on the normal solution is negligible under the elastic condition.
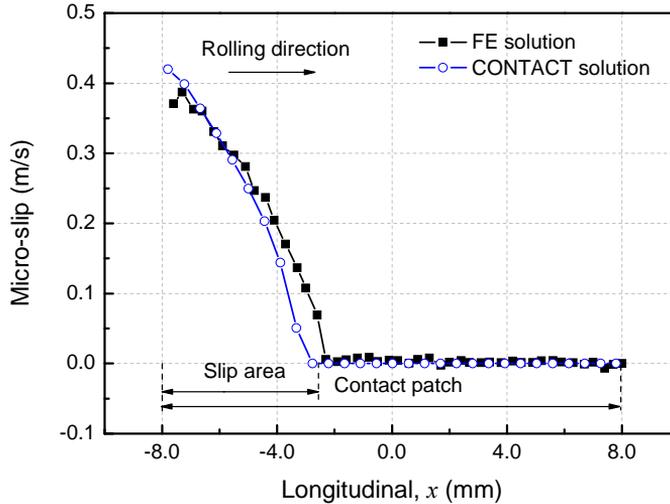
Figure 3.1.9. Distribution of micro-slip (absolute values) along the longitudinal axis when $\mu = 0.3$.

3.1.4 The influence of element size on the solution

Validity of the model has been shown above with the 0.33×0.33 mesh. Such a fine mesh is time-consuming and may be unnecessary for applications such as those encountered in [3.1.18] and [3.1.19]. Taking the model of Figure 3.1.1 as an example, it takes about 27 hours for a single 3.73 GHz Pentium 4 processor to simulate the rolling over a distance of 0.22 m under the default rolling speed, see Table 3.1.3. It is well known that the following two factors, among others, determine the computation time of an explicit FE simulation: 1) number of elements, which decreases with the increase in the size of the elements; 2) size of the time step, which is approximately proportional to the size of the smallest element. Therefore, a coarser mesh should be used whenever possible. By doubling the length of the elements the model size can be reduced by 7/8 in 3D situation, and the time step can be doubled in the meantime, so that the total computation time can be reduced by 15/16.

In this section, the three meshes, that is, 0.33×0.33, 0.63×0.63, and 1.3×1.3, of the same model are employed to assess the influence of element size on the accuracy of the solution and computational costs. With all the other parameters being the same, the rolling contact with a traction coefficient of 0.3 is solved with the three meshes. The results are summarized in Table 3.1.3 for comparison together with their computation time. Note that the difference of the computation time between the three meshes is not as large as estimated above. This is because the finest element size does not influence the model size as effectively as estimated due to the employment of non-uniform meshing; see the element numbers listed in Table 3.1.3.

Table 3.1.3 The contact solution comparison (traction coefficient of 0.3) when a different mesh is employed.

Parameters		Size of finest element in mesh (mm)			Max Diff of size 0.63 w.r.t. 0.33	Max Diff of size 1.3 w.r.t. 0.33
		0.33	0.63	1.3		
Contact patch	Semi-axis a (mm)	8.00	7.83	8.48	−2.0%	6.0%
	Semi-axis b (mm)	6.03	5.95	7.07	−1.3%	17%
	Area (mm^2)	151	146	188	−3.3%	24%
Maximum pressure (MPa)		1344	1323	1297	−1.6%	3.5%
Area ratio between Adhesion area and contact patch (%)		60.0	59.6	62.5	−0.7%	2.9%
Maximum surface shear stress (MPa)		575	580	485	8.7%	−16%
Maximum relative speed (m/s)		0.39	0.41	0.52	5.0%	33%
Element number (k)		219	116	53		
Calculation time (h)		27.05	6.29	2.37		

Note: The calculation time is for a single 3.73 GHz Pentium 4 processor to simulate the rolling over a distance of 0.22 m under the rolling speed of 140 km/h.

3.1.4.1   The normal solution

From Table 3.1.3 it can be seen that the element size has a considerable influence on the calculated size of the contact patch. Comparing the size and shape of the contact patches shown in Figures 3.1.4(b) and 3.1.10(b), the 1.3 × 1.3 mesh approximates the contact ellipse with less accuracy and the contact patch is obviously larger than those of the 0.33 × 0.33 mesh and the Hertz solution. However, the contact patch obtained from the 0.63 × 0.63 mesh is very close to that of the 0.33 × 0.33 mesh; see Table 3.1.3 or compare Figures 3.1.4(b) and 3.1.10(a).
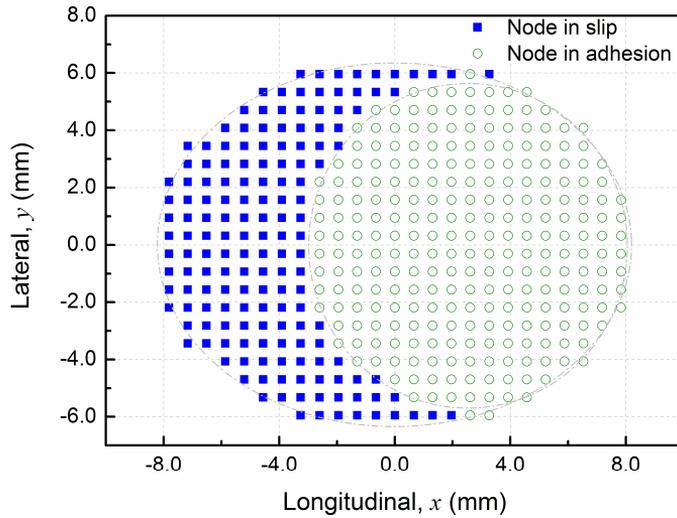
(a)
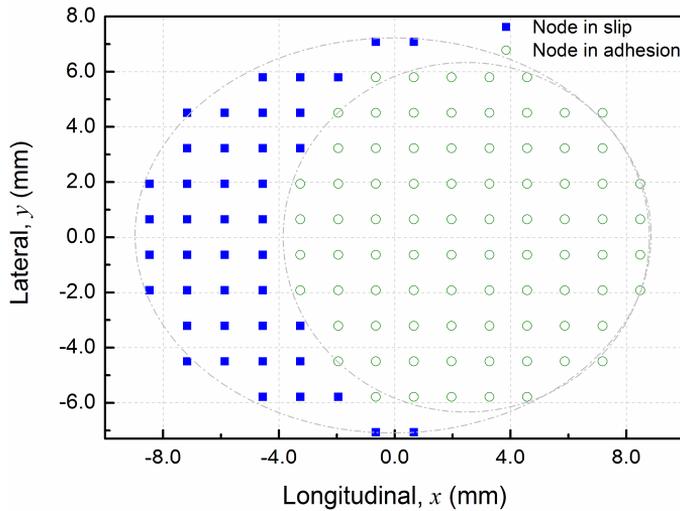
(b)

Figure 3.1.10. The adhesion-slip area distributions obtained with the transient FE simulations with different meshes, $\mu = 0.3$. (a) Solution of the 0.63×0.63 mesh and (b) solution of the 1.3×1.3 mesh.

The pressure distributions along the longitudinal axis are given in Figure 3.1.11(a). Generally speaking, the difference relative to the Hertz solution increases with the enlargement of the element size. The following causes contribute to this trend: (1) The position of nodes at which the pressure is calculated for the longitudinal axis deviate from the axis. In the FE model the pressure is derived from the normal nodal force. When the element size becomes large, the distance between the longitudinal axis and the nearest nodes to the axis may become large for discretization reason, so that the error in the

calculated pressure at the longitudinal axis increases. (2) Different element sizes lead to slightly different results for the dynamic relaxation at the moment the contact solutions are obtained. (3) The changes in size and actual shape of the contact patch caused by the increased element size can also influence the pressure distribution, especially at the edges of the contact patch, which can be seen in Figure 3.1.11. Nevertheless, this influence is small since the pressure distribution at the edge of the contact patch is not important due to its small magnitude.

(a)

(b)
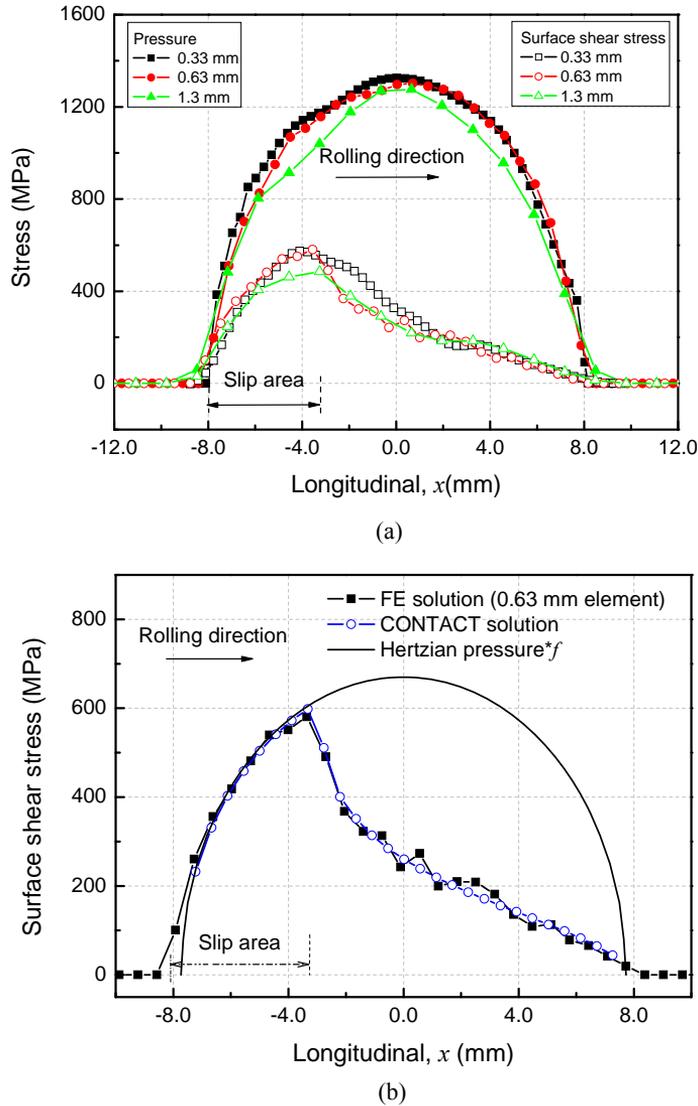
Figure 3.1.11. (a) The pressure and surface shear stress distributions (absolute values) along the longitudinal axis when different meshes are employed, $\mu = 0.3$. (b) Surface shear stress distributions of the solutions by CONTACT and by the FE model when the element size is 0.63 mm.

The pressure distributions obtained from the 0.33 × 0.33 mesh and the 0.63 × 0.63 mesh are very close to each other (see Figure 3.1.11(a)); the maximum pressure of the 1.3 × 1.3 mesh is reduced by 3.5% with respect to that of the 0.33 × 0.33 mesh and by 3.2% with respect to that of the Hertz solution. In contrast, along the longitudinal axis, the error of the solution of the 1.3 × 1.3 mesh relative to the Hertz solution is up to 16% at $x = -4$ mm. This error may be mainly attributed to the fluctuation of the stress between $x = -6$ and -2 mm. Since the largest error is 16%, it may be concluded that the 1.3 × 1.3 mesh should be accurate enough for the solution of many engineering problems.

3.1.4.2   The tangential solution

The calculated areas of adhesion and slip when element sizes of 0.63 and 1.3 mm are used are shown in Figure 3.1.10. When compared with Figure 3.1.4(b) (element size of 0.33 mm), it can be seen that they agree well in shape and position of the adhesion and slip areas, although the number of nodes involved is much reduced. Furthermore, the relative size of the adhesion and slip areas is also in good agreement, as shown in Table 3.1.3. In addition, the isolated nodes in adhesion in the slip areas, seen in Figure 3.1.4(b), disappeared when the element size was increased. The reason for this is not yet clear.

The surface shear stress distribution along the longitudinal axis is also plotted in Figure 3.1.11 together with the pressure. It is seen that the solutions of the three meshes generally agree well, with some difference in magnitude. When the element size increases from 0.33 to 0.63 mm, the surface shear stress in the trailing part of the adhesion area is decreased. Consequently, the surface shear stress distribution obtained from the FE model with an element size of 0.63 mm agrees better with the CONTACT result; see Figures 3.1.7 and 3.1.11(b). The reason for this is not yet clear and it needs further investigation. On the other hand, the surface shear stress distributions corresponding to element sizes of 0.63 and 1.3 mm agree well in the adhesion area (see Figure 3.1.11(a)).

In the slip area, the maximum surface shear stress differs by 16% when the element size increases from 0.63 to 1.3 mm, as shown in Figure 3.1.11(a) and listed in Table 3.1.3. This error of the 1.3 × 1.3 mesh in the shear stress is caused by the error in the pressure mentioned above, which is determined by the employment of Coulomb's law of friction and can be seen in the same figure.

Figure 3.1.12 shows the comparison of the micro-slip along the longitudinal axis. It can be seen that the micro-slip results agree well for all three cases even if the element size changes significantly, except at the trailing edge of the contact patch. In the case of 1.3 × 1.3 mesh, the micro-slip at the trailing edge is much higher than those of the other two cases, due to the larger calculated contact patch. One important application of the micro-slip is for the calculation of frictional power by multiplying the local micro-slip with the shear stress. Considering that the surface shear stress is very small at the edge of the contact patch, the error in the micro-slip would have only a small influence on the frictional power.
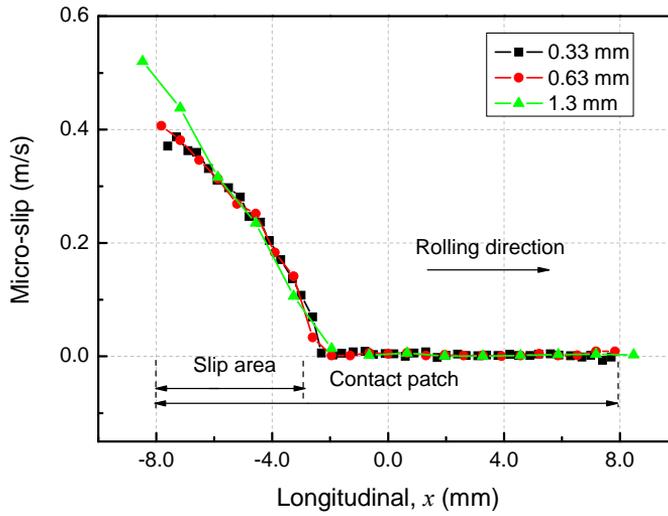
Figure 3.1.12. Distribution of micro-slip (absolute values) along the longitudinal axis when different meshes are employed, μ=0.3.

In view of the above discussions, it can be concluded that for the tested case the 0.63 × 0.63 mesh has an accuracy comparable to that of the 0.33 × 0.33 mesh. For engineering applications the accuracy of the 1.3 × 1.3 mesh should usually be acceptable. In addition, by using the 1.3 × 1.3 mesh the computation time is reduced to about 1/12 of that of the 0.33 × 0.33 mesh (see Table 3.1.3). For the same mesh size, a solution by CONTACT on the same computer needs less than 0.1 second. As expected, CONTACT is much faster than the FE model.

Considering the dimensions of the contact patch listed in Table 3.1.1, the element size of 0.63 mm is about 1/20 of the minor axis of the contact ellipse, which is 11.72 mm according to the Hertz theory. Hence, it may be tentatively concluded that with an element size of about 1/20 of the minor axis of the contact patch, an accurate solution can be obtained with the FE model, while with an element size of 1/10 of the minor axis, the accuracy may be acceptable for many engineering problems. More numerical experiments with extensive variations in load and contact geometry are needed to confirm this conclusion.

3.1.5   Discussions

To the best of the authors' knowledge there is currently no means to validate directly, experimentally or analytically, the solution of transient rolling contact of high frequency interaction. The presented FE model is therefore validated in the quasi-static case.

Based on the comparisons in Section 3.1.3, it is found that the results of the FE model agree well with those of the Hertz theory and CONTACT, where appropriate, in size and shape of the contact patches, in division of the contact patch into adhesion and slip areas, and in distributions of the pressure, the surface shear stress, and the micro-slip.

This shows that the FE model is valid for accurate solution of the frictional rolling contact between wheel tread and rail head.

As shown in Figure 3.1.1, the rail head is fixed on a rigid foundation. From the results shown above, it can be concluded that the wave reflection at the fixture has no significant effect on the results. In the FE approach, the normal and tangential contact problems are solved simultaneously. The elastic solution verifies that the tangential contact stress has negligible influence on the normal solution when the same material is used for both bodies in contact. Furthermore, the good agreement between the results of the FE model, the Hertz theory, and CONTACT confirms that half-space is indeed a good approximation for contact on the rail head.

It should be mentioned that fluctuations still exist in the obtained stress distributions, as observed in Figures 3.1.3, 3.1.7, and 3.1.11, although dynamic relaxation was applied to achieve a quasi-static state. These fluctuations or irregularities might make the figure less appealing, but such high frequency vibration is inherent to the rolling contact of continuum. Therefore, the solution by this method should be more close to the reality. This is particularly true in the case of contact at squats [3.1.18], and may also be true for wheel–rail interaction at other short-wavelength track defects. In the present work, only elastic solutions are presented, but it can be expected that the stress fluctuation due to vibration should be reduced in the presence of plastic deformation, which can dissipate some of the kinetic energy.

Finally, although a specific problem of wheel–rail rolling contact is treated in this paper, the same FE modelling approach can also be applied to solve general rolling contact problems. More complicated material model such as plasticity and strain rate dependency can also be readily included in the model, owing to the explicit integration solver.

3.1.6   Conclusions and further work

A 3D transient FE model, which is intrinsically suitable for the solution of dynamic problems in the high frequency range owing to its explicit integration scheme, is developed for the solution of frictional rolling contact problems.

- By approaching a quasi-static state, the FE model is validated against Hertz theory and CONTACT in both normal and tangential solutions, for the case of wheel tread–rail head contact.
- It is tentatively concluded that when the element size is about 1/20 of the minor axis of the contact patch, an accurate contact solution can be obtained, while when the size is 1/10 of the minor axis, the accuracy may be acceptable for many engineering problems. More numerical experiments are needed for a firm conclusion.
- A small fluctuation of stress is observed in the FE solutions, due to imperfect dynamic relaxation. In applications of the model to transient and dynamic problems, such fluctuation is inherent and realistic. However, in applications to static problems, error estimation is needed, or a better relaxation is desired before a solution is sought. These should be among the further work.

- The results and discussion show that the presented model is a valid and promising tool for the solution of frictional rolling contact, in statics and dynamics, in elasticity and plasticity. Furthermore, arbitrary contact geometry can be taken into account owing to the nature of finite element methods.

References

[3.1.1] H. Poritsky, N.Y. Schenectady, Stress and deflections of cylindrical bodies in contact with application to contact of gears and of locomotive wheels, J. Appl. Mech. 17 (1950), 191–201.

[3.1.2] J.J. Kalker, Three-Dimensional Elastic Bodies in Rolling Contact, Kluwer Academic Publishers, Dordrecht, 1990, 137–184.

[3.1.3] R. Lewis, R.S. Dwyer-Joyce, Wear at the wheel/rail interface when sanding is used to increase adhesion, J. Rail. Rapid. Transit. 220 (2006), 29–41.

[3.1.4] H. Hertz, Über die berührung fester elastischer körper, J. reine und angewandte Mathematik 92 (1882), 156–71.

[3.1.5] F.W. Carter, On the action of a locomotive driving wheel, Pro. R. Soc., London, A. 112 (1926), 151–157.

[3.1.6] P.J. Vermeulen, K.L. Johnson, Contact of non-spherical bodies transmitting tangential forces, J. Appl. Mech. 31 (1964), 338–340.

[3.1.7] S. Bogdanski, M. Olzak, J. Stupnicki, Numerical stress analysis of rail rolling contact fatigue cracks, Wear 191 (1996), 14–24.

[3.1.8] M. Wiest, E. Kassa, W. Daves, J.C.O. Nielsen, H. Ossberger, Assessment of methods for calculating contact pressure in wheel–rail/switch contact. Wear 265 (2008), 1439–1445.

[3.1.9] M. Pau, F. Aymerich, F. Ginesu, Distribution of contact pressure in wheel–rail contact area, Wear 253 (2002), 265–274.

[3.1.10] T. Telliskivi, U. Olofsson, Contact mechanics analysis of measured wheel–rail profiles using the finite element method. Proc. Inst. Mech. Eng., Part F 215 (2001), 65–72.

[3.1.11] A. Sladkowski, M. Sitarz, Analysis of Wheel–rail interaction using FE software Wear 258 (2005), 1217–1223.

[3.1.12] W. Yan, F.D. Fischer, Applicability of the Hertz contact theory to rail-wheel contact problem, Archive of applied Mechanics 70 (2000), 255-268.

[3.1.13] A. Saulot, L. Baillet, Dynamic finite element simulations for understanding wheel–rail contact oscillatory states occurring under sliding conditions, J. Tribol. ASME 128 (2006), 761–770.

[3.1.14] M. Bijak-zochowski, P. Marek, Residual stress in some elasto-plastic problems of rolling contact with friction, J. Mech. Sci. 39 (1997), 15–32.

[3.1.15] Y.C. Chen, L. Chen, Effects of insulated rail joint on the wheel/rail contact stresses under the condition of partical slip, Wear 260 (2006), 1267–1273.

[3.1.16] J Sandström, A Ekberg, Numerical study of the mechanical deterioration of insulated rail joints, Proc. Inst. Mech. Eng. Part F: J. Rail Rapid Transit 223 (2009), 265–273.

[3.1.17] J. Zhang., S. Sun, X. Jin, Numerical simulation of two-point contact between wheel and rail, Acta Mechanica Solida Sinica 22 (2009), 352–359.

[3.1.18] Z. Li, X. Zhao, C. Esveld, R. Dollevoet, M. Molodova, An investigation into the causes of squats: correlation analysis and numerical modeling. Wear 265 (2008), 1349–1355.

[3.1.19] Z. Li, X. Zhao, R. Dollevoet, M. Molodova, Differential wear and plastic deformation as causes of squat at track local stiffness change combined with other track short defects, Veh. Syst. Dyn. 46 (Supplement 1) (2008), 237–246.

[3.1.20] X. Zhao, Z. Li, C. Esveld, R. Dollevoet, The effects of the tangential traction on wheel–rail dynamic contact, WSEAS Trans. Appl. Theor. Mech. 2 (2007), 52–59.

[3.1.21] R. Courant, K. Friedrichs, H. Lewy, On the partial differential equations of Mathematical Physics, Math. Ann. 100 (1928), 32 – 74.

[3.1.22] D.J. Benson, J.O. Hallquist, A single surface contact algorithm for the post-buckling analysis of shell structures, Computer Methods in Applied Mechanics and Engineering 78 (1990), 141-163.

3.2 A 3D finite element solution of frictional wheel–rail rolling contact in elasto-plasticity[2]

ABSTRACT

Wear, rolling contact fatigue and plastic deformation are the major failure modes of railway wheels and rails. Proper analyses of the failure mechanisms as well as improvement in design and maintenance require an accurate evaluation of the stress and strain states. Solution of frictional rolling contact between wheel and rail in elasto-plasticity seems, however, still to lack in the literature. This paper presents such a solution. A 3D finite element model is built up to simulate the rolling contact. The focus is on the tangential problem, namely the distributions of surface shear stress and micro-slip as well as the distinction of areas of adhesion and slip in the contact patch. With the presented model the assumptions of half space, linear elasticity and quasi-static state, which are often employed in existing solution of rolling contact, are dropped. A bi-linear elasto-plastic material with isotropic hardening is employed. It is found that when plastic flow occurs the contact patch in the rail top changes from an ellipse into an egg shape; the cross influence between the normal and tangential contact problems become stronger: the normal solution is not independent of the tangential solution any more, and the tangential solution is greatly affected by the normal solution. The model can be applied to investigation of the relationships between material properties, plastic deformation, frictional work, wear and crack initiation. Such relationships may help in better understanding the occurrence of corrugation, head checks, and squats, and may further be used for design of fatigue resistant materials and profiles of wheel and rail.

---

[2] This section has been submitted in 2012 for publication to Wear as: X. Zhao and Z. Li, A 3D finite element solution of frictional wheel–rail rolling contact in elasto-plasticity.

3.2.1 Introduction

Frictional rolling contact plays a critical role in many wheel–rail related problems such as wear, rolling contact fatigue (RCF), plastic deformation, vehicle dynamics and vehicle–track interaction. However, there are no yet effective experimental techniques for the determination of contact stress and strain states, especially under dynamic conditions. Therefore, emphasis has been placed on computer based numerical approaches. Hereinafter the frictional rolling contact between wheels and rails is referred to as rolling contact for short.

Traditionally rolling contact has been divided into the normal and the tangential problems for solution; the normal problem is solved first. This is justified because under elastic conditions with planar contact patch the friction force transmitted between the contact bodies usually has negligible influence on the shape of the contact patch and the distribution of pressure [3.2.1]. The division was necessary because of the limitations of the analytical and numerical methods employed. In the normal problem the contact patch and pressure are determined. In the tangential contact problem friction is considered, and a solution is sought for the distributions of surface shear stress and micro-slip, as well as the related areas of adhesion and slip in the contact patch.

Micro-slip is the relative velocity of two material particles in contact: one is in the rail surface and the other in the wheel. In the literature on rolling contact, 'micro-slip' is often called 'slip' for short [3.2.2]. Since 'slip' may also mean creepage in Railway Engineering and in Tribology [3.2.3], 'micro-slip' is used in this paper to avoid confusion. 'Slip area' is, however, used when the area of micro-slip is referred to.

The normal problem of contact between bodies of quadratic surfaces was solved analytically by Hertz [3.2.4]. Rolling contact was treated by Carter [3.2.5] in two dimensions, in which the surface shear stress in the contact patch satisfies the Coulomb friction law everywhere. By assuming an elliptical adhesion area, Vermeulen and Johnson [3.2.6] obtained an analytical solution for 3D rolling contact between bodies of quadratic surfaces without spin. In order to treat more general cases, the employment of numerical approaches seems inevitable. Currently, the available numerical approaches for contact problems are mainly the boundary element (BE) method and the finite element (FE) method.

Using a BE approach with analytical influence coefficients for half space, Kalker [3.2.2] has developed several numerical methods to solve the 3D rolling contact problem with arbitrary longitudinal and lateral creepages as well as spin, with the assumption of half-space contact bodies, linear elasticity and statics. The most advanced has been implemented in the computer code CONTACT, with which problems of arbitrary planar contact patch can be treated. Nowadays, CONTACT is widely employed to obtain detailed solutions of wheel–rail contact, when the tangential problem has to be solved.

Based on CONTACT, Li [3.2.7] extended Kalker's work from half-space to quasi-quarter spaces and from constant spin to varying spin for curved contact geometry like the conformal contact between wheel flange root and rail gage corner. It has been implemented in the program WEAR. Although the initial intention of the extension was

for severe wear simulation [3.2.7], it has recently been successfully applied to analysis of loading conditions for rolling contact occurrence [3.2.8] and for the design of an anti-head checking profile [3.2.9].

With the FE method, contact geometry can be simulated more realistically than with the BE method. Further, material non-linearity can be included in an FE model by employing a proper constitutive equation; dynamic behavior related to the contact can also be taken into account, for instance, by performing a transient analysis. FE solutions of normal contact problems in statics can be found in [3.2.10-3.2.14].

For frictional rolling the Arbitrary Lagrangian Eulerian (ALE) method, developed by Oden and Lin [3.2.15], Bass [3.2.16], Faria et al. [3.2.17], Nackenhorst [3.2.18], Ziefle and Nackenhorst [3.2.19], and Hu and Wriggers [3.2.20], has become established for tire–road contact where the tire is deformable and the road could be considered as being rigid. It offers a time independent formulation and is advantageous for stationary rolling [3.2.18].

For wheel–rail contact the deformability of both the wheel and the rail must be considered simultaneously due to their similar modulus. Furthermore, the contact surfaces may not be simulated as planes in the presence of geometry defects in the contact surfaces, in the case of squats, for instance. The ALE approach was applied to wheel–rail rolling in [3.2.21]. Remarkable difference, however, still exists compared to well established results. One problem is related to the regularization of stick-slip behavior. Regularization is employed to avoid non-differentiability of Coulomb's law at the onset of slip. The regularized constitutive interface laws physically have the drawback that they only describe the stick-slip motion in an approximate fashion. For a value of the regularization parameter that is too large, such a model may not be able to predict real stick-slip motions [3.2.22]. Future work on reliable, stable and efficient numerical algorithm is needed for solution of wheel–rail frictional rolling by ALE [3.2.23].

Solution of frictional rolling contact between wheel and rail in elasto-plasticity seems still lack in the literature. In the present work, a 3D FE model is employed to solve the wheel–rail rolling contact problem in elasto-plasticity. The model was developed from the model presented in [3.2.24] which simulated the high frequency vehicle–track interaction at squats and was validated in [3.2.25] against Hertz's theory and Kalker's CONTACT for quasi-static elastic rolling for both the normal and tangential solutions. This paper studies in details the influence of plastic deformation on both the normal and the tangential solutions. The rolling of a wheel on a smooth rail, with Coulomb friction in between, is considered under different traction levels. The elasto-plasticity of wheel–rail material is taken into account with a bi-linear material constitutive relation with isotropic hardening. At the end of the paper, features of the employed FE approach are summarized and conclusions are drawn.

3.2.2   The FE model

A schematic diagram of the transient FE model is shown in Figure 3.2.1. The wheel is represented by a cylinder with a radius of 460 mm, while the rail top is assumed to have a constant radius of 300 mm in the lateral direction. The contact occurs in the middle of the

rail top. The contact surfaces of both the wheel and the rail are smooth. As illustrated in Figure 3.2.1, a 3D right-handed Cartesian coordinate system (*Oxyz*) is defined, of which the origin *O* is located at the centre of the solution zone (position B) and in the middle of rail top. The *x* axis is defined along the rolling direction (i.e. the longitudinal direction), and the *z* axis is in the vertical direction. A solution is output for analysis at the instant when the centre of the contact patch is at *O*.

An explicit time integration scheme with a central difference method is employed to perform the transient analyses. To ensure convergence, the Courant stability condition [3.2.26] must be fulfilled. This requires that the time step is small enough so that sound wave may not cross the smallest element within one time step. This also means that as long as the elements are sufficiently small the FE model will automatically include in its solution all the relevant vibration modes of the structure and the continua. To minimize such vibrations to approximately achieve a quasi-static rolling state, the rail head is supported continuously on a rigid foundation (see Figure 3.2.1), and a dynamic relaxation is applied.

The wheel and the rail head are meshed with 8-node solid elements. In order to achieve a high accuracy of the solution with a reasonable model size, non-uniform meshing is used, with the finest element in the solution zone, as shown in Figure 3.2.2. For this work, the dimension of the faces of the finest elements in the contact surface is 0.33×0.33 mm. The wheel–rail interaction is calculated using a surface-to-surface contact algorithm [3.2.27].
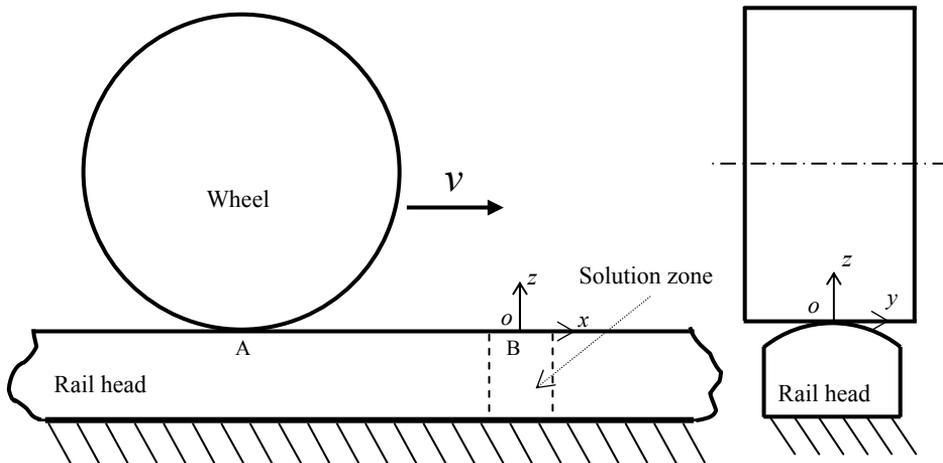
Figure 3.2.1. A schematic diagram of the 3D transient FE model. The point of first contact (assuming the normal load is zero) is in the lateral center of the rail top (y = 0).

The model parameters are listed in Table 3.2.1. Hardening of the wheel and rail materials, modeled with a constant tangent modulus in a bi-linear material model, is assumed to be isotropic. For comparison, the corresponding elastic solutions are also obtained, which are different from the elasto-plastic ones only in that no plasticity is included in the material model. A relatively high friction coefficient of 0.5, which

corresponds to dry and clean wheel–rail contact condition [3.2.28], is chosen in order to allow a wide range of friction exploitation level and a proper amount of plastic deformation in the solutions for discussion. Different friction exploitation levels are simulated by specifying the corresponding traction effort, varying from zero to full level. Coulomb's law of friction is employed. No lateral creepage or kinematic spin is involved. Spin due to contact geometry is automatically taken into account by the solution method.
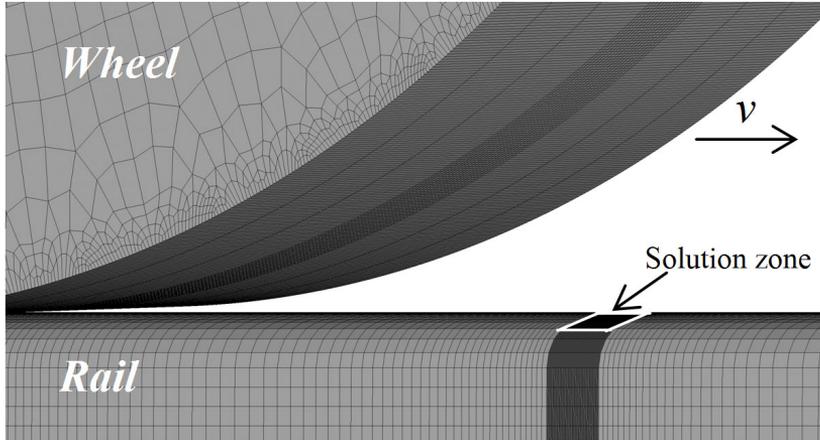
Figure 3.2.2. Meshing of the 3D transient FE model.

The Coulomb's coefficient of friction (*f*) is distinguished from friction exploitation level which is measured by traction coefficient (μ). The Coulomb's law is applied locally so that in each element in contact

$$|\boldsymbol{\tau}| \leq fp \qquad (3.2.1)$$

where p is the pressure and $\boldsymbol{\tau}$ is the surface shear stress. In contrast, the traction coefficient is defined as

$$\mu = F_L / F_N \leq f \qquad (3.2.2)$$

where, F_N is the total normal force and F_L is the resultant tangential (creep) force.

Table 3.2.1 Values of parameters involved

Parameters		Values
Wheel diameter, ϕ (m)		0.92
wheel load, F_w (kN)		140
Friction coefficient, f		0.5
Wheel–rail material	Young's modulus, E (GPa)	210
	Poisson's ratio, υ	0.3
	Density, ρ (kg/m³)	7800
	Yield strength, σ_Y (MPa)	500
	Tangent modulus, T (GPa)	21

3.2.3 Solutions in elasticity and elasto-plasticity

In this section a tractive case is simulated with a traction coefficient of 0.3 as a median value. Note that the results shown below are all for the rail.

3.2.3.1 Normal contact

a) Contact patch

The contact patch is determined as follows:

$$\text{A surface node is in the contact patch if:} \qquad F_{n\_N} > 0 \qquad (3.2.3)$$

Where, F_{n_N} is the nodal force in the direction normal to the local surface, being non-negative because no tension is transmitted through contact; n denotes nodal number.

Figure 3.2.3 shows the obtained contact patches. It is seen that, compared to the corresponding elastic solution, the contact patch changes from an ellipse to an egg shape due to plastic deformation, i.e. the trailing edge becomes more obtuse. At the same time, the area of the contact patch is increased and its longitudinal position is slightly shifted forward.
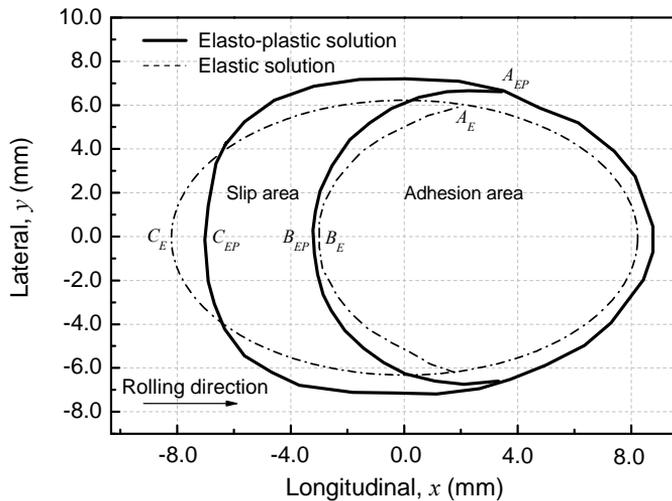
Figure 3.2.3. Changes of the contact patch and the slip-adhesion areas due to plastic deformation, (taken from Figure 3.2.9).

In order to explain the difference between the contact patches of the elastic and elasto-plastic solutions, the contact geometry variation caused by plastic deformation need be examined. Figure 3.2.4(a) shows the vertical plastic deformation of the rail top along the longitudinal axis of the contact patch (at $y = 0$, it is called the longitudinal axis hereinafter). Here, the vertical plastic deformation is defined as the permanent displacement in the vertical direction (U_{Pz}):

$$\text{At a node:} \qquad U_{Pz} = U_{EPz} - U_{Ez} \qquad\qquad (3.2.4)$$

where, U_{EPz} and U_{Ez} are the vertical displacements in the elasto-plastic and elastic solutions, respectively.

As anticipated, negative plastic deformation occurs in the vertical direction as the wheel passes, i.e. the rail surface sinks. Following a material particle P which flows through the contact patch (see Figure 3.2.4(a)), its deformation is purely elastic before entering the contact at P_1 and in the front part of the contact patch between P_1 and P_2. At P_2 the yield stress σ_Y is reached. Subsequently, plastic deformation accumulates when P flows from P_2 till P_3 due to continuous rise in the effective stress (see Von-Mises (V-M) stress results shown in Figures 3.2.12 and 3.2.17). Correspondingly, the contact surface descends from P_2 to P_3. At P_3 the effective stress starts to decrease, so that no further plastic deformation occur between P_3 and P_4. At P_4 the particle exits the contact, while the plastic deformation remains. The dashed line in Figure 3.2.4(a) shows the corresponding total deformation U_{EPz}, the ordinate of which is on the right side of Figure 3.2.4(a). Obviously the elastic deformation component is much larger than the plastic one. Figure 3.2.4(b) shows the effective plastic strain along the cross-section of $x = -6.0$ mm.

The contact geometry change caused by plastic deformation may be plotted schematically in Figure 3.2.5. The dashed line represents the elastic situation, in which the rail surface has a constant height. In contrast, when elasto-plastic material is used, the height of the rail surface is reduced after contact; see the solid line. Assuming that a contact compression of C_0C_E is required to sustain the wheel load under the elastic condition and the contact patch of $P_{1E}P_{4E}$ is resulted, a larger contact compression C_0C_{EP} is needed to balance the same wheel load in the elasto-plastic solution. Consequently, the contact patch in the elasto-plastic solution ($P_{1EP}P_{4EP}$) shifts forward with respect to the one under the elastic condition ($P_{1E}P_{4E}$). Note that P_{4EP} is always on the right side of P_{4E}. This is because the plastic deformation (U_{P2}) is greater than compression difference ΔU:

$$\Delta U = C_0 C_{EP} - C_0 C_E = U_{P2} + U_{E2} - U_{E1} \qquad\qquad (3.2.5)$$

where $U_{E2} < U_{E1}$ due to the existence of plastic deformation. Only when U_{P2} equals ΔU, being impossible if plastic deformation exists, P_{4EP} has the same longitudinal position as P_{4E} at the position of $P_{4EP}^{\#}$.

On the other hand, it can also be expected that the contact patch is enlarged by plastic deformation in the lateral direction due to the increased contact compression. Furthermore, the trailing edge of the contact patch is more obtuse in the elasto-plastic solution than in the elastic solution, leading to the fact that the trailing edge shifts forward more than the leading edge in the longitudinal direction when plastic deformation occurs, i.e. a contact patch of egg shape is generated. The reason behind this phenomenon is the conformity increase between the contact bodies after plastic deformation. All the above-discussed changes of contact patch can be observed in Figure 3.2.3.
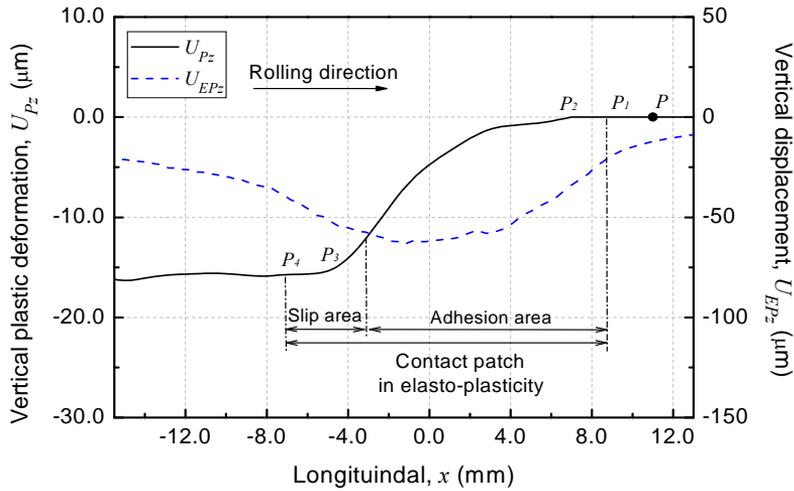
(a)
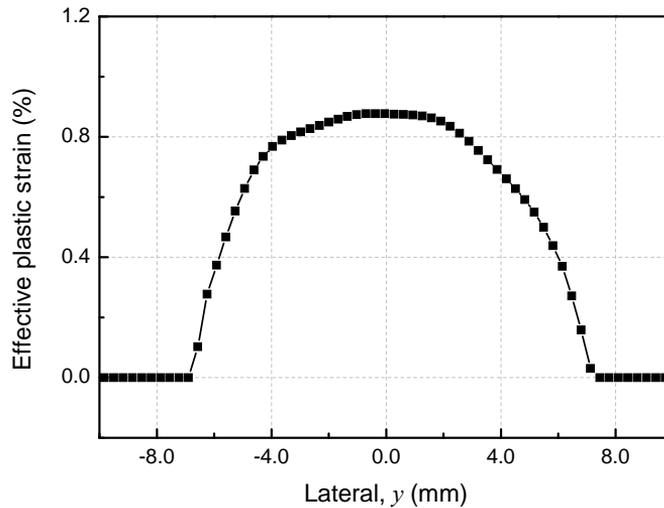
(b)

Figure 3.2.4. Plastic deformation of the rail top. (a) Vertical plastic deformation along y = 0.0 mm and (b) effective plastic strain along x = −6.0 mm.

It should be noted that the plastic deformation distribution shown in Figure 3.2.4(a) slightly fluctuates in the zone where the wheel has passed. This is attributed to the high frequency vibrations involved in transient simulations. Since such high frequency vibrations are inherent to rolling contact between bodies of continua, the solution by the FE method should represent the reality. As will be seen later, fluctuations caused by the same reason also exist in other results. In addition, the plastic deformation of the wheel is ignored in Figure 3.2.5 to ease the explanation. Further discussion about this point will be given in Section 3.2.5.
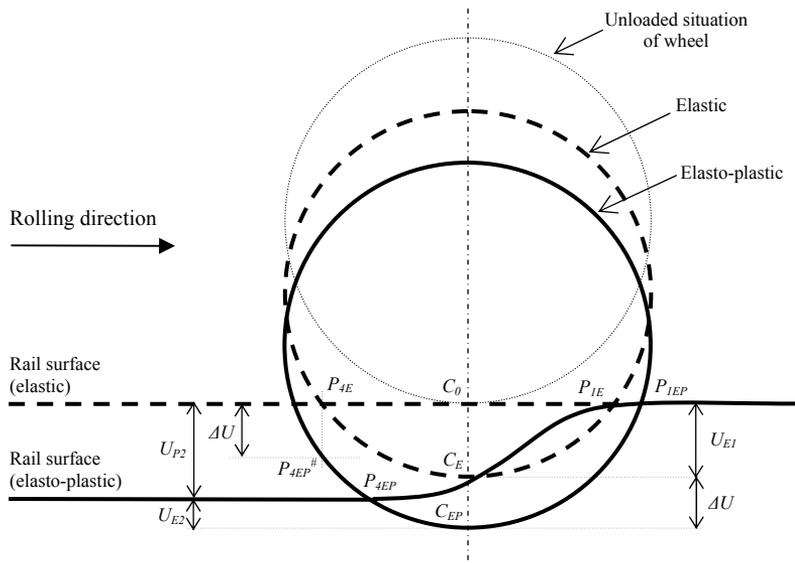
Figure 3.2.5. Schematic comparison of elastic and elasto-plastic rolling contact. Elastic deformation is ignored.

b) Pressure distribution

The pressure contours of the elastic and elasto-plastic solutions are given in Figure 3.2.6 for comparison. It is seen that the maximum pressure is significantly reduced by plastic deformation. The enclosed area of the contour line of 1.1 GPa is much larger in the elastic solution, and the area with a pressure higher than 1.3 GPa does not exist in the elasto-plastic solution (Figure 3.2.6(b)). In the meantime, the shape of the contour lines is also changed in the way of the contact patch does; compare for example the contour lines of 180 MPa in Figure 3.2.6. Note that the contour lines are smoother under the elasto-plastic condition due to the fact that the high frequency vibrations, the cause of the fluctuation as mentioned above, is attenuated to a certain extent by plasticity.

For more detailed comparison, the magnitude distributions of pressure along the longitudinal axis $y = 0$ are shown in Figure 3.2.7(a). It can be seen that for the elasto-plastic solution the pressure is lower in the rear part than that of the elastic solution, whereas it is higher in the front part due to the forward shift of the contact patch. The maximum pressure is reduced and shifted forward (see also Figure 3.2.6). The forward shift of the contact patch can also be seen in Figure 3.2.7(a).

Figures 3.2.7(b) and 3.2.7(c) further show the pressure distributions along two lateral cross-sections at which the longitudinal coordinates are $x = -4.3$ and 5.7 mm, respectively. In the rear part of the contact patch (Figure 3.2.7(b)), the pressure in the middle is reduced by plastic deformation, while increased on both sides. This is attributed to the increased conformity and the consequent wider contact patch. In contrast, the pressure is increased in the whole range of the front part of the contact patch (Figure 3.2.7(c)) because of the forward shift of the contact patch.
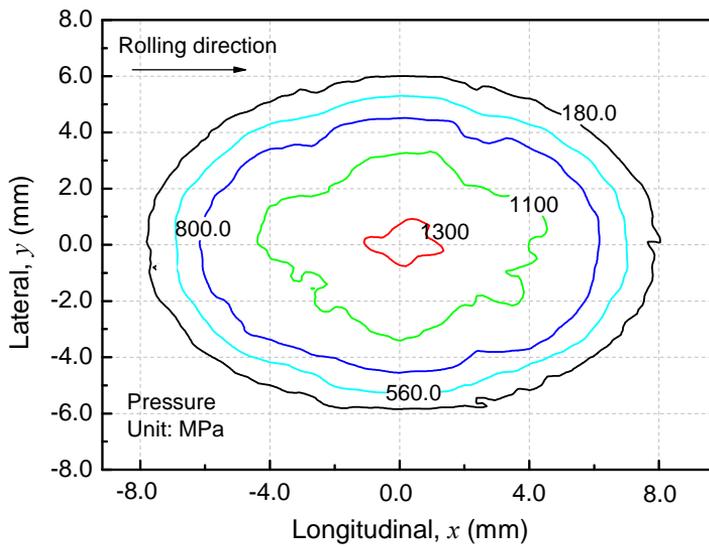
(a)
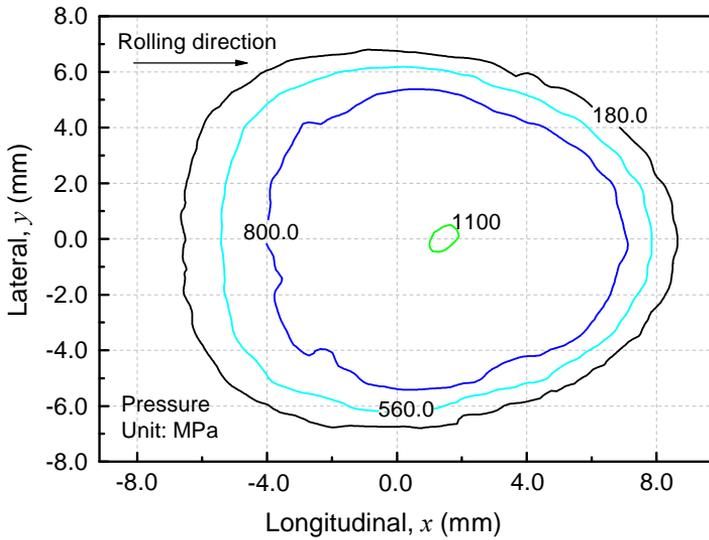
(b)

Figure 3.2.6. Pressure contours in the contact patch. (a) Elastic solution and (b) elasto-plastic solution.
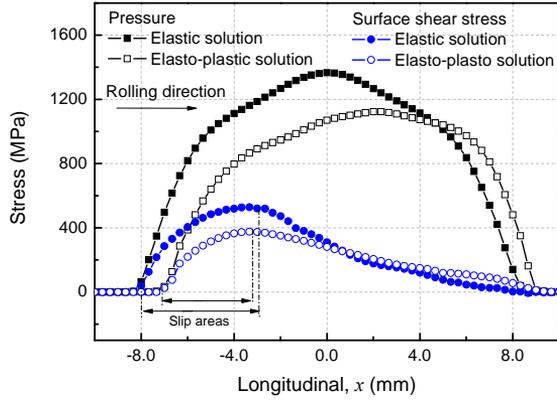
(a)
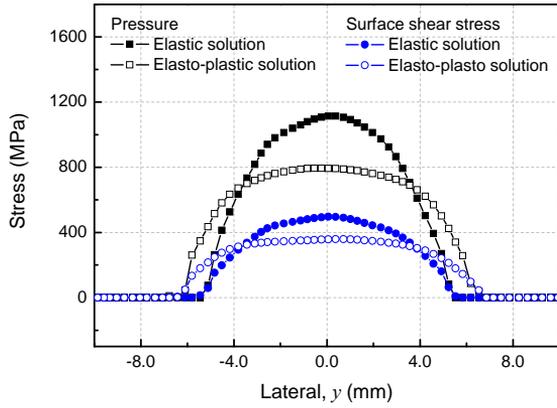
(b)
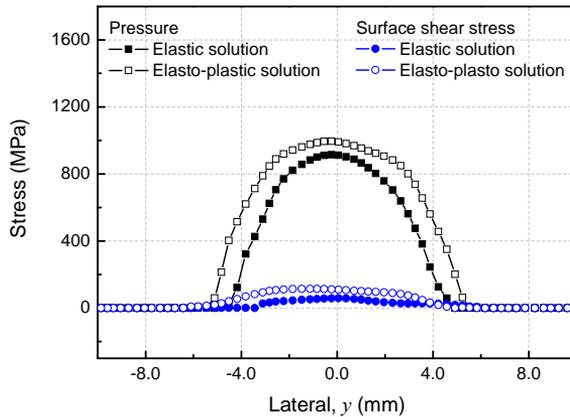
(c)

Figure 3.2.7. Distributions of pressure and surface shear stress (magnitude) in the contact patch. (a) Along the longitudinal axis ($y = 0$), (b) along the lateral cross-section of $x = -4.3$ mm and (c) along the lateral cross-section of x = 5.7 mm.

3.2.3.2 Tangential contact

a) Areas of adhesion and slip

The areas of adhesion and slip are determined according to Coulomb's law:

A node is in adhesion if: $f\,F_{n\_N} - |F_{n\_T}| > \varepsilon_T$ (3.2.6)

A node is in slip if: $f\,F_{n\_N} - |F_{n\_T}| \leq \varepsilon_T$ (3.2.7)

where F_{n_T} is the nodal forces in the tangential direction and $\varepsilon_T > 0$ is a tolerance. In the present work, ε_T is 0.3% of the maximum tangential nodal force in the contact patch. In other words, a node could be on the wrong area of slip or adhesion if the difference between the tangential nodal force and its limit determined by the Coulomb's law is close to 0.3% of the largest tangential nodal force in the contact patch.

The division of the areas of slip and adhesion is presented in Figure 3.2.8. The adhesion area is also enlarged by the plastic deformation. The size of the slip area is, however, almost the same for the two cases (the numbers of nodes in slip are 557 and 558, respectively). A block or bullet in the figure represents a node in the slip area or adhesion area. The "wrong" block of slip around the border between the areas of slip and adhesion, and the "wrong" bullets of adhesion in the slip area near the border of the contact patch are due to numerical error and ε_T. This error seems quite reasonable and acceptable as is discussed in [3.2.25].

To be more detailed, the area of slip as a whole is stretched a bit in the rolling direction from $x = [-8.0, 1.3]$ mm (between A_E and C_E in Figure 3.2.3) in the elastic solution to $x = [-7.0, 3.0]$ mm (between A_{EP} and C_{EP} in Figure 3.2.3) in the elasto-plastic solution. Along $y = 0$ where the stress is the highest, the slip area shrinks from $x = [-8.0, -3.0]$ mm (between B_E and C_E in Figure 3.2.3) to $x = [-7.0, -3.3]$ mm (between B_{EP} and C_{EP} in Figure 3.2.3).

b) Surface shear stress distribution

Figure 3.2.9 shows the vector graph of the surface shear stress in the contact patch. The arrows point in the direction of the stress and their lengths are proportional to its magnitude. The border of the slip and adhesion areas is also plotted. Although the shape of the contact patches is different, the direction of the stress in the two plots is similar – it is basically in the opposite direction of rolling since no lateral force is applied. The stress on the two sides has small lateral components towards the center line $y = 0$ due to the small geometry spin caused by the circular contact surface of the rail. They cancel each other so that equilibrium in the lateral direction is maintained.

The magnitude variation of the surface shear stress in the longitudinal and lateral directions is plotted in Figure 3.2.7 together with the pressure distributions. The surface shear stress is, like the pressure, reduced by plastic deformation in the rear part of the longitudinal axis where slip occurs (Figure 3.2.7(a)). This is attributed to the fact that in the slip area where the friction reaches saturation the magnitude of the surface shear stress is only determined by the pressure when the friction coefficient is constant. For the

same reason, the magnitude of the surface shear stress in the slip area follows a similar relation with the pressure along a lateral direction, as shown in Figure 3.2.7(b).

(a)

(b)
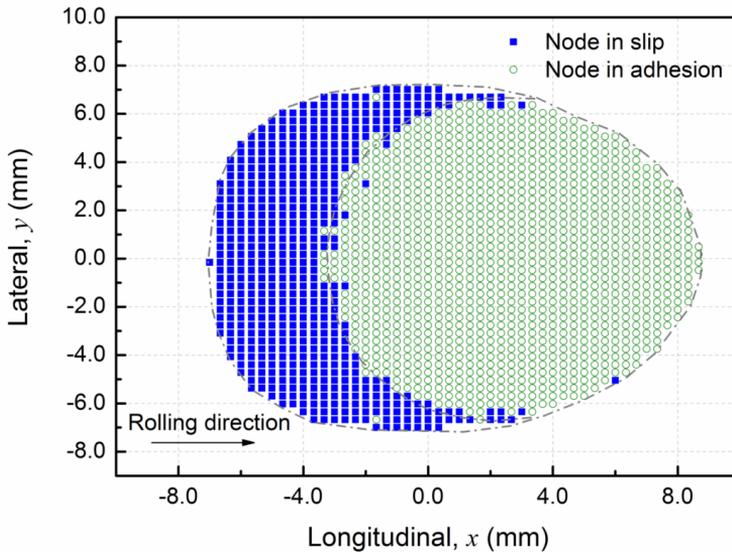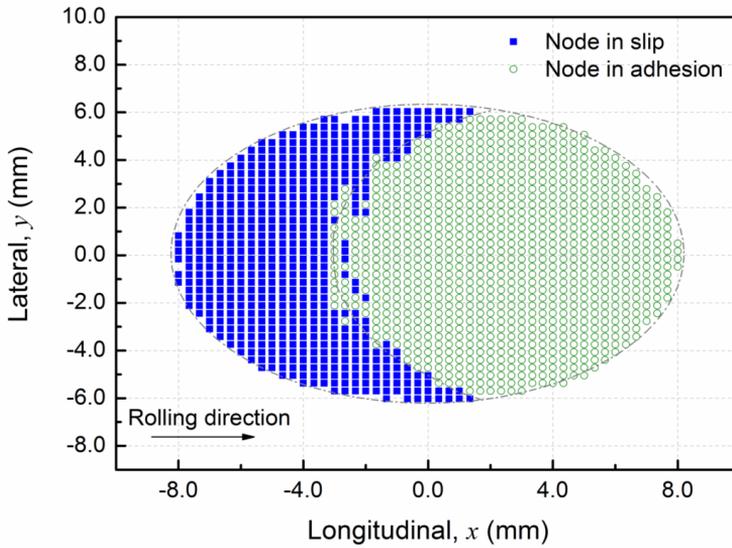
Figure 3.2.8. Slip-adhesion area distribution. (a) Elastic solution and (b) elasto-plastic solution.

In the adhesion area the surface shear stress monotonically decreases along the rolling direction for both the elastic and elasto-plastic solutions, see Figure 3.2.7(a), while the pressure distribution is not monotonous. Since the maximum surface shear stress, which is at the slip-adhesion border, is higher in the elastic solution and the adhesion area

is longer under elasto-plastic condition, the surface shear stress descends slower when plastic deformation occurs, as shown in Figure 3.2.7(a). Figure 3.2.7(c) shows the results at $x = 5.7$ mm, where the surface shear stress is increased by the forward extension of the contact patch and the consequent extension of the pressure distribution, see Fig 7(a).
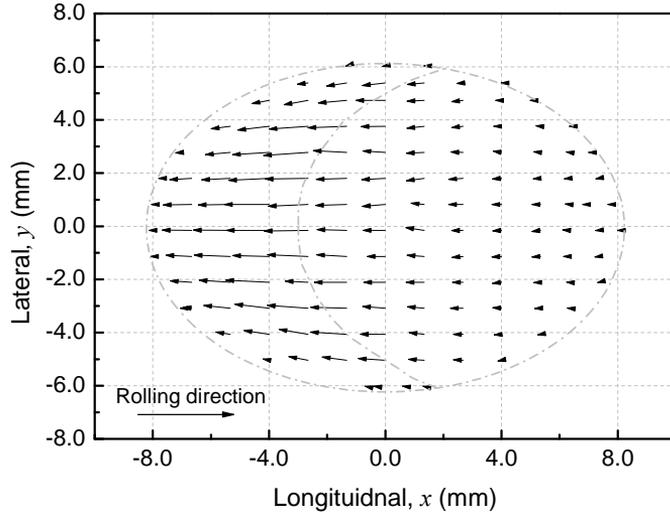
(a)
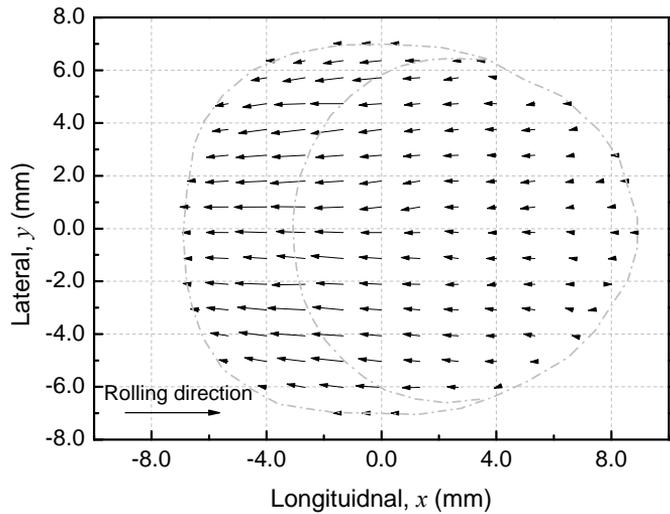
(b)

Figure 3.2.9. A vector graph of surface shear stress. (a) Elastic solution and (b) elasto-plastic solution.

c) Micro-slip distribution

The vector graphs of the micro-slip in the contact patch are shown in Figure 3.2.10. The arrows are in the opposite direction of the surface shear stress of Figure 3.2.9. Magnitude

distribution of the slip in the longitudinal axis is presented in Figure 3.2.11(a). As expected, it decreases in magnitude along the rolling direction and becomes zero in the adhesion area. In the lateral direction it is shown in Figure 3.2.11(b) at three cross-sections with $x = -0.3$, -3.4 and -5.5 mm, respectively. Due to numerical error the curves are not completely symmetric about $y = 0$.

It is seen in Figure 3.2.11(b) that the micro-slip is not uniform in the lateral direction, but has a saddle-shape. One reason for this is that the pressure and the surface shear stress both reach their maxima at $y = 0$, see Figures 3.2.7(b) and 3.2.7(c), so that relative motion of contacting particles is the most constrained there. Towards the sides of the saddle the micro-slip grows because the constraint by the contact stresses becomes weak. As a cross-section in the slip area approaches the trailing edge of the contact patch, the depth of saddle decreases as shown in Figure 3.2.11(b) owing to the release of constraints.

The saddle is deepened by plastic flow, which takes place in the opposite direction of micro-slip and therefore partly cancels it (compare the elastic and elasto-plastic solutions in Figure 3.2.11). Therefore, the adhesion area extends and the trailing edge of the adhesion area shifts backward when plastic deformation occurs (Figure 3.2.3 and 11(a)). As plastic deformation is the largest at $y = 0$ (Figure 3.2.4(b)), the cancellation of micro-slip by plastic flow, i.e. the difference between the elastic and elasto-plastic solutions, is the most there.

It should be noted that plastic flow does not occur in the whole contact patch. As explained in Figure 3.2.4(a), it starts at P_2, accumulates till P_3 is arrived at, and then the plastic deformation remains in P_3P_4. Therefore, plastic flow is only one of the reasons for the change in the micro-slip distribution shown in Figures 3.2.10 and 3.2.11. Another influential factor for the change is the modification of contact patch and the slip area by the plastic deformation. Such an influence can be understood by looking at the linear theory of Kalker [3.2.2]: the creepage, which is a macro embodiment of the micro-slip, reduces with the increase of contact patch in size if the shape of the contact patch and the transmitted friction force both stay constant.

For the problem at hand, towards the sides of the saddle the micro-slip grows more under the elasto-plastic condition than under the elastic condition (see Figure 3.2.11(b)), due to the larger contact patch and lower pressure. Thus, the highest value of the micro-slip in the elasto-plastic solution in each of the three cross-sections is always larger than that of the corresponding elastic solution. The larger micro-slip on the sides of the saddle, together with the wider contact patch and its more obtuse trailing edge, explains the changes of slip area discussed in Section 3.2.3.2 *a)*.

(a)

(b)
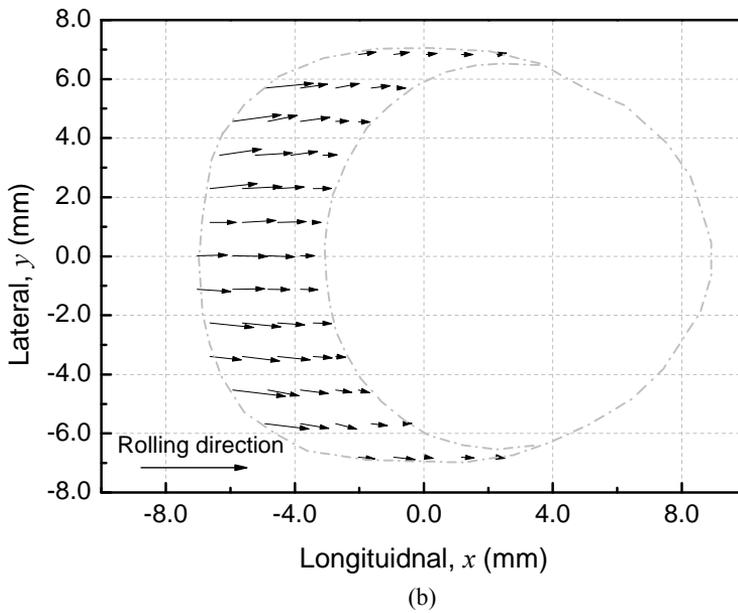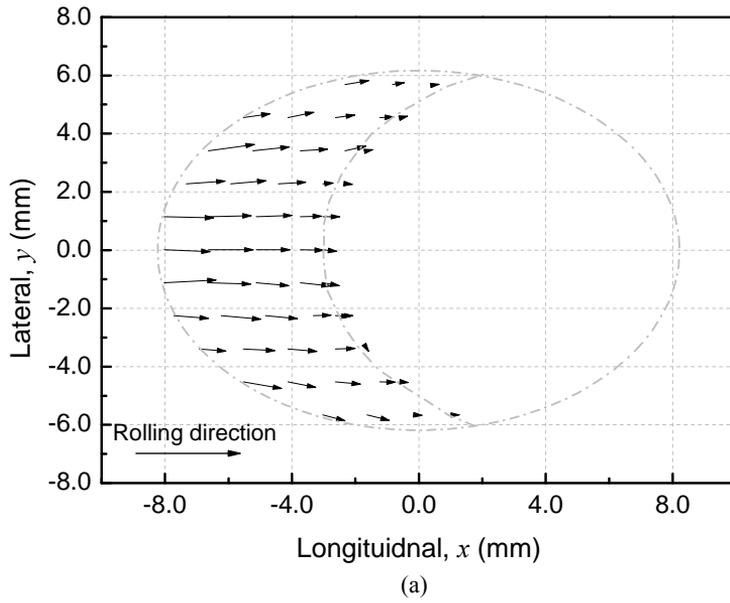
Figure 3.2.10. A vector graph of micro-slip. (a) Elastic solution and (b) elasto-plastic solution.
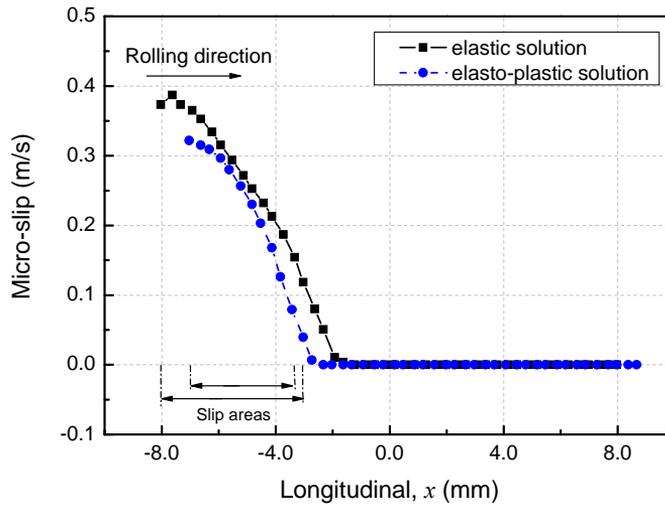
(a)
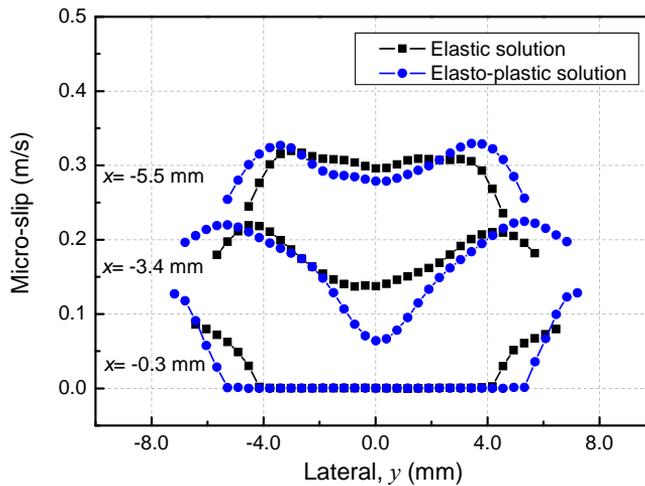
(b)

Figure 3.2.11. Distribution of micro-slip (magnitude). (a) Along the longitudinal axis (y = 0) and (b) along the lateral cross-sections of x = −0.3, −3.4 and −5.5 mm.

3.2.3.3   V-M stress

As anticipated, the V-M stress distribution on the contact surface of the rail is significantly modified by the occurrence of plastic deformation (see Figure 3.2.12). The maximum V-M stress in the contact patch is considerably lower in the elasto-plastic solution than in the elastic one. Residual stress is clearly left behind: in Figure 3.2.12(a) the stress returns to zero after a wheel passage, i.e. at $x < −15.0$ mm, while in Figure 3.2.12(b) the maximum residual stress remains at about 360 MPa (see also Figure 3.2.18).
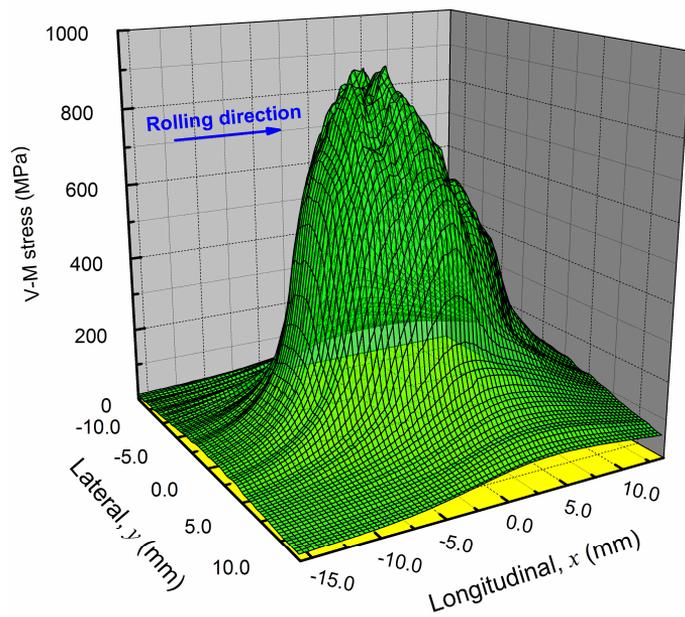
(a)
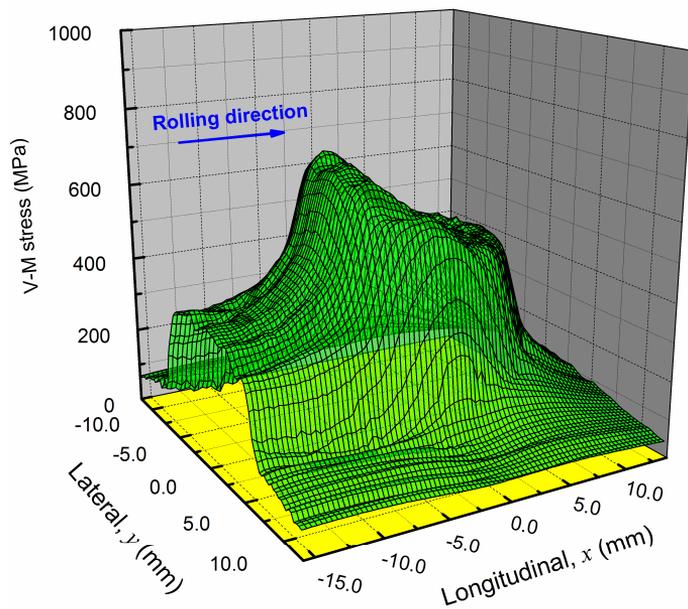
(b)

Figure 3.2.12. V-M stress distribution on the rail surface (the range of the contact patch can be seen from Figure 3.2.3). (a) Elastic solution and (b) elasto-plastic solution.

Further, the longitudinal location of the maximum V-M stress moves backward in the elasto-plastic solution compared to that in the elastic solution, as shown in Figure 3.2.12. This is mainly because that the effective stress that the rail surface material can sustain under the elasto-plastic condition increases from P_2 to P_3 (Figure 3.2.4(a)) due to hardening of the material. In addition, the change in the contact geometry and the generation of residual stress also play a role. More discussions about the V-M stress distribution are presented later in Section 3.2.4.4. It should be noted that the V-M stress results of the contact surface shown in this paper are all at 0.7 mm below surface, where the integration points of the surface layer of elements are located.

3.2.4 Different friction exploitation

Friction exploitation level can greatly vary between wheels and rails. Two cases with traction coefficients of 0 and 0.5 are studied in this section. All the other parameters are kept the same as for the preceding section.

3.2.4.1 Contact patch

The contact patches of the three cases are plotted in Figure 3.2.13. The obtained contact patches coincide with each other under the elastic condition, which confirms that the normal solution for concentrated contact problem is independent of the friction force transmitted in the contact patch if the two bodies are elastically similar [3.2.1]. However, when plasticity is taken into account, the contact patch tends to become larger in size and shift forward more with the increase in traction coefficient. This is attributed to the fact that plastic deformation, which modifies the contact patch as discussed in Section 3.2.3.1. *a)*, increases in magnitude with traction coefficient (see Figure 3.2.14).

Furthermore, it can be observed from Figure 3.2.13 that the contact patch is very close to each other for the cases with traction coefficients of 0.3 and 0.5 under the elasto-plastic condition, especially in its width. Taking into account errors related to the discrete characteristic of the FE approach, it may be concluded that with the increase in traction coefficient, its influence on the contact patch gradually decreases under the elasto-plastic condition.
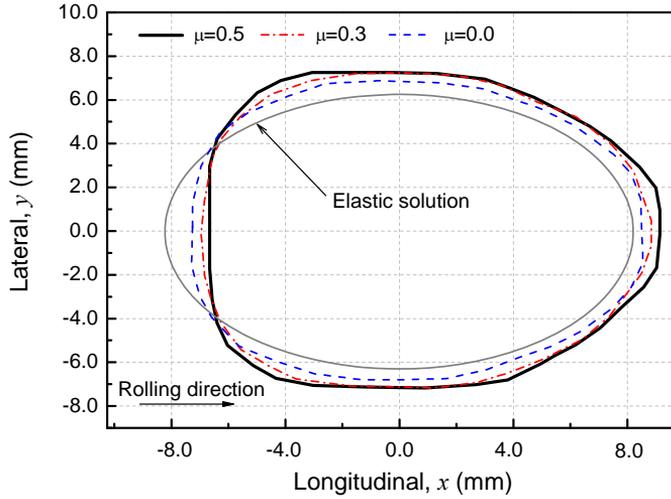
Figure 3.2.13. Variation of contact patch with traction coefficient. The contact patches coincide with each other under the elastic condition.
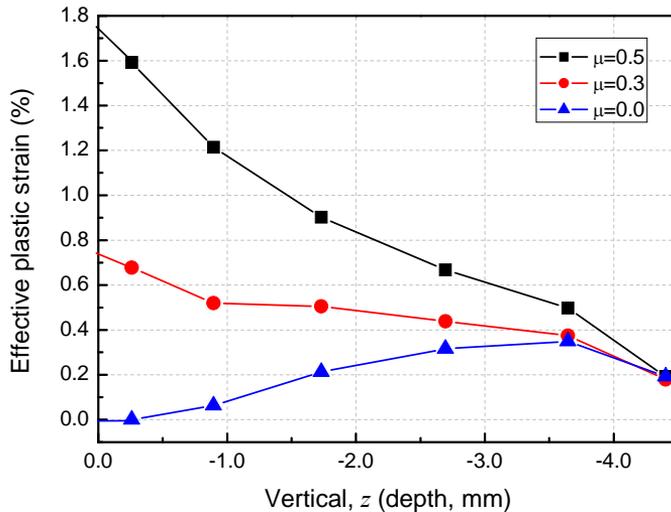
Figure 3.2.14. Distributions of effective plastic strain after a wheel passage along depth z at x = 0, y = 0.

3.2.4.2   Pressure

Figure 3.2.15 shows the pressure distributions along $y = 0$ under the different traction levels. Like for the contact patches in Figure 3.2.13, the elastic solutions also coincide with each other. For the elasto-plastic solutions the maximum pressure decreases in magnitude with the increase in traction force and shifts forward. This can be explained by the mechanism discussed in Section 3.2.3.1 *b)* with the consideration of the different plastic deformation levels.
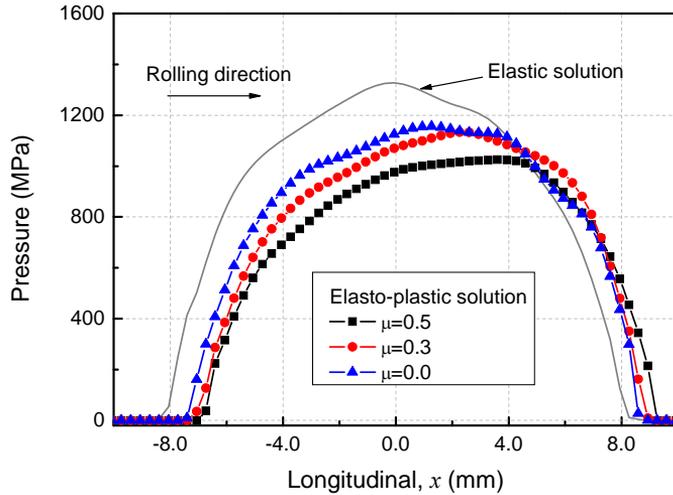
Figure 3.2.15. Pressure distribution along the longitudinal axis (y = 0) when traction coefficient changes.

3.2.4.3   Surface shear stress

The surface shear stress distributions along the longitudinal axis are plotted in Figure 3.2.16. The maximum surface shear stress is significantly reduced by plastic deformation when $\mu = 0.3$ and 0.5. The reason for this has been discussed in Section 3.2.3.2 *b)* For the case of $\mu = 0.0$, however, the influence of plasticity is not noticeable because adhesion area covers the entire contact patch. It should be mentioned that there is still non-zero surface shear stress in the case of $\mu = 0.0$, which ought to be attributed to the high frequency vibrations involved in the simulations and maybe to some extent to the geometry dissimilarity between the contact surfaces.

Furthermore, the surface shear stress distributions of $\mu = 0.3$ and 0.5 coincide with each other in the slip area under the elastic condition when taking numerical error into account, as it should be with Coulomb's law of friction. Nevertheless, for the elasto-plastic solution the surface shear stress is lower in the case of $\mu = 0.5$ at the same location due to the higher plastic deformation.
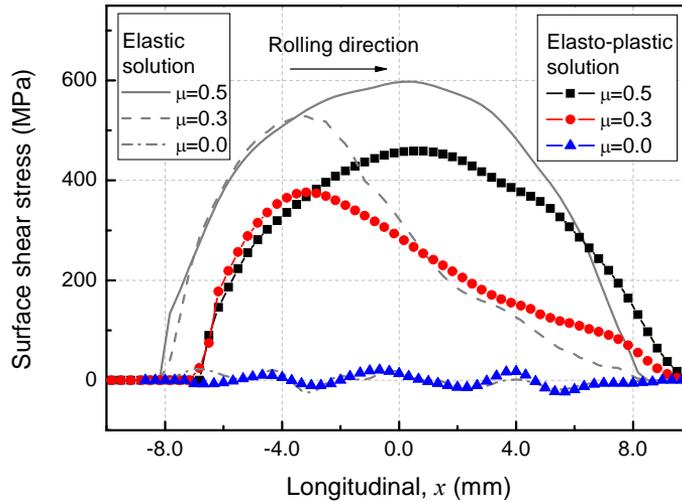
Figure 3.2.16. Surface shear stress distribution along the longitudinal axis (y = 0) when traction coefficient changes.

3.2.4.4 V-M Stress

The V-M stress distributions along the longitudinal axis $y = 0$ are shown in Figure 3.2.17. Their maxima increase with traction coefficient for both the elastic and elasto-plastic solutions. When $\mu = 0$, the elastic distribution is approximately symmetric about $x = 0$. For the elasto-plastic case the maximum shifts forward, due to the same reason of the forward shift of the maximum pressure (see Figure 3.2.6(b)). In both cases the distribution is somewhat distorted by the dynamic effect of the model.

For $\mu = 0.3$ and 0.5 under elastic condition, the distributions of the V-M stress follows roughly those of the surface shear stress, compare Figure 3.2.17 with Figure 3.2.16. This is determined by the fact that the V-M stress results shown in the figure are at depth close to the surface (0.7 mm below surface as mentioned above). The location of the maximum V-M stress shift forward when the slip area extends from the trailing edge towards the leading edge with the increase in traction coefficient. Under elasto-plastic condition, however, the maximum V-M stress moves backward. This can be attributed to work hardening of material. For example, the plastic deformation reaches its maximum at $x = -5.3$ mm (P_3 in Figure 3.2.4(a)) when $\mu = 0.3$, and the maximum V-M stress is also at $x = -5.3$ mm (see Figure 3.2.17). As high traction coefficient will lead to more plastic flow, the location of P_3 will correspondingly shift backward and the material there will be harder. Therefore, the maximum V-M stress of $\mu = 0.5$ is larger than that of $\mu = 0.3$, and is located more closely to the trailing edge. At the same time, larger deformation also leaves higher residual stress behind, see Figure 3.2.17.

In Figure 3.2.14 it is observed that sub-surface plastic deformation may occur even if surface deformation is zero. This resulted in changes of contact patch and V-M stress distribution when $\mu = 0$ (Figures 3.2.13 and 3.2.17, respectively), i.e. when the surface shear tress should theoretically be zero, and is actually much lower than the yield stress.
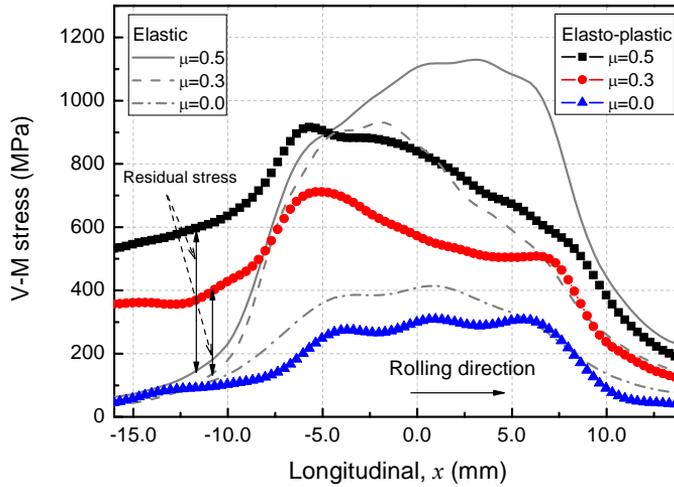
Figure 3.2.17. V-M stress distribution along the longitudinal axis (y = 0) when traction coefficient changes.

3.2.5 Discussions

So far in this paper, discussion has been made only for the rail, though the same material is applied to the wheel. Due to the action-reaction relationship, the observed physical phenomena are also applicable to the wheel. By using different material model, the solution method can also be used to cases where one of the contact bodies is elasto-plastic and the other is purely elastic.

It is observed that plastic deformation significantly increases with traction force, especially in the surface layer, see Figure 3.2.14. This leads to the dependence of normal solution on the friction force transmitted in the contact patch, see Figures 3.2.13 and 3.2.15. In contrast, the normal contact and tangential contact can be decoupled under the elastic condition. Owing to the non-linear characteristic of the wheel–rail contact and the material non-linearity involved, the dependence is also non-linear.

Although residual stress is computed as a result of the simulated rolling contact, pre-existing residual stress has not been considered yet. Such residual stress is important in cyclic plasticity and the related fatigue analysis. To simulate cyclic loading, computation time has to be considered. Currently, a typical simulation of the presented model, which simulates one wheel passage, takes about 27 hours for a PC workstation of Inter Pentium 4 CPU 3.73 GHz (3.5 GB of RAM). For future applications of the FE model on cyclic loading, computers with higher performance have to be employed to make the computation time acceptable. Note that the calculated residual stresses in this paper may not be realistic because the lateral shift of contact patch in different wheel passages is not considered and a simplified material model (i.e. a bi-linear model) is employed.

In elasticity, energy is dissipated in the contact patch only by frictional work which is proportional to the product of micro-slip and surface shear stress. In elasto-plasticity the dissipation consists of two parts: frictional work and work of plastic deformation. Frictional work causes wear of the contact surfaces. Plastic deformation may on one hand lead to fatigue crack initiation, and on the other hand reduce wear level by increasing the contact conformity. When the plastic flow is large enough, wear level becomes insufficient to prevent the occurrence of crack initiation from the rail surface, and RCF such as head checks and squats may occur. Although the significant influence of plastic deformation on surface shear stress and micro-slip has been presented above, its quantitative effect on frictional work is not included in this paper. In the future, by further calculating the frictional work for cyclic plasticity, the presented method can be applied to understand the relationships between material properties, plastic deformation, frictional work, wear and crack initiation, especially the competition between wear and crack initiation. Such relationships may help in better understanding the occurrence of head checks, as well as short pitch corrugation and squats, and may further be used for design of fatigue resistant materials and profiles of wheel and rail.

Besides the traction force, work hardening and residual stresses, many other factors can also considerably affect the magnitude of plastic deformation and its distribution, e.g. complex contact geometry, axle load level, the friction coefficient between wheel and rail, vehicle dynamics and vehicle–track interaction. In applying the 3D FE model of this paper to investigations of the occurrence and growth of squats, all the above-listed factors have been properly included according to the loading conditions at squats. The simulations unveiled some interesting phenomena about squats [3.2.24, 3.2.29], which were proved later by track monitoring [3.2.29, 3.2.30]. The success in squat investigation confirms that the transient FE model is a valid and promising tool in solving frictional rolling contact problems with the consideration of plasticity, complex contact geometry and dynamics.

It should be noted that the current FE model focuses on the rolling contact state with only the longitudinal friction force, which simulates the practice of traction or braking. In future work, efforts will also be placed on improving the FE model to take lateral creepage and spin into account. As suggested above, the small fluctuation of the contact solutions should be attributed to the high-frequency vibrations that are inherent to the rolling contact between bodies of continua. Thus, the occurrence of plastic deformation can reduce the level of the fluctuation. Studies are needed to further verify this. In addition, numerical errors of the FE method might also have played a role in the phenomenon, which should also be investigated in the future work.

Finally, the main characteristics\advantages of the 3D FE model, in terms of contact solution, are summarized as follows.

- The tangential solution is obtained together with the normal one.
- Arbitrary contact geometry can be treated in three dimensions because a fine mesh of solid elements can be applied. For instance, the rolling of a wheel over a 3D squat is treated [3.2.24].
- The transient rolling contact is solved in the time domain. Vibrations of the structure and continua are included in the solution together with the associated wave propagation due to the small time step required by the explicit integration scheme.

- Different friction forces transmitted in the contact patch can be simulated by specifying the corresponding traction or braking effort.
- Material non-linearity can be included.

3.2.6 Conclusions and remarks

A 3D transient FE approach is employed to solve the frictional rolling contact problem in elasto-plasticity. By applying it to contact between wheel and rail, the following conclusions can be drawn compared with the elastic solution:

1) The contact patch is increased in size, shifted forward in the rolling direction, and changed from an ellipse to an egg in shape.
2) In the rear part of the contact patch, the pressure is reduced along the lateral direction in the middle but increased at both sides. In the front part of the contact patch, in contrast, it is increased in the whole range.
3) The maximum pressure is reduced in magnitude and is moved forward in its longitudinal position.
4) The magnitude of the surface shear stress varies following the trend of pressure in the slip area, whereas it decreases along the longitudinal axis with a lower descending rate in the adhesion area.
5) The micro-slip distribution is also changed, which is attributed to plastic flow of material, as well as modifications of the contact patch and of the slip area.
6) The normal problem is non-linearly related to the tangential problem under plastic condition.
7) With the increase of traction coefficient, higher V-M stress is reached in the contact patch mainly due to the material hardening. At the same time, larger residual stress is also left behind.

The presented FE modeling can be applied to general rolling contact of Coulomb friction with the consideration of plasticity, complex contact geometry and dynamics. Employing computers of higher performance, cyclic plasticity can also be studied by simulating cyclic loading, in which influence of pre-existing residual stress can be considered. Furthermore, detailed frictional work and plastic deformation can be obtained, which makes it possible to investigate the relationships between material properties, plastic deformation, frictional work, wear and crack initiation. Such relationships may help in better understanding the occurrence of corrugation, head checks, and squats, and may further be used for design of fatigue resistant materials and profiles of wheel and rail.

Further, the transient wheel–rail rolling contact on a curved track can be simulated with the FE modeling by further taking into account lateral creepage and spin.

Reference

[3.2.1] H. Poritsky, N.Y. Schenectady, Stress and deflections of cylindrical bodies in contact with application to contact of gears and of locomotive wheels, Journal of Applied Mechanics 17 (1950), 191–201.

[3.2.2] J.J. Kalker, wheel–rail rolling contact theory, Wear 144 (1991), 243–261.

[3.2.3] R. Lewis, R.S. Dwyer-Joyce, Wear at the wheel/rail interface when sanding is used to increase adhesion, J. Rail Rapid. Transit. 220 (2006), 29–41.

[3.2.4] H. Hertz, Über die berührung fester elastischer körper, J. reine und angewandte Mathematik 92 (1882), 156–171.

[3.2.5] F.W. Carter, On the action of a locomotive driving wheel, Proc. R. Soc. Lond. A 112 (1926), 151–157.

[3.2.6] P.J. Vermeulen, K.L. Johnson, Contact of non-spherical bodies transmitting tangential forces, J. Appl. Mech. 31 (1964), 338–340.

[3.2.7] Z. Li, J.J. Kalker, Simulation of severe wheel–rail wear, Proceedings of the International Conference on Computer Aided Design, Manufacture and Operation in The Railway and Other Advanced Mass Transit Systems, Lisbon, Portugal, September 2–4, 1998, 393–402.

[3.2.8] R. Dollevoet, Z. Li and Arias-Cuevas (2010b), A method for the prediction of head checking initiation location and orientation under operational loading conditions. Proc. IMechE, Part F: J. Rail and Rapid Transit, 224 (F5) (2010), 369–374.

[3.2.9] R. Dollevoet, Design of an anti head check profile based on stress relief, PhD thesis, Z. Li as co-promotor, 2010, TU Twente.

[3.2.10] W. Yan, F.D. Fischer, Applicability of the Hertz contact theory to rail-wheel contact problem, Archive of applied Mechanics 70 (2000), 255–268.

[3.2.11] T. Telliskivi, U. Olofsson, Contact mechanics analysis of measured wheel–rail profiles using the finite element method, Proc Instn Mech Engrs, Part F 215 (2001), 65–72.

[3.2.12] M. Pau, F. Aymerich, F. Ginesu, Distribution of contact pressure in wheel–rail contact area, Wear 253 (2002), 265–274.

[3.2.13] A. Sladkowski, M. Sitarz, Analysis of wheel–rail interaction using FE software, Wear 258 (2005), 1217–1223.

[3.2.14] M. Wiest, E. Kassa, W. Daves, J.C.O. Nielsen, H. Ossberger, Assessment of methods for calculating contact pressure in wheel–rail/switch contact, Wear 265 (2008), 1439–1445.

[3.2.15] J.T. Oden and T.L. Lin, On the general rolling contact problem for finite deformations of viscoelastic cylinder, Computer methods in applied mechanics and engineering, 57 (1986), 297–376.

[3.2.16] J.M. Bass, Three-dimensional finite deformation, rolling contact of a hyperelastic cylinder formulation of the problem and computational results, Computers & Structures, 26 (1987), 991–1004.

[3.2.17] L.O. Faria, J.T. Oden, B. Yavari, W.W. Tworzydlo, J.M. Bass and E.B. Becker, Tire modeling by finite elements, Tire science and technology, 20 (1) (1992), 33–56.

[3.2.18] Nackenhorst, On the finite element analysis of steady state rolling contact. In M.H. Aliabadi and C.A. Brebbia (editors), Contact Mechanics - Computational Techniques, Computational Mechanics Publication, Southampton, Boston, 1993, 53–60.

[3.2.19] M. Ziefle, U. Nackenhorst, Numerical techniques for rolling rubber wheels: treatment of inelastic material properties and frictional contact, Comput. Mech, 42 (2008), 337–356.

[3.2.20] G. Hu and P. Wriggers, On the adaptive finite element methods of steady-state rolling contact for hyperelasticity in finite deformations, Comput. Methods Appl. Mech. Engrg. 191 (2002) 1333 – 1348.

[3.2.21] S. Damme, U. Nackenhorst, A. Wetzel, B. Zastrau, On the numerical analysis of the wheel–rail system in rolling contact, in: K. Popp, W. Schielen (Eds.), System Dynamics and Long-term Behavior of Railway Vehicles, Track and Subgrade, Springer, Berlin, Heidelberg, 2003, 155–174.

[3.2.22] Peter Wriggers, Computational contact mechanics, 2nd edition, Springer, 2006.

[3.2.23] U. Nackenhorst, The ALE-formulation of bodies in rolling contact – theoretical foundations and finite element approach, Comput. Methods Appl. Mech. Engrg, 193 (2004), 4299–4322.

[3.2.24] Z. Li, X. Zhao, C. Esveld, R. Dollevoet, M. Molodova, An investigation into the causes of squats: correlation analysis and numerical modeling, Wear 265 (2008), 1349–1355.

[3.2.25] X. Zhao, Z. Li, The solution of frictional wheel–rail rolling contact with a 3D transient finite element model: validation and error analysis, Wear 271 (2011), 444–452.

[3.2.26] R. Courant, K. Friedrichs, H. Lewy, On the partial differential equations of Mathematical Physics, Math. Ann. 100 (1928), 32 – 74.

[3.2.27] D.J. Benson, J.O. Hallquist, A single surface contact algorithm for the post-buckling analysis of shell structures, Computer Methods in Applied Mechanics and Engineering 78 (1990), 141–163.

[3.2.28] O. Arias-Cuevas, Z. Li and R. Lewis, Investigating the lubricity and electrical insulation caused by sanding in dry wheel–rail contact, Tribology letter, 37 (2010), 623–635.

[3.2.29] Z. Li, X. Zhao, R. Dollevoet and J. Liu, The determination of a critical size for rail top surface defects to grow into squats, CM2009, Florence Italy September 15–18 2009, 379–388.

[3.2.30] Z. Li, R. Dollevoet, M. Molodova and X. Zhao, Squat growth–some observations and the validation of numerical predictions, Wear 271 (2011), 148–157.

Summary and recommendations

Compared to Hertz theory and Kalker's CONTACT, the 3D FE model is validated in elasticity and in statics for both the normal and tangential contact solutions. Considering that the FE model can take into account actual contact geometry, material non-linearity, and transient effects, the newly developed modeling approach provides a valid and promising tool to solve the problem of frictional rolling contact.

Elasto-plastic solutions of frictional rolling contact are also obtained by the validated FE model. Simulations of different traction efforts show that plastic flow and the resulted changes of contact geometry significantly modify the normal and tangential solutions. As a result, the normal problem, which is decoupled from the tangential problem in the elastic situation, becomes non-linearly related to the tangential problem when material behaves plastically.

4 Squat initiation, growth and wheel–track dynamic forces

In this chapter, the FE model of rolling contact presented in Chapter 3 is integrated into the vehicle–track interaction system. Rail surface defects are introduced to simulate the dynamic vehicle–track interaction at squats. Dynamic wheel–rail contact force is mainly studied. Taking into account field observations, a growth process of squats is postulated in Section 4.1 based on dynamic contact forces obtained from the FE model. A correlation analysis on squats in the Netherlands is also presented. Section 4.2 presents validations of the findings in Section 4.1. Afterwards, the dynamic contact force is further analyzed in the frequency domain in Section 4.3. Two vibration components are found to be related to squats. Moreover, dynamic forces at the discrete supports of rail are also examined for the case of squats in Section 4.3.

| Chapter 1: Introduction to squats and their consequences. | Chapter 2: Literature review and outline of this dissertation. |

| Chapter 3: Solution of frictional rolling contact. |

| **Chapter 4: Squat initiation, growth and wheel–track dynamic** |

| Chapter 5: A minimum critical size for squats to initiate. |

| Chapter 6: Conclusions and recommendations. |
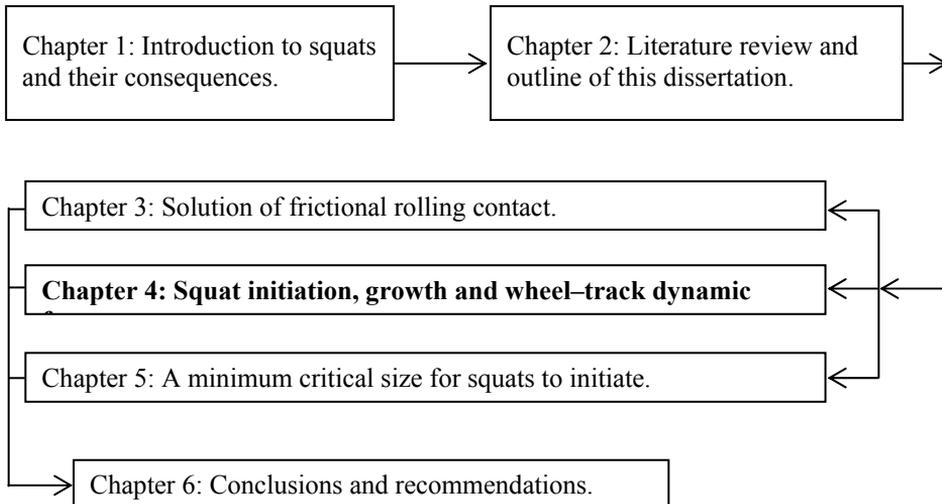
4.1 An investigation into the causes of squats— Correlation analysis and numerical modeling[3]

ABSTRACT

Nowadays squats are a severe rolling contact problem on some railways. The root causes of it are not yet clear. From railway operation and maintenance points of view, root causes have to be known so that effective maintenance policies and plans can be made based on measurements and prediction. In the present work first a correlation analysis is performed based on measured data and field observations, and then a relation between squat occurrence and some parameters of the vehicle–track system is identified. Subsequently a transient finite element model of vehicle–track interaction is built up to analyze the contact forces and the resulting stresses and strains. The numerical results show good correspondence with observed phenomena at squats, and suggest certain relation between squat development and some eigen frequencies of the coupled vehicle–track system. A squat growth process is accordingly postulated.

---

[3] This section has been published as: Z. Li, X. Zhao, C. Esveld, R. Dollevoet and M. Molodova, *An investigation into the causes of Squats: Correlation Analysis and Numerical Modeling*, Wear 265 (2008), 1349–1355.

4.1.1 Introduction

Squat is a type of rolling contact fatigue (RCF) which was hardly known in Europe until about 30 years ago [4.1.1, 4.1.2]. Now they have become one of the major RCF damages on some railways [4.1.3]. A fundamental characteristic of squats is local plastic deformation of the rail top surface. Excessive dynamic wheel–rail contact force is apparently the direct cause; changes in the traffic–track interactive systems in the past decades should lie behind its occurrence. These changes include higher axle load and traffic density, different maintenance policies and deterioration in track quality, new wheel and rail materials, stiffer concrete sleepers, etc. Squats can be roughly classified into three categories—light, moderate and severe [4.1.3], or classes A, B and C, correspondingly. For classes A and B squats grinding of the rail top may be an effective corrective cure; with severe squats, replacement of rail is often inevitable. Squats are therefore not only a threat to operation safety, but also very costly for maintenance.

So far research on squats has mainly been on the metallurgical aspect and on crack growth. Clayton et al. [4.1.1] reviewed a metallurgical research program at British Rail on surface-initiated rail problems; squats were found a problem of great concern on some routes in the UK. Longitudinal/vertical sectioning rail specimens showed surface-initiated cracks which can branch downwards, and when reaching a critical size brittle fracture can result in a broken rail. It was found that fatigue life is reduced as contact pressure is increased, and creepage has played an important role. A factor of 10 differences has been observed in the rolling contact fatigue resistance of two tested materials.

In Ref. [4.1.4] Bold et al. studied the growth rate of squat-type shallow-angled surface-initiated rolling contact fatigue cracks. At the same time Smallwood et al. [4.1.5] developed optimization techniques for minimizing contact stresses to delay the initiation of RCF cracks which are a precursor to squat, the cracks being found on the running surface of high rails of curved tracks.

In 1990s the European Rail Research Institute (ERRI) carried out a RCF research program [4.1.2], in which Bogdański et al. [4.1.6] modeled a squat as a plane oblique semi-elliptical crack. The state of stress in the vicinity of the crack front is determined, and the values and ranges of the stress intensity factors at the crack front are calculated. In Ref. [4.1.7], by combining crack front loading histories with mixed-mode fatigue crack growth rate data Bogdański and Brown further analyzed the growth of squat-type cracks.

Dang Van and Maitournam [4.1.8] presented for the case of squats calculations of stresses and strains in the rail subjected to repeated moving contacts. Stationary methods were employed. Busquet et al. [4.1.9] computed plastic flows in the near-surface layer as a function of traction coefficient; surface contact load distribution was based on the solution of Kalker.

For RCF problems prevention is a safer and more economic solution. This is particularly true for squats in view of their very detrimental effects on the track. For such purpose their root causes have to be known, their development process needs to be clarified and the related detection methods should be developed. A RCF research project

has therefore been launched by the Dutch Railways seeking the root causes of squats and quantitative relations between these causes and squat initiation.

The present work is concerned with the squats which are found in the running band in the middle of the rail top. This kind of squats grows from small rail top local geometrical irregularities, with or without cracks in their early stage. The mechanism of their development is not yet clear.

4.1.2   Correlation analyses

4.1.2.1   Squats in the Netherlands

Squats in the Netherlands occur mainly on tangent tracks and shallow curves, and on switch and crossing works, in a seemingly random way. Class A squats (Figure 4.1.1(a)) are usually found at rail surface irregularities such as welds, indentations, wheel burns, short wave corrugation, etc. In more advanced stages (classes B and C, see Figure 4.1.1(b) and (c)) they are often seen with widening and sinking of the wheel–rail contact zone, which causes large impact forces between wheel and rail, resulting in noise, vibration and rapid deterioration of the track. Many of them have accompanying cracks, which, if developing downward, may cause catastrophic rail fracture. Owing to the large forces associated with the geometry deviation, classes B and C squats grow continuously and in an accelerated way. That means shakedown limit of rail material is exceeded and ratcheting (incremental plastic flow) of the rail take place.

Among the large number of class A squats only a small part grow into classes B and subsequently C [4.1.3]. It is important to mention that though indentations are an important source of class A squats, most indentations do not grow into squats. An example is at level crossings where the most indentations are found: only few squats classes B and C are found there. Squats have also been occasionally observed at locations of cross-sectional change of rails, such as at switch and crossings, and at insulated joints. All these observations indicate that in the vehicle–track interaction system there are other influential factors which, together with surface irregularities, determine the initiation and growth of squats.

Since 2002, squats have been recorded with their severities, locations and photos of the rail tops in the Netherlands in a data base called IRISsys which also contains track geometry irregularity data obtained by a measuring train running twice a year over the entire Dutch network.

4.1.2.2   Correlation analysis

To isolate the root causes of squats, a correlation analysis [4.1.10] was performed based on field inspections and on examining 122 km of rails with IRISsys by checking the rail top conditions and track geometry data around squats.
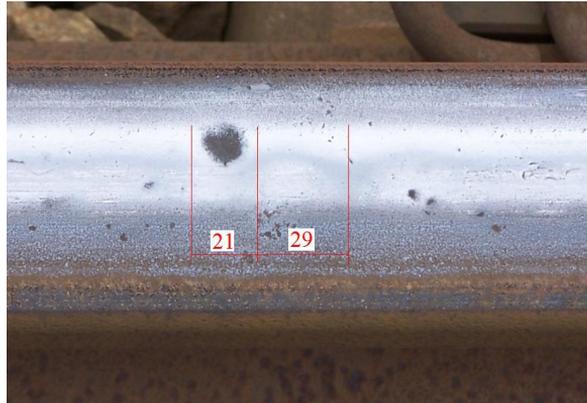
(a)
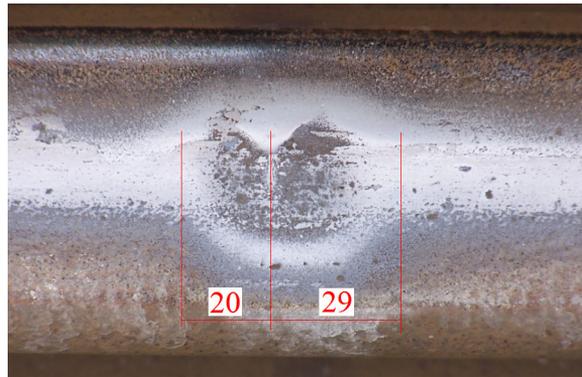
(b)
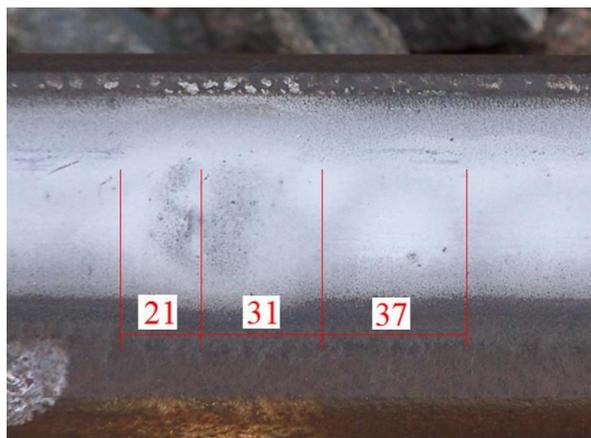
(c)

Figure 4.1.1. Squats in different severity phases: (a) Class A, (b) Class B and (c) Class C. Traffic from left to right. The numbers are estimated distance in mm, see discussion in section 4.1.3. Note the photos are taken from 3 different squats.

The relationship of squat occurrence with track irregularities such as alignment, cross-level, gauge and vertical profile was analyzed; effects of rail vertical wear and rail foot inclination were examined. There has been no clear correlation found in this regard, as is expected, because these irregularities are of the long wave type.

However correlation has been identified with short wave irregularities. If divide the analyzed rail into two parts, as shown in Figure 4.1.2, it is found that about 74% of the squats occur on part I, and the rest are on part II. In other words 74% of squats are found on the half rail above the sleepers. This may suggests that the stiffness and damping characteristics of the track, particularly those of the rail and the rail pad, may have played a role.
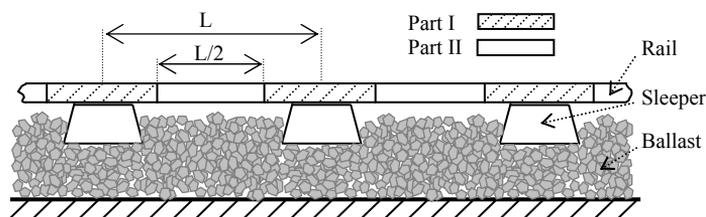
Figure 4.1.2. The definition of parts I and II of rail for the correlation analysis. L is a sleeper span.

It is also found that corrugation of various severities can be seen in the neighborhood of about 72% of the squats. This may suggest certain relationship between squats and corrugation: corrugation and squats may have occurred due to the same stiffness and damping properties of the track, though they may have developed independently; or the rail surface irregularities at some corrugation have caused large wheel–rail forces, therefore caused the squats around the corrugation; or it is a consequence of the two mechanisms combined: track local eigen characteristics cause both corrugation and squats to develop, and rail surface irregularities at corrugation speed up the development of squats.

It is noticed that the majority of classes B and C squats bear a typical shape shown in Figure 4.1.1(b) and (c), i.e. the dark widened and sunken part of the rail surface is divided into two by a piece of shining out-jutting material in the middle. At the most advanced stage, this shining part may disappear. According to Clayton et al. [4.1.1], a network of cracks may be found beneath the dark surface. Another outstanding feature shown in Figure 4.1.1 is the wave pattern immediately following the squats. This phenomenon can be found accompanying many class C squats, to a lesser extent with class A, while not yet seen by the author with class B.

Welds, both thermite and flash butt, are vulnerable to squats. Out of a field survey [4.1.10] of 65 squats listed in Table 4.1.1, 17% are found at welds. Among the 16 squats recorded in the section between Barneveld Noord and Apeldoorn where many more squats are found, 6 are at welds. This is perhaps an extreme case. If it is removed from the statistics, there will still be 5 at welds out of the 49, which is 10% of the total. Figure 4.1.3 shows two squats at thermite welds. The high percentage of squats at welds may be explained by two main factors: material strength/hardness and longitudinal rail top profile. Figure 4.1.4 illustrates the hardness distributions of thermite and flash butt welds by a

welding company. Part, if not all, of the weld and the heat affected zone have lower hardness than the parent material. Figure 4.1.5 is the rail longitudinal profile measured at a weld.

Table 4.1.1 Statistics of squats at welds

Track locations	Number of squats	Squats at welds
Amsterdam CS -Sloterdijk	3	0
Den Haag Laan van NOI	5	0
Den Haag Mariahoeve	15	1
Barneveld Noord - Apeldoorn	16	6
Ede-Wageningen	14	1
Rotterdam CS - Schiedam	5	0
Zoetermeer	7	3
Sum	65	11

(a)

(b)

Figure 4.1.3. Squats at thermite welds, traffic from lower to upper side. Spalling already occurs due to large plastic deformation.

Comparing Figures 4.1.3–4.1.5, resemblance can be seen between the hardness variation and the rail top longitudinal profiles. When a new weld is made with large geometrical deviation, high contact forces occur. A new welding geometry assessment norm has been developed in the Netherlands [4.1.11]. It should be noted that the rail longitudinal profile at a weld shown in Figure 4.1.5 is just a typical one. Weld geometry differs from one to another. This one is chosen because the 80 mm or so profile at the center is typical of the rail top geometry at classes B and C squats (see Figures 4.1.1 and 4.1.3). Usually the vertical profile deviation at classes B and C squats is larger than shown in Figure 4.1.5.

It is observed that more squats are found near stations and signals where traction and braking efforts are high. This has not yet been statistically verified.
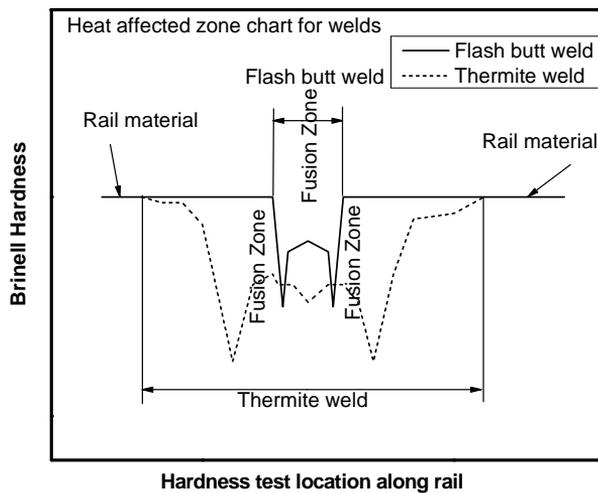
Figure 4.1.4. A Schematic hardness distribution of welds by a welding company
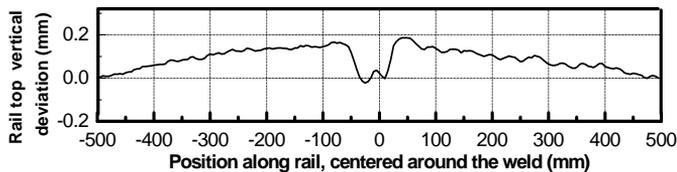
Figure 4.1.5. Vertical rail top profile measured at a weld

4.1.3   Numerical analysis

4.1.3.1   The finite element model

To isolate the root causes in the vehicle–track system by parameter variation, a transient finite element model is built up (Figure 4.1.6). Due to the short wave characteristic, only one wheel on one rail is considered. The length of the model covers five sleeper spans.

In Table 4.1.2 the model data are listed. Sprung mass is lumped into Mc with first suspension stiffness Kc and damping Cc adopted from parameters of the ICM-type Dutch intercity passenger train. Track parameters are taken from [4.1.12] for the Dutch railway. A passenger car wheel is modeled with a nominal radius of 0.46 m, tread conicity 1/20 and mass 431 kg. The rail is UIC60, with 1/20 inclination. Since squats are found in the middle of rail top, the vehicle–track system is modeled only in the vertical direction. The wheel and rail FE model is three-dimensional, so that the transient stress-strain state in the contact can be fully taken into consideration. The wheel flange shape is ignored because it is irrelevant.

(a)

(b)
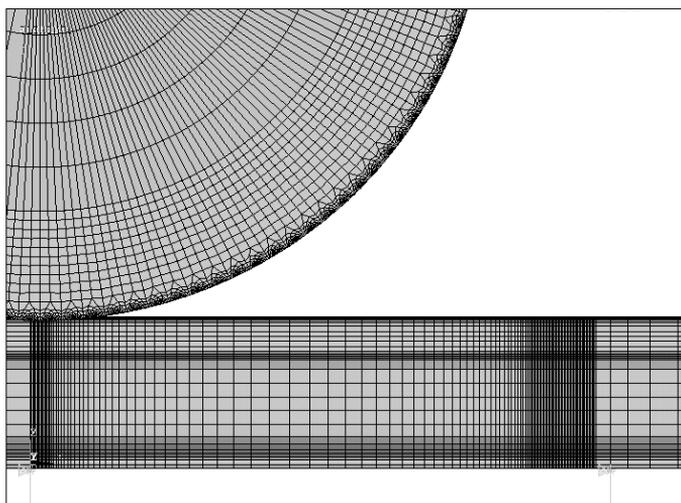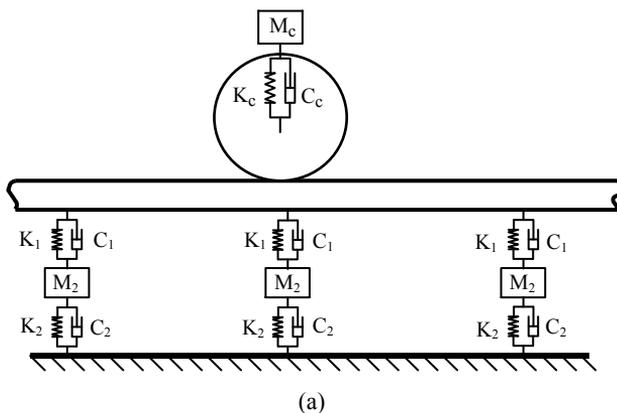
Figure 4.1.6. The finite element model [*]. (a) Global overview and (b) the mesh.

---

[*] During developing the model, influence of the length of the model was investigated and the length employed in this dissertation was found to be sufficient for the investigated topics. Symmetry boundary conditions are applied to the rail ends.

Table 4.1.2 Vehicle, track and material data used

Components	Parameters	Values
Sprung mass	Mass	10,000 kg
1st suspension	Stiffness	1.15 MN/m
	Damping	2500 Ns/m
Rail pad	Stiffness	1300 MN/m
	Damping	45 KNs/m
Sleeper	Mass	200 Kg
	Spacing	0.6 m
Ballast	Stiffness	45 MN/m
	Damping	32 KNs/m
Wheel and rail material	Young's modulus	210 GPa
	Poisson's ratio	0.28
	Mass density	7,800 kg/m^3
	Yield stress [1]	1.12 GPa
	Yield stress [2]	0.8 Gpa
	Tangent modulus	21 GPa

[1] work hardened rail; [2] work hardened welds

4.1.3.2 The contact forces

Two distinctive cases of rail surface irregularities are modeled (see Figure 4.1.7). For case 1, the irregularity is between A_{1G} and B_{1G}. It is artificial, corresponding to the typical dimension of a class A squat. For case 2 the irregularity is between A_{2G} and C_{2G}. It is taken from the measured profile of Figure 4.1.5. For the latter case the geometry of the squat is asymmetry. Its traffic direction is not known, and is tentatively selected as shown in Figure 4.1.7(b). It is shown below that this tentative selection does not affect the analysis and conclusions of the present work.

The wheel is initially at the position shown in Figure 4.1.6(b) which is 0.48 m away from the defect. This distance is chosen to minimize the influences of initial conditions and the computation time. The smallest element size for both wheel and rail is 1.3 mm, which should produce a reasonably accurate stress and strain state in the area in question. After finding its equilibrium position, the wheel moves with a typical Dutch line speed of 140 km/h from time t=0 toward the irregularities, which are located in the densely meshed area on the right side of Figure 4.1.6(b). At about 14 ms and 12.75 ms, the wheel arrives at somewhere before the beginning points A_{1G} and A_{2G} of cases 1 and 2, respectively. Due to the lower contact stiffness at the sharp points A_{1G} and A_{2G}, which in reality is smoothed due to deformation, the contact force begins to fall before these two points.
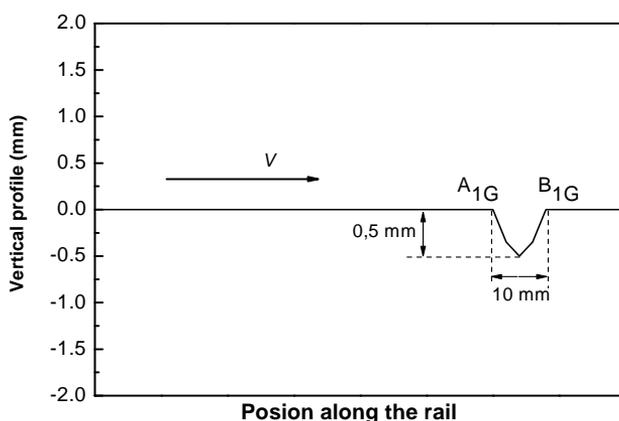
(a)
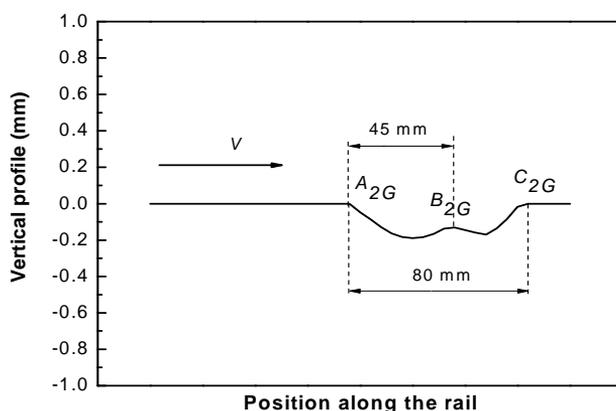
(b)

Figure 4.1.7. The two rail surface irregularity cases simulated. (a) Case 1, an artificial circular irregularity simulating class A squats, without the wave pattern that often follows a squat. (b) Case 2, measured geometry taken from Figure 4.1.5.

Figure 4.1.8 shows the normal contact forces. The maximum force is about 1.8 times of the static load. Note that the unsprung mass of a half wheelset is usually higher than the value used here when with brake discs and driving system directly mounted on it, and especially for locomotive wheelsets. Therefore for such wheelsets the dynamic force should also be correspondingly higher. Sprung mass does not show significant influences on the reaction of squats as is expected due to its low frequency characteristic. The small variation of the force before the A_1 and A_2 points is attributed to the moving wheel on the elastically supported rail.

It is interesting to note that for case 1, the largest force is not B_1 peak, which corresponding to B_{1G}, but thereafter. As shown in Figure 4.1.8(a), the distance between contact force peaks A_1 and B_1 is about 20 mm, and between B_1 and C_1 is about 30 mm. These distances coincide very well with the longitudinal lengths of the squat (21 mm) and the first wavelength after it (29 mm) shown in Figure 4.1.1(a) (indicated by the red lines

and the numbers). In addition, they agree also very well with the distances in Figure 4.1.1(b) between the red lines, namely 20 mm and 29 mm. In estimating the distances, the width of the contact patch and the effect of the relatively sharp edge at A_{1G} and B_{1G} have been taken into account.

The same is true for case 2. Looking at Figure 4.1.8(b), the distances between the contact force peaks, which are A_2–B_2 (25 mm), B_2–C_2 (35 mm) and C_2–D_2 (37 mm), correspond again very well to the wavelength of the squat shown in Figure 4.1.1(c) (21 mm, 31 mm and 37 mm, respectively, indicated by the red lines).
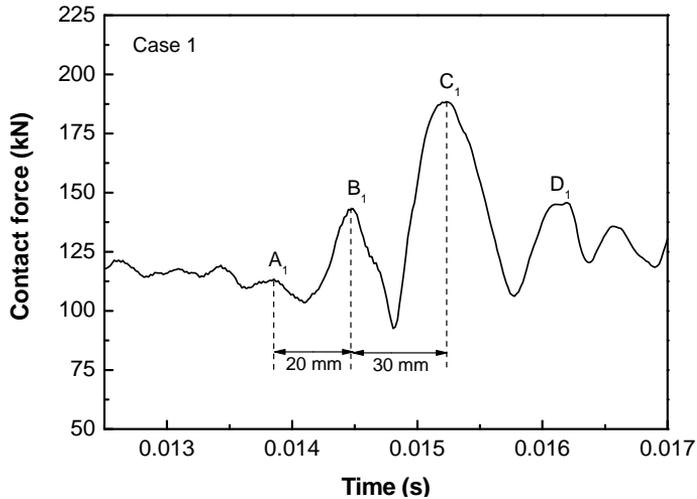
(a)
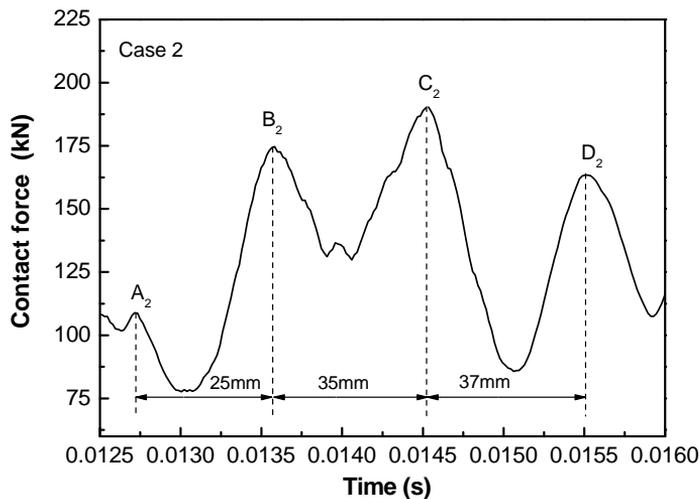
(b)

Figure 4.1.8. Contact forces for (a) case 1 and (b) case 2 respectively. They modify the contact geometry.

Comparing the wavelengths in Figure 4.1.8(b) with those in Figure 4.1.7(b), it seemed that it would, for the same force wave of Figure 4.1.8(b), be more reasonable if the traffic was from the right to the left. To clarify this, another calculation was performed by reversing the rolling direction of case 2, i.e. the wheel rolls over the squat from C_{2G} to A_{2G}. The results show that the dynamic force wave in this case is almost the same as that of the wheel rolling from A_{2G} to C_{2G}, with only a small difference in force amplitude. This means that in this case the force wave, especially its wavelength, is mainly determined by the local eigen system, irrespective of traffic direction, therefore irrespective of geometry. In other words, it would mean that squats started from different initial rail top irregularities will grow in a way so that their appearance at the same later stage of their development will be similar. This reasoning is indeed supported by the following two facts observed by the authors: (1) initiating squat geometry varies largely, and (2) almost all squats of classes B and C bear close similarity to those shown in Figure 4.1.1(b) and (c). This shows that the tentative selection of traffic direction for case 2 does not affect validity of the analysis and conclusions.

4.1.3.3   A postulated squats growth process [*]

In view of the observations and discussion above, especially the comparison between the distances indicated in Figure 4.1.1 and the wave lengths of the contact forces out of the simulation, it may now be postulated that during the development of a squat, A_{1G} remains at the leading edge of the squat, B_{1G} becomes gradually the middle out-jutting V shape shining part corresponding to the position of the middle red line in Figure 4.1.1(b). B_{1G} deforms then further into the central shining strip of the squat at the location of the second vertical red line of Figure 4.1.1(c) (counted from the left), which corresponds to B_{2G}. The C_1 peak contact force, corresponding to the wave pattern at the right-most red line in Figure 4.1.1(a), causes the rail at this place to deform gradually into part of the squat, corresponding to the location indicated by the right-most red line in Figure 4.1.1(b). By this process, a class A squat of case 1 geometry (Figure 4.1.7(a)) has developed into a class B squat, the geometry of which is already similar to that of case 2 (Figure 4.1.7(b)). A class B squat continues to grow into class C. Its V shape shining part shrinks into the central shining strip, which at the most advanced stage of a squat may be wiped out by the large contact force (Figure 4.1.9).

According to this process, the wave pattern following a class A squat may disappear at certain stage of its development when the first wave (indicated by the right-most red line in Figure 4.1.1(a)) has developed into part of the squat (the part between the middle and right red lines of Figure 4.1.1(b)), while the next contact force peak D_1 is not yet large enough to cause sufficient plastic deformation to form another visible wave to the right side of the right-most red lines of Figure 4.1.1(a) and (b). Compare the peaks D_1 and D_2 in Figure 4.1.8, and the respective corresponding positions in Figure 4.1.1 for the absence and presence of wave pattern. This seems indeed to be the case: the authors have so far not been able to find wave pattern which follows a class B squat like the one in Figure 4.1.1(b).

---

[*] The damage mechanism behind the postulated growth process may be differential wear and/or plastic deformation, which is explained in detail in: Z. Li, X. Zhao, R. Dollevoet, M. Molodova, Differential wear and plastic deformation as causes of squat at track local stiffness change combined with other track short defects, Veh. Syst. Dyn. 46 (Suppl) (2008) 237–246.

Figure 4.1.9. A severe squat. The rail has to be replaced very soon.

It is noticed in Figure 4.1.8 that peaks B_1, B_2 and C_2 are due to forced vibration by the contact geometry. On the other hand the peaks C_1, D_1 and D_2 are due to free vibration of the coupled vehicle–track system. In view of the process postulated above this would suggest that the development of squats is influenced by certain eigen frequencies of the system. The wavelength is between 20 mm and 40 mm, bearing some resemblance with short pitch corrugation. The geometrical wave is smeared when the squats are developed into a advanced stage and if the wavelength is shorter than 30 mm or so, i.e. the wave pattern of Figure 4.1.1(a) is smeared into being part of the squat, and the V shape shining part of Figure 4.1.1(b) is smeared into the middle strip in Figure 4.1.1(c), which may finally be wiped out at more advanced stages, see, e.g. Figure 4.1.9. Further investigation is needed into this phenomenon.

4.1.3.4   Parameter variation studies

To examine the stress and strain states a bi-linear elastic-plastic material is chosen for the rail. Under repeated wheel passages the rail surface is work hardened. With a three-dimensional finite element model, it is too time consuming to simulate cyclic plasticity of wheel–rail contact. According to Ref. [4.1.12], the highest rail surface hardness measured after many wheel passages is 336 HV, which is equivalent to a yield stress of 1.12 GPa. This value is used in the present work to take the cyclic hardening process into account. A lower yield stress of 0.8 GPa is also tentatively chosen to represent weld for comparison.

Figure 4.1.10 shows the peak von-Mises effective stresses in the strip of surface elements along the travelling direction where the highest stress value is found in the entire model. The peak stress of an element is the highest stress the element experiences in the time history (see Figure 4.1.11(a) for an example). It can be seen from the smoothed curve that the peak stress variation follows more or less the shape of the rail top profile. As can be expected the stress state of the two cases are the same when it is below the yield stress of a weld, and the maximum stress of the weld is much lower than that of the normal rail due to the larger plastic flow. The location of the highest stress in Figure 4.1.10(b) corresponds well to the spalling spot of Figure 4.1.3(a). In Figure

4.1.10(b) the stress corresponding to B_2 point is already close to the yield stress. If the yield stress is lower, or the profile deviation is larger, or the axle load, esp. the unsprung mass is higher, plastic flow may take place there, the situation of Figure 4.1.3(b) occurs. The computed stresses in Figure 4.1.10 fluctuate significantly; this may happen in reality, see the alternating shining and dark spots in Figures 4.1.1 and 4.1.3. Sometimes multiple ripples are seen in a squat.
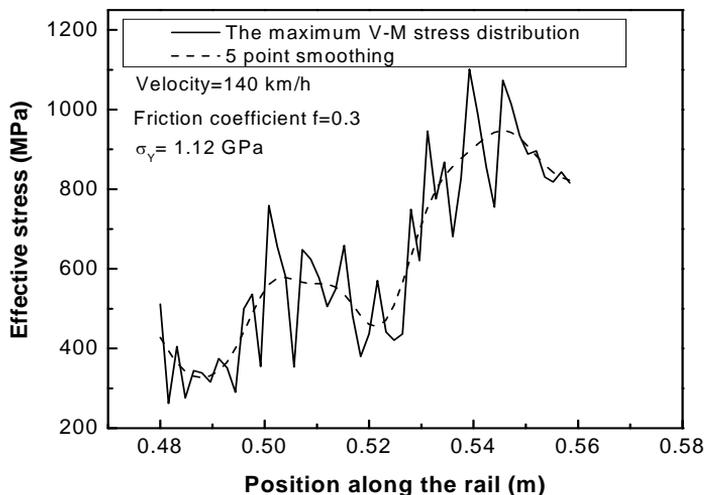
(a)
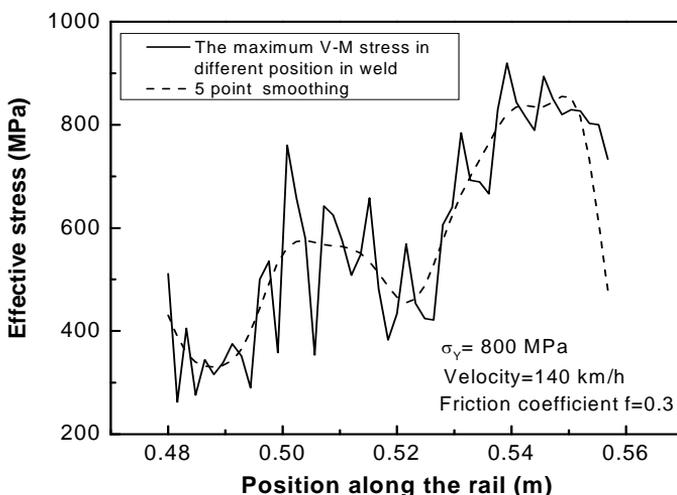
(b)

Figure 4.1.10. The peak Von-Mises effective stress in the strip of surface elements, among which the highest effective stress is found. (a) With yield stress 1.12GPa and (b) with yield stress 0.8GPa. The sharp fall of the smoothed results is due to smoothing error. Origin of abscissa is at starting point of wheel.

In Figure 4.1.11 the histories of von-Mises effective stress and the effective plastic stain in element 82911 with different yield stresses, which is in the area with the highest stress and strain, are compared. The weld, with a lower yield strain, experiences much larger plastic flow under the same loading condition.

In the above calculations, the wheel has been set to sliding with a coefficient of friction equal to 0.3. In Figure 4.1.12, a comparison is made with friction coefficient equal to 0. With the same loading condition, the same element undergoes no plastic deformation any more. This indicates the great influence of traction and braking on squats development. As is expected, the different coefficient of frictions has negligible effect on the normal contact force.
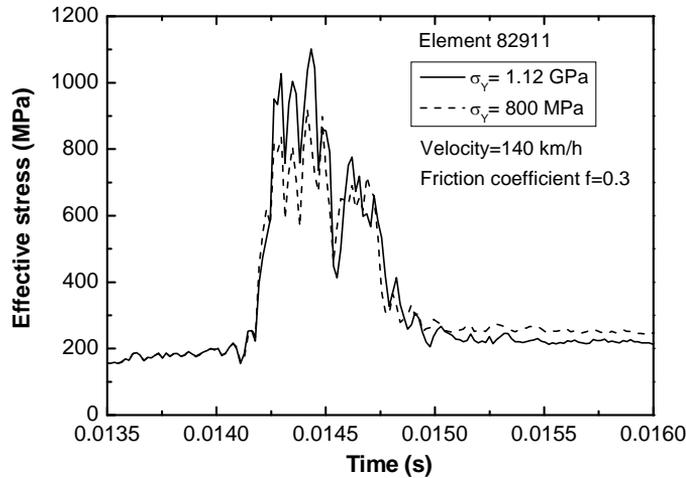
(a)
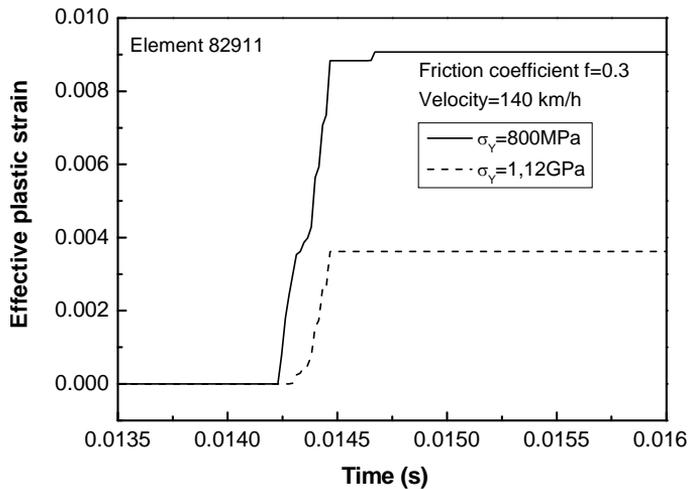
(b)

Figure 4.1.11. Comparison of effective stress and plastic strain histories. (a) Histories of effective stress and (b) histories of plastic strain.
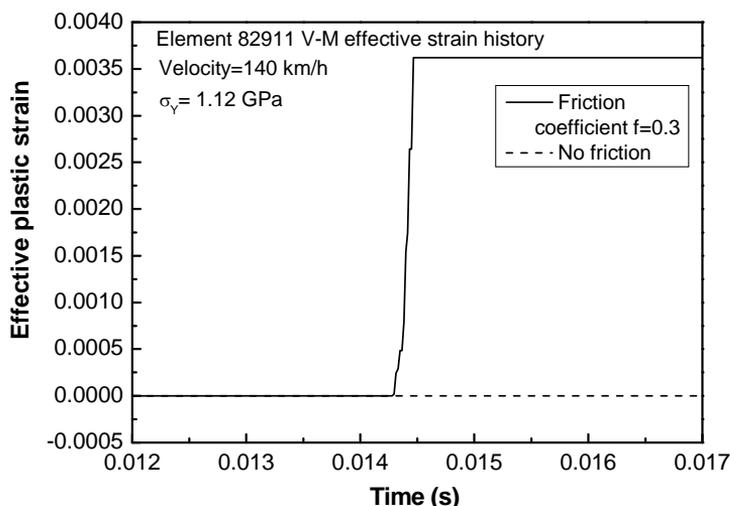
Figure 4.1.12. Comparison of plastic strain under different friction conditions

4.1.4   A general discussion

In the numerical analysis the additional components which are directly mounted on wheel such as brake discs and driving devices have not been included. And locomotive wheels have larger radius. These mean higher unsprung mass, and may increase the dynamic contact forces significantly. Also in reality the rail top profile deviation may be much larger than modeled in the present work. The effects of all these, together with other factors in the vehicle–track system, such as properties of the fastening system and ballast, sleeper mass and stiffness, should be analyzed in further work.

The underlying cracks accompanying many squats, indicated by the dark marks, and the residual stresses are not taken into account; they have certainly influence on the contact forces and the stress and strain states in the rail. Further, the initiation and propagation of the cracks should be analyzed under three-dimensional transient/impact conditions.

The numerical modeling has produced normal contact force magnitude and wavelength which correspond well with squat dimension and the wave pattern following them. Based on this observation, a squat growth process has been postulated. This has yet to be verified. A field squat monitoring is been planned. Crack depth, rail head transverse and longitudinal profiles, axle box acceleration and track local dynamic properties will be measured at a number of squats. It is hoped that this will deliver sufficient data for validation and for future analyses.

The present work does not explain how class A squats like Figure 4.1.1(a) come into exist. In the FE model, the geometry of Figure 4.1.7 is applied across the transverse profile of the rail; this would have led to some error in the results. Further refinement is needed.

4.1.5   Conclusions

The correlation analyses have been performed based on data collected from track inspections and from IRISsys. It is concluded that squat occurrence can be related to track short wave irregularities, especially those on the rail top such as indentations, weld and corrugations. Material strength, unsprung mass, traction and braking, sleeper spacing and fastening system properties also play important roles.

A transient finite element model has been set-up to analyze these factors. Results in contact force, stress and strain levels, and their positions show good correspondence with field observations. A squat development process has been postulated based on the correspondence between on the one side the contact force magnitude and its wavelength and on the other side the squat dimension and the wave pattern following them. The wavelength is between 20 mm and 40 mm, which corresponds to 950–1900Hz for a rolling speed of 140 km/h, and bears some resemblance to short wave corrugation. This suggests that certain eigen frequency of the coupled vehicle–track system should have influenced the development of squats.

Further numerical analyses for parameter variation are needed to take more factors into account and to reveal more details. Monitoring and measuring are being planned to collect data for validation and further analyses.

References

[4.1.1]   P. Clayton, M.B.P. Allery, P.J. Bolton, Surface damage phenomena in rails, in J. Kalousek, R.V. Dukkipati, G.M.L. Gladwell (Ed), Proceedings of the conference on Contact Mechanics and Wear of Rail/wheel Systems, Vancouver, British Columbia, July 6-9, 1982, 419-443, University of Waterloo Press, 1982

[4.1.2]   D.F. Cannon, H.Pradier, Rail rolling contact fatigue – Research by the European Rail Research Institute, Wear 191 (1996) 1-13.

[4.1.3]   J. Smulders, Management and research tackle rolling contact fatigue, Railway Gazette International, June 2003, 439 – 442.

[4.1.4]   P. E. Bold, M. W. Brown, R. J. Allen, Shear mode crack growth and rolling contact fatigue, Wear 144 (1991) 307-317.

[4.1.5]   R. Smallwood, J. C. Sinclair, K. J. Sawley, An optimization technique to minimize rail contact stresses, Wear 144 (1991) 373-384.

[4.1.6]   S. Bogdanski, M. Olzak, J. Stupnicki, Numerical modelling of 3D rail RCF 'squat'-type crack under operating load, Fatigue Fract. Eng. Mater. Struct. 21 (1998) 923–935.

[4.1.7]   S. Bogdanski, M. W. Brown, Modelling the three-dimensional behavior of shallow rolling contact fatigue cracks in rails, Wear 253 (2002) 17-25.

[4.1.8]   K. Dang Van, M.H. Maitournam, On some recent trends in modelling of contact fatigue and wear in rail, Wear 253 (2002) 219–227.

[4.1.9]   M. Busquet, L. Baillet, C. Bordreuil, Y. Berthier, 3D finite element investigation on the plastic flows of rolling contacts—correlation with railhead microstructure observations, Wear 258 (2005) 1071–1080.

[4.1.10] Z. Li, Correlation Analysis of Squats, Internal project report, Railway Engineering group, TU Delft, 5 March, 2006.

[4.1.11] M. Steenbergen, C. Esveld, R.Dollevoet, New Dutch assessment of rail welding geometry, European Railway Review 11 (2005).

[4.1.12] M. Hiensch, J. C.O. Nielsen, E. Verheijen, Rail corrugation in The Netherlands—measurements and simulations, Wear 253 (2002) 140–149.

4.2 Validation of numerical predictions about squat growth[4]

ABSTRACT

This section presents several observations that validate the numerical predictions about squats shown in Section 4.1. It includes the postulated growth process of squats, the relationship between squat growth and friction, the relationship between dynamic contact force and wave pattern, and the high frequency interaction at squats.

4.2.1 Validation of the postulation on squats growth process

4.2.1.1 Example 1

In Figure 4.2.1 one can see how a lung-like squat was developing from a small defect, in the way postulated in Section 4.1. It is not clear what kind of defect precisely it was in the photo of March 2006. It could be an indentation, or it could be remains of a (network of) cracks which were not completely ground away. The rail was ground not long before March 2006; grinding marks are clearly visible in the upper photo of Figure 4.2.1. But even if it was remains of cracks, the cracks should be small and shallow because otherwise black spot caused by the subsurface cracks should be visible in the photos. Further the small V-shaped crack of February 2008 seems not related to any of the crack-like features in the photos of March 2006 and June 2007. It may therefore be concluded that the nature of the defect in March 2006 did not have influence on its growth into the typical squat of February 2008. Further monitoring of this squat was impossible because the rail was removed from the track due to other severe squats nearby.
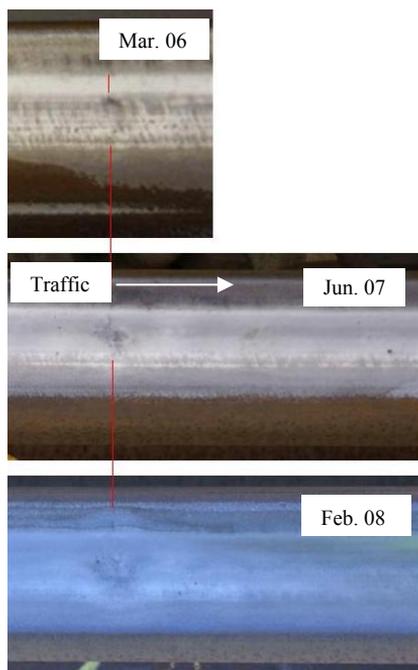
Figure 4.2.1. A growing squat with the typical two-lung shape being formed. A small crack is visible in the middle of the squat in February, 2008. It is not clear if there was already a crack in March 2006.

4.2.1.2   Example 2

Figure 4.2.2 shows the growth of one more squat. According to the photo of June 2007, there should be no surface crack at that time. In October 2008, the second lung of the squat seemed being formed to the right side of the first lung, following the postulated growth process. It is not clear if the crack-like marks in October 2008 were real initiating cracks or they were just some scratches caused by something entrapped there between the wheels and the rail. The rail of this squat was also removed from the track due to severe squats in the neighborhood.
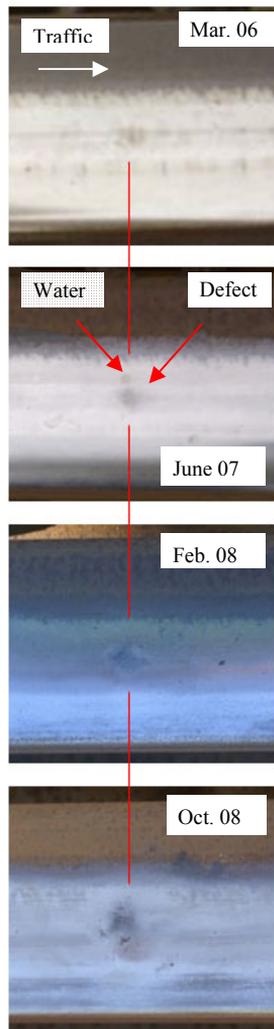
Figure 4.2.2. A squat growing from a small defect. The two-lung shape started to form in Oct. 2008. No visible crack in June 2007. There was a water mark in the photo of June 2007, which should not be taken as part of the defect, as indicated in the photo.

4.2.1.3 Example 3

Figure 4.2.3 shows how a lung-like squat which grew from small to large. It is noted that the squat in the upper photo of Figure 4.2.3 bear good resemblance with those in the lowest photos of Figures 4.2.1 and 4.2.2 in the senses of their size and the embryonic two-lung shape.
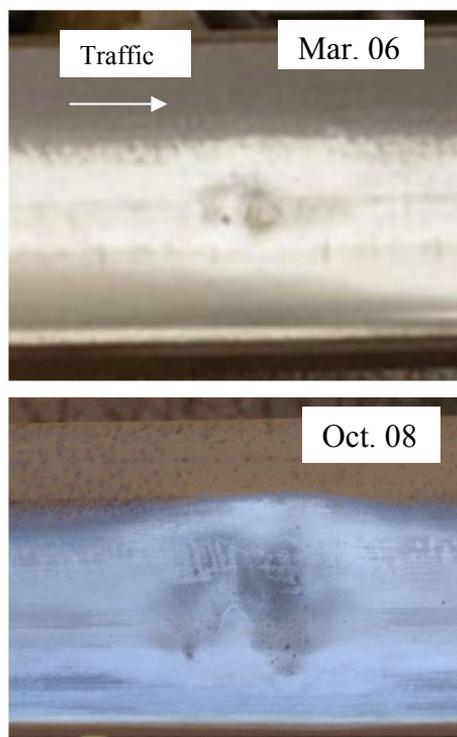
Figure 4.2.3. A growing squat. Cracks existed already in the middle in March 2006.

4.2.1.4 Summary

As can be seen from the discussion above the growth of a squat concerns, among other things, two phenomena: the roughly constant wavelength of the W, and the way the W (or the two lungs) is formed. The wavelength is determined by the eigen characteristics of the system, so that the two lungs of a mature squat have roughly the same size. A squat without cracks will grow from the trailing edge of the initial light squat forward in the direction of the traffic, i.e. the initial defect develops into the first half of the squat, and the dynamic contact force excited by the first half makes the second lung.

In conclusion the postulated growth process is validated for squats initiating and growing from defects without cracks or with small cracks. In the presence of deep indentation and large cracks, the growth process will be somewhat altered, but the two-lung shape is still retained by virtue of the wavelength of the dynamic contact force, and

the wavelength is usually not significantly changed if the indentations or cracks are not very severe.

4.2.2 Validation of the relationship between squat growth and friction

In Section 4.1 it was predicted that higher friction level would result in higher plastic strain, which means more squat occurrences and faster squat growth. Figure 4.2.4 shows an observation between braking and squat occurrence on the monitored track section. The upper curve shows a typical speed curve of the trains along the track. This straight track is just before a curve and the trains brake before entering the curve, as can be seen from the speed decrease. The track was ground in later 2005. The middle and lower plots shows all the moderate and severe squats recorded on the track in the period between March 2006 and October 2008. Light squats are not included because there were too many of them and not all of them were counted during the monitoring.

It can be seen that there were many more squats between 25.9 and 26.0km where the gradient of the speed was among the highest and where the speed was low. High deceleration means high tangential force at the wheel–rail contact. Low speed means more frequent activation of the Anti-lock Brake System (ABS) system. It has been observed that for a major type of the trains running on this track the ABS appears to be activated during braking more frequently at lower speed. Because ABS activation is often accompanied by high wheel slip, it means that between 25.9 and 26.0km not only the tangential force could be as high as the available friction level but also the wheel slip was high. This should explain the much higher squat occurrence there: the high slip promoted squat initiation, and the high tangential force promoted squat growth. It should, however, be noted that no wheel skidding marks were confirmed on the rail surface during the three rounds of monitoring in the surveyed period, though something that looked like slip marks were occasionally observed.
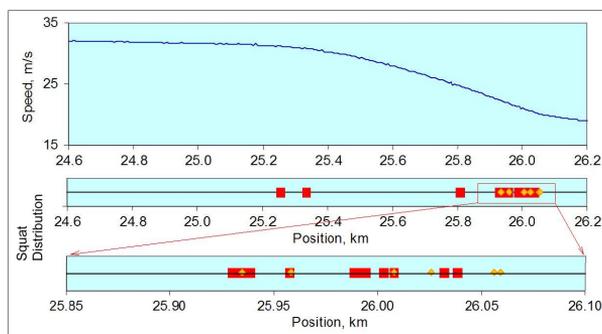
Figure 4.2.4. Upper plot is a typical speed curve along the track between Maarheez and Weert, the Netherlands; middle and lower plots show the locations where moderate and severe squats were recorded in the period between June March 2006 and October 2008. Traffic was from left to right. A red square indicates a severe squat, and a yellow diamond indicates a moderate squat

4.2.3 Validation of the relationship between dynamic contact force and wave pattern

In Section 4.1, a relationship between the dynamic contact force and the wave pattern that often follows squats was revealed, and a similarity between the wave pattern and short pitch corrugation was observed.

It also reported that a correlation analysis finished in March 2006 based on data collected by measuring trains on the ProRail network showed that 72% of the squats were related to corrugation. It was realized later that the corrugation included actually both the real one and those which were corrugation-like wave pattern caused by squats. The two could not be separated during the analysis of March 2006 because from the data available the traffic direction was not always clear.

By January 2007 a second correlation was finished based on a field survey carried out on the ProRail network between February and April 2006. 43 track sections of a total length of 114 km were visited. The tracks were so chosen that they were distributed more or less uniformly over the entire ProRail network, so that statistically reliable results could be obtained. The survey showed that among all the squats 33% was really caused by corrugation, while another 41% had only corrugation-like wave pattern after them. Figure 4.2.5 shows the difference between corrugation and squat-caused corrugation-like wave pattern. Note that on the surveyed tracks the traffic was one-directional.

(a) Squats caused by corrugation

(b) Squat causes corrugation-like wave pattern

Figure 4.2.5. Short pitch corrugation and corrugation-like wave pattern after a squat. The wavelength of the wave pattern is usually in the range of 20 – 40 mm. Traffic is from left to right.

This means that the percentage of squats together with corrugation and corrugation-like wave pattern was 74%. This is in agreement with the 72% of the previous correlation analysis. It is worthy pointing out that the wave pattern caused by squats may disappear

and re-appear. This was predicted in Section 4.1 along with the postulation of the squat growth process.

4.2.4   Validation of the predicted high frequency interaction

In Section 4.1 it was predicted that squats initiation and growth are related to some characteristic wavelength of the wheel–track interaction system. The wavelength is 20–40mm, which corresponds to the frequency range of 950–1900 Hz when the train is moving at a speed of 140 km/h, the typical speed of the Dutch passenger train traffic.

(a)
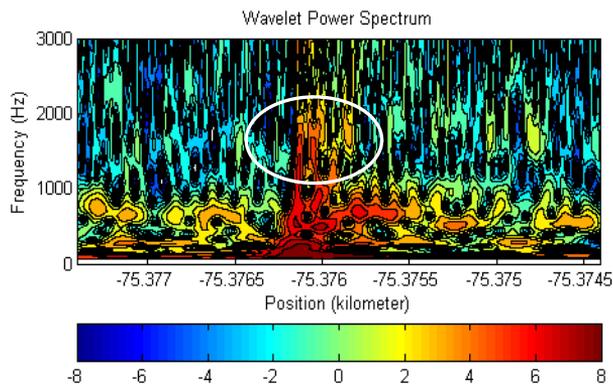
(b)

Figure 4.2.6. Characteristic frequency content in vertical axle box acceleration measured at a squat. (a) A squat and its wave pattern on the right side of it. The traffic was from left to right. (b) The wavelet power spectrum measured at the squat shown in (a). The colour scale in the lower bar indicates energy density. Large scale number means more energy in the corresponding frequency contents. The abscissa of is negative because the measuring train was travelling in the direction of descending kilometre.

The existence of such frequency contents has been confirmed in measurement of axle box acceleration, see Figure 4.2.6. Because the speed of the train for this measurement was around 110 km/h, the corresponding frequency was lower, see the high

energy density (encircled by the white oval) at around 1500 Hz. This component of the vibration should have been excited by the dynamic wheel–rail interaction at the squat which in turn promoted the growth of the squat. The related wave pattern can be seen in Figure 4.2.6(a).

4.3 Wheel–rail impact and the dynamic forces at discrete supports of rails in the presence of singular rail surface defects[5]

ABSTRACT

A validated three-dimensional (3D) transient finite element model is used to evaluate the wheel–rail impact at singular rail surface defects and the resulted high-frequency dynamic forces at the discrete supports of the rail. A typical ballasted railway track is modeled, in which the supports of the rail are composed of the fastenings, the sleepers, and the ballast. The primary suspension of the vehicle is considered. To include all the important eigen characteristics of the vehicle–track system, the wheel set, the rail, and the sleepers are all meshed using 3D solid elements. The transient wheel–rail rolling contact is solved using a surface-to-surface contact algorithm in the time domain. By simulating the steady-state rolling of a wheel set on a smooth rail, the vertical force distribution at the discrete supports is first compared with Zimmermann solution. Afterward, rail surface defects are applied to calculate the resulted dynamic forces at the wheel–rail interface and at the discrete supports of the rail under different rolling speeds. The obtained dynamic responses confirm the necessity of using such a detailed model for the investigations.

4.3.1 Introduction

Singular rail surface defects (SRSDs), which occur in the rail top with a length of 2–8 cm, may under certain conditions come into being in an initially smooth rail. Among them, squats, a kind of rolling contact fatigue as shown in Figure 4.3.1, have become one of the main concerns on the Dutch railway network [4.3.1]. Once these defects exist, the related wheel–rail impact will excite high-frequency dynamic forces at the wheel–rail interface (950–1900 Hz for squats under a typical Dutch line speed of 140 km/h) [4.3.1] and at the discrete supports of rails, due to their characteristic of short wavelength. In a typical ballasted railway track, a discrete support is composed of two fastenings, a sleeper and the ballast under the sleeper. Field observations have shown that these dynamic forces can exacerbate the deterioration rate of the ballast and the fastening and may cause hung sleepers, sleeper cracking, walking, and skewing [4.3.2–4.3.5]. To understand these phenomena and find possible solutions to them, the high-frequency dynamic responses of the vehicle–track system to SRSD need to be calculated, especially the dynamic forces at discrete supports of rails. This is the goal of this work.
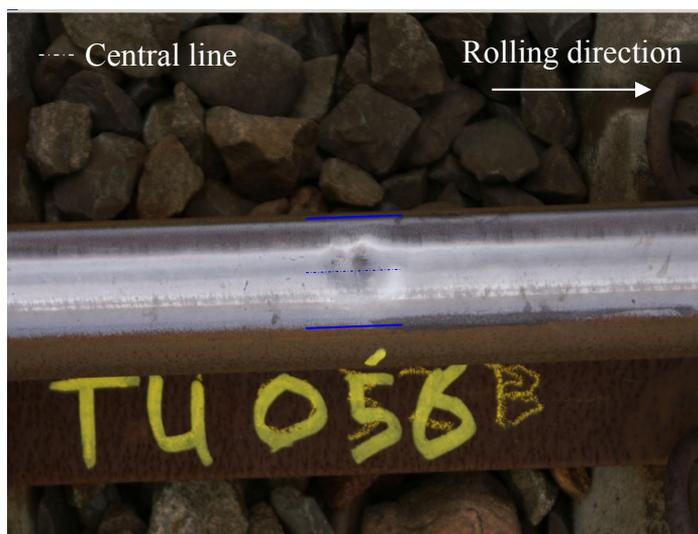
Figure 4.3.1. A typical mature squat in relatively late stage on the Dutch railway network. The rail shown in the figure is 54E1.

In many published work such as [4.3.2–4.3.4, 4.3.6–4.3.13], wheels are approximated as rigid masses and connected to flexible rail beams through simplified contact springs when the dynamic vehicle–track interaction has to be treated. Hereinafter, this widely used approach is referred to as mass–spring–beam approach. According to the benchmark tests published in references [4.3.4, 4.3.6] for models of this approach, the calculated dynamic forces vary significantly with the used models. In authors' opinion, these discrepancies are due to the different assumptions and simplifications used: (a) rails are represented by different beam elements (Bernoulli beam, Timoshenko beam, etc.); (b) the vehicle and track structures are simplified to different extents; and (c) in addition to

the diversity in physical modeling mentioned earlier, the algorithms of solving the formulated mathematic problems also varies. It should be noted that this situation partly results from the fact that each model has its own focus in application, e.g. some models are used to predict the impact at rail joint [4.3.2, 4.3.9] and another ones are for corrugation [4.3.7, 4.3.8].

This situation suggests that special care should be taken in simplifying the system when the mass–spring–beam approach is used for a certain application, and verification of the results is necessary. Nevertheless, for the verification, there are still no reliable means to measure accurately the impact forces at the wheel–rail interface nowadays, especially when the distributions of stresses and strains in the contact patch are concerned.

Recently, other modeling approaches such as the ones in references [4.3.14] to [4.3.18] have also been developed to overcome the drawbacks of the mass–spring–beam approach. Among these, influence of the wheel set flexibility was examined in references [4.3.14] to [4.3.16], whereas detailed structures of the wheel and rail were included in references [4.3.17] and [4.3.18] to take into account their flexibilities more accurately. In reference [4.3.15], it was also reported that under certain conditions, the rotation of the wheel may also play an important role in the high-frequency vehicle–track dynamics.

For predicting the dynamic responses of the vehicle–track system to SRSD, which is in the high frequency range as mentioned earlier, the mass–spring–beam approach may not be suitable any more. The reasons are as follows: on one hand, the detailed structure of the system and the continua of the wheel and rail are ignored, whereas they may play important roles in the high-frequency dynamics through certain local structural vibrations in the vicinity of the contact patch and the stress wave propagation in the wheel and rail continua; on the other hand, the contact filter effect related to the finite contact patch may be important in the presence of SRSD of certain sizes [4.3.19, 4.3.20] but cannot be included in a contact spring due to the assumption of infinitesimal contact patch.

Furthermore, the mass–spring–beam approach embodies only the vertical wheel–rail interaction (related to the normal contact) [4.3.21, 4.3.22], whereas the dynamics in the tangential plane remains an open problem, especially in the high-frequency range in which the non-steady-state contact [4.3.7, 4.3.8] must be included. Such a situation probably results from the fact that the non-linear characteristic of the friction force (or the creep force) cannot accurately be represented in the simplified contact springs. In contrast, it has been observed that the tangential load transmitted at the wheel–rail interface plays a crucial role in the occurrence of squats [4.3.1, 4.3.23].

With the intention of investigating the wheel–rail impact at SRSD and the resulted high-frequency dynamic responses of the vehicle–track system, a three-dimensional (3D) transient finite element (FE) model of the vehicle–track interaction has been built up [4.3.1]. The relevant structures of the vehicle and the track are both considered in the model together with the actual geometries of the wheel set and the rail. Thus, the local structural vibrations and the stress wave propagation in the continua are included in the result. Moreover, a detailed 3D contact model is integrated to calculate the transient frictional rolling of a wheel set on a rail in the time domain for any desired loads and speed. Both the dynamic vertical and tangential contact forces can be obtained. The

contact filter effect is taken into account automatically by the finite size of the contact patch.

Based on an analysis of the dynamic contact forces at SRSD obtained from the 3D FE model in combination with field observations, a growth process of squats was postulated in reference [4.3.1] and subsequently verified by field monitoring tests [4.3.23]. Measured axle box acceleration has also been used to validate the 3D FE model [4.3.23, 4.3.24]. Therefore, this model is valid for the calculation of the high frequency vehicle–track interaction. In addition, the 3D FE model has also been validated in the normal and the tangential contact solutions against Hertz theory and Kalker's CONTACT [4.3.25] and in predicting the influence of friction force on the occurrence of squats [4.3.23]. Note that another important reason for using such a detailed model is to obtain distributions of the dynamic stresses at SRSD, which is necessary for analysis of rolling contact fatigue and cannot be realized by the mass–spring–beam approach.

In the past work performed with the 3D FE model, focus has been on the wheel–rail interface, such as their interaction forces and stresses [4.3.1, 4.3.25], to understand the occurrence of SRSD, and on the axle box acceleration [4.3.24, 4.3.26], to develop a train-borne measurement method. With the goal of assessing the influence of SRSD on track deterioration, the excited dynamic vertical forces at discrete supports of rails are examined in this work with the same FE approach. In this article, a defect profile adjusted from a measured profile at the squat shown in Figure 4.3.1 and the measured profile are simulated in sections 4.3.3.2 and 4.3.3.3, respectively, to investigate the dynamic vertical forces in the presence of SRSD and to examine the influence of defect geometry. Moreover, the rolling speed is also varied to study its effects.

4.3.2   Model description

4.3.2.1   The 3D transient FE model

The 3D transient FE model of the vehicle–track interaction system is schematically shown in Figure 4.3.2(a). A car of a Dutch double-decker passenger train (VIRM) with new wheels and a typical ballasted track of the Dutch railway are considered. As the high-frequency dynamic forces, such as the ones related to SRSD, are mainly attributed to the flexibilities of wheel set and track [4.3.3, 4.3.14], the sprung mass of the vehicle is lumped into Mc that is connected to the wheel set through the primary suspension (Kc and Cc). Considering that the flexibilities of the vehicle–track system can effectively filter out the transmission of the high-frequency dynamic forces from one wheel–rail interface into the other, only a half wheel set and a half straight track are modeled in view of the symmetry of the system. The rail profile is the standard 54E1 with an inclination of 1:40. The tread conicity of the wheel is also set to be 1:40. Thus, the wheel–rail contact occurs in the middle of the rail top, which is required for the research into the squat (Figure 4.3.1).

To include all the necessary eigen characteristics of the vehicle–track system related to SRSD, the wheelset, the rail, and the sleepers are all meshed using 3D solid elements with the consideration of their actual geometries (Figure 4.3.2(b)). The total numbers of the elements and nodes are 260 071 and 295 849, respectively. The minimum element

size is 1.1 mm, which exists in part of the rail top where the defect is applied. Table 4.3.1 lists the values of the involved parameters, from which it can be seen that no material nonlinearity is included. This is valid for the present work due to the negligible influence of plastic deformation on dynamic vehicle–track interaction (in a macro scale). Note that a dynamic factor of 30% is included in determining the static load to take into account the dynamic loads in low frequencies and the un-uniform distribution of car-body weight on different wheels.
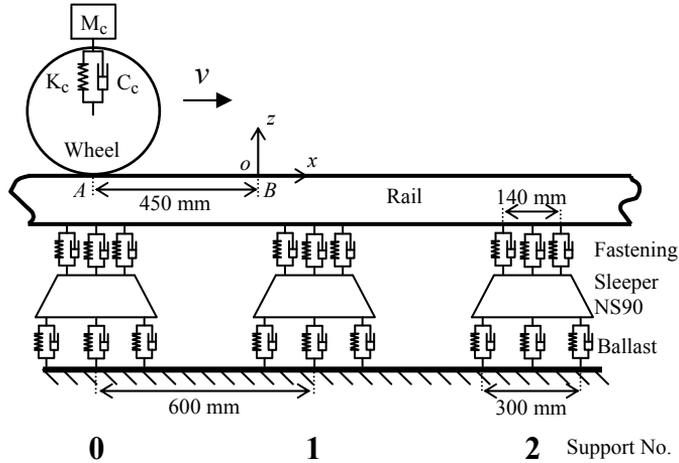
(a)
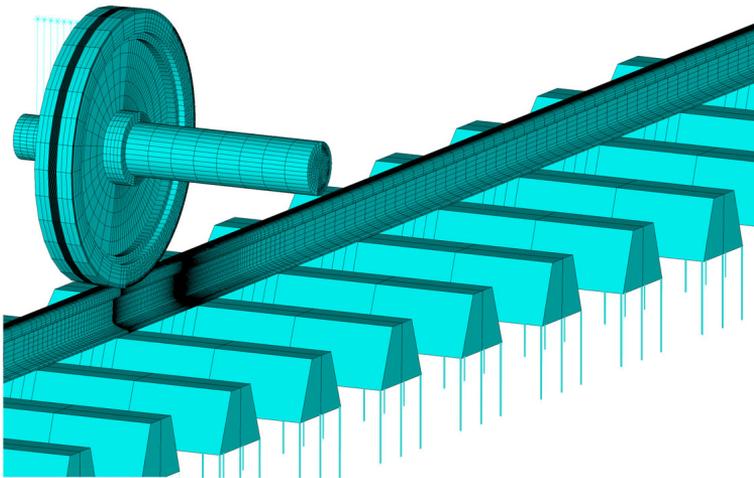
(b)

Figure 4.3.2. The representation of the 3D vehicle–track FE model. (a) A schematic diagram and (b) The mesh.

For calculation, the lateral movement of the wheel set is constrained by applying symmetric boundary conditions, and the rail supports are only allowed to move vertically. Being different from the model in reference [4.3.1], a fastening and the ballast under a sleeper are both modeled with 12 parallel springs and dampers (three columns in the

longitudinal and four rows in the lateral), so as to simulate more realistically the constraints at the bottoms of the rail and of the sleepers. By this, the pressure distribution at the rail seat of sleepers can also be obtained with reasonable accuracy. Another two important improvements of the current model with respect to the one in reference [4.3.1] are that the length of the simulated track is extended to 9.14 m or 15 sleeper spans, and the sleepers are modeled with solid elements instead of lumped mass points.

For convenience, the discrete supports of the rail are numbered according to their positions. As indicated in Figure 4.3.2(a), the numbering of the supports increases along the rolling direction and the one under the initial position of the contact patch centre, i.e., position A, is set as zero. Therefore, the supports before position A are counted as negative. Correspondingly, Fastening i (i is integer), Sleeper i, and Ballast i are numbered for those at Support i. A 3D right-handed Cartesian coordinate system ($Oxyz$) is defined as illustrated in Figure 4.3.2(a). For easy reference, the origin (O) is located at the centre of the vertical-lateral profile of the rail top and is 0.45 m away from position A in the longitudinal direction. The z axis is in the vertical direction.

Table 4.3.1. The values of the parameters in the model.

Parameters		Values	Parameters		Values
Wheel diameter		0.92 m	Wheel &rail material	Young's modulus	210 GPa
Static wheel load		116.8 kN		Poisson's ratio	0.3
Primary suspension	Stiffness	880 kN/m		Density	7800 kg/m^3
	Damping	4000 Ns/m			
Railpad	Stiffness	1300 MN/m	Sleeper material	Young's modulus	38.4 GPa
	Damping	45000 Ns/m		Poisson's ratio	0.2
Ballast	Stiffness	45 MN/m		Density	2520 kg/m^3
	Damping	32000 Ns/m		Mass	280 kg

Note: The stiffness and damping of the rail pads and ballast are taken from reference [4.3.27].

4.3.2.2   The rail surface defect and a typical simulation

Two-dimensional (2D) profiles of SRSDs in a longitudinal-vertical plane are recorded by RailProf during monitoring tests. For a straight rail of 54E1, the profile is always measured along the central line of the rail top (Figure 4.3.1) due to limitation of Railprof. Clearly, if an SRSD deviates from this measuring line in location, its critical geometry may not be recorded by such a measurement. Taking the mature squat shown in Figure 4.3.1 as an example, its two-lung appearance [4.3.23] is not revealed clearly in the recorded profile shown in Figure 4.3.3(a) (Defect M0, M: measured). From the color at the bottom of the squat in Figure 4.3.1, it can be seen that the maximum or critical geometric deviation at the squat should be larger than the one of Defect M0 (the deeper parts of a squat look darker as the possibility of being in contact is lower).

In such case, the measured profile should be adjusted based on other available information such as visual appearance and measurements of series of SRSDs. In reference [4.3.26], an approach of profile adjustment was developed and validated by matching axle box acceleration numerically reproduced from a squat to measured signal. In this work, the same approach is used to adjust Defect M0 with some randomness, leading to Defect M1 as the critical geometric deviation of the squat. Focus of this article is, therefore, placed on the dynamic forces excited by Defect M1, which is applied on the

rail top with its leading edge located at the origin of *Oxyz* (Figure 4.3.3(a)). Furthermore, to derive the 3D geometry required, it is assumed that depth distribution of the squat is parabolic along the lateral direction with the deepest position located in its lateral centre ($y = 0.0$), as shown in Figure 4.3.3(b).

(a)

(b)
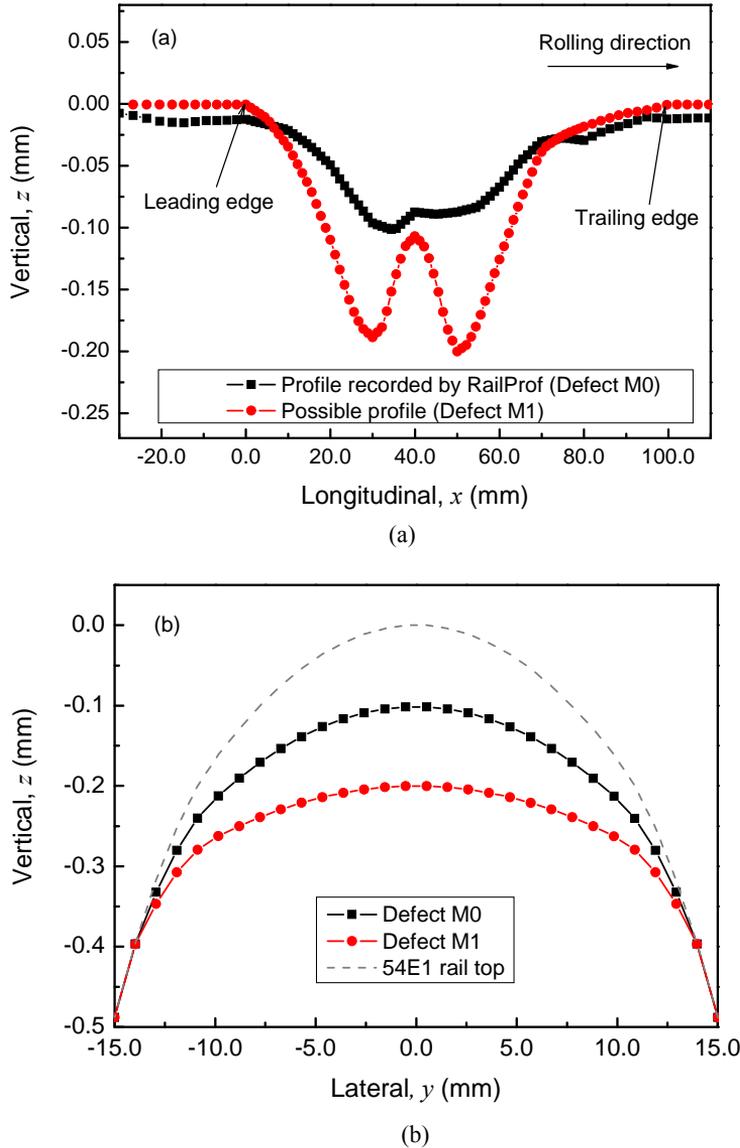
Figure 4.3.3. The simulated defect models. (a) The longitudinal-vertical profile at $y = 0.0$ and (b) profiles at the deepest lateral-vertical cross-section. The symbols in the figures represent the nodes in the mesh.

In each simulation, the equilibrium state of the system under gravity is first reached at the initial position of the wheel (position *A* in Figure 4.3.2(a)). Then, the movement of the wheel in rotation and forward translation is applied as the initial conditions, and

subsequently, the wheel is set to roll in the presence of friction along the rail from position *A* toward the defect. The effects of transient wheel rotation are, therefore, included inherently in the simulation. The contact patch moves along the contact surfaces of the wheel and rail as the wheel rolls forward. If a drive or braking torque is further applied onto the wheel axle, a specified acceleration or deceleration can be simulated.

A detailed contact model with a surface-to-surface contact algorithm [4.3.28] is used to solve the transient wheel–rail rolling contact in the time domain. For calculation, the potential contact areas are defined large enough to cover the rolled contact surfaces of both the wheel and the rail. The rolling distance *AB* (Figure 4.3.2(a)) is designed for the system to reach a steady state before entering the defect. When the wheel rolls over the defect, the dynamic forces at the wheel–rail interface and the discrete supports are obtained. For comparison, a reference case (Case R) in which no defect is applied is also simulated.

In this work, the typical passing speed of the trains at the location of the squat shown in Figure 4.3.1, i.e., 100 km/h, is assumed to be the default rolling speed. Based on the Dutch network situation, another three different rolling speeds, namely 140, 70, and 40 km/h, are also simulated to examine their influences. From now on, the default speed is simulated unless specified.

It is well known that the normal contact is independent of the tangential contact if the contact bodies are composed of the same elastic material. Therefore, the presented results in this work, which are only for the vertical (normal) direction, are valid for both free and frictional rolling. The dynamic tangential forces transmitted at the wheel–rail interface, which can also be obtained from the 3D FE model, are not discussed in this article because the rail supports are only simulated in the vertical direction.

4.3.2.3   The sleeper FE model

The sleepers are more coarsely meshed than the wheel and the rail in the detailed 3D FE model of Figure 4.3.2 to reduce the model size. Therefore, another FE model of a sleeper with fine mesh, as shown in Figure 4.3.4, is used for stress calculation of sleepers in static condition with the pressure results exported from the vehicle–track simulations as input. The sleeper type of NS90, which is widely used on the Dutch network, is modeled with its actual geometry. Only half length of a sleeper is considered to keep consistence with the vehicle–track FE model in Figure 4.3.2, and special care is taken in meshing the rail seat area. For calculations, a fixed boundary condition is applied to the sleeper bottom.

4.3.2.4   Zimmermann solution

In this work, Zimmermann solution, which is proposed to evaluate the longitudinal distribution of the vertical force at discrete supports in European standards (such as in CEN/TC 256/SC 1/WG 16N XXX1E), is also obtained for comparison in this work. The next paragraphs briefly introduce the process of getting Zimmermann solution for readers' convenience.

Simulating the track as the discretely supported model shown in Figure 4.3.5, the vertical force $F(x_i)$ at the *i*th support is governed by:

$$F(x_i) = k_d w(x_i) \tag{4.3.1}$$

where, k_d is the stiffness of each support, and $w(x_i)$ and x_i are the deflection of the rail in the vertical direction at the ith support and the longitudinal coordinate of the ith support, respectively.
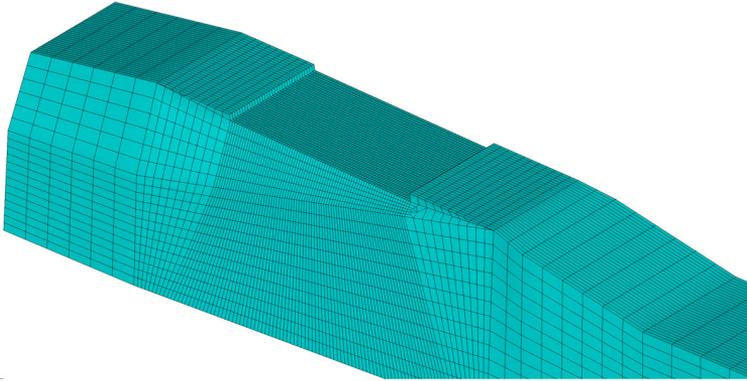
Figure 4.3.4. A 3D FE model of sleeper NS90.

For calculating the vertical deflection of the rail analytically, the discrete supports of the rail are further approximated as Winkler support with the stiffness (k) calculated by k_d and a (the distance between two supports as shown in Figure 4.3.5) in Equation (4.3.2).

$$k = k_d / a \tag{4.3.2}$$

Setting the origin of the coordinate system at the position where the contact force is applied, the rail deflection is governed by Eqs. (4.3.3)–(4.3.5) under the assumption of Winkler support, in which the rail is only considered by its bending stiffness (EI).

$$w(x) = \frac{Q}{2kL} \eta(x) \tag{4.3.3}$$

$$\eta(x) = e^{-x/L}[\cos(x/L) + \sin(x/L)], \; x \geq 0 \tag{4.3.4}$$

$$L = \sqrt[4]{\frac{4EI}{k}} \tag{4.3.5}$$

where, Q is the wheel load, L is the characteristic length of the track model, and x is the longitudinal coordinate.
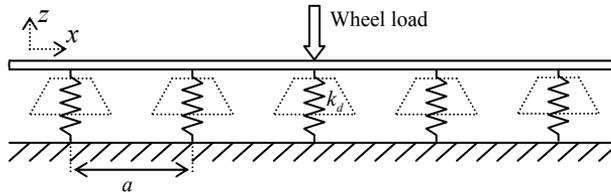
Figure 4.3.5. The discretely supported track model.

4.3.3 The dynamic vertical force at discrete supports

4.3.3.1 Comparison with Zimmermann solution

The vertical force distributions at the discrete supports obtained from Case R are compared to the Zimmermann solution, as shown in Figure 4.3.6. The presented results correspond to two different instants, in which the centre of the contact patch is located at the positions of $x = -150$ mm and $x = 75.5$ mm, respectively. Referring to Figure 4.3.2(a), it is found that the position of $x = -150$ mm is located in the middle of Supports 0 and 1. For direct comparisons with the Zimmermann solutions, the vertical force at every support obtained from Case R is converted into force ratio with respect to the wheel–rail vertical contact force at the chosen instant. Only five sleepers that are loaded most heavily are plotted in Figure 4.3.6. The position of the contact patch centre is, hereinafter, referred to as the position of the wheel for short.

The difference between the Zimmermann and the FE solutions is 11.0% and 8.5% at Sleeper 1 in Figures 4.3.6(a) and (b), respectively. From engineering point of view, the agreement is good between the two solutions. Considering that the Zimmermann approach is based on statics, this agreement on one hand confirms that the steady-state rolling of the wheel on the rail is approximately reached in Case R when the wheel rolls over a half sleeper bay. On the other hand, it also indicates that the vertical force distribution at discrete supports obtained from the transient FE simulation is valid.

4.3.3.2 Defect M1

a) The dynamic vertical forces

The dynamic vertical forces excited by Defect M1, which are extracted by subtracting the corresponding results of Case R from the resultant forces at the defect, are plotted versus the longitudinal position of the wheel in Figure 4.3.7(a). It can be observed that negative values corresponding to the load reduction with respect to steady state (Case R) occur. Note that no loss of contact occurs between the wheel and the rail for all simulated cases presented in this paper. Hereinafter, the dynamic vertical force is referred to as the dynamic force for short considering that the vehicle–track interaction is only analyzed in the vertical direction in this work.

From Figure 4.3.7(a) it can be seen that a dynamic force with a peak magnitude (the largest magnitude) of 67.1 kN is excited by Defect M1 at the wheel–rail interface. Performing a fast Fourier transform (FFT) analysis, it is found that this dynamic force mainly oscillates with two frequencies, which are about 297 and 1061 Hz, respectively, as indicated as f_1 and f_2 in Figure 4.3.7(b). Considering their corresponding wavelengths of about 93 and 26 mm under the simulated rolling speed, the two vibration components can also be easily identified in Figure 4.3.7(a). Furthermore, the force oscillation corresponding to the vibration at f_2 agrees well with the geometric deviation of the squat shown in Figure 4.3.3, which explains the growth of a squat [4.3.1]. It should be noted that this agreement does not necessarily mean that the dynamic force variation just simply follows the geometric deviation at the simulated defect. About this point, more discussion is given in Section 4.3.4. Additionally, the sample interval of the FFT analysis

is 2.3×10^{-5} s for the results with the default rolling speed. As the rolling speed varies, the interval also changes correspondingly.

(a)

(b)
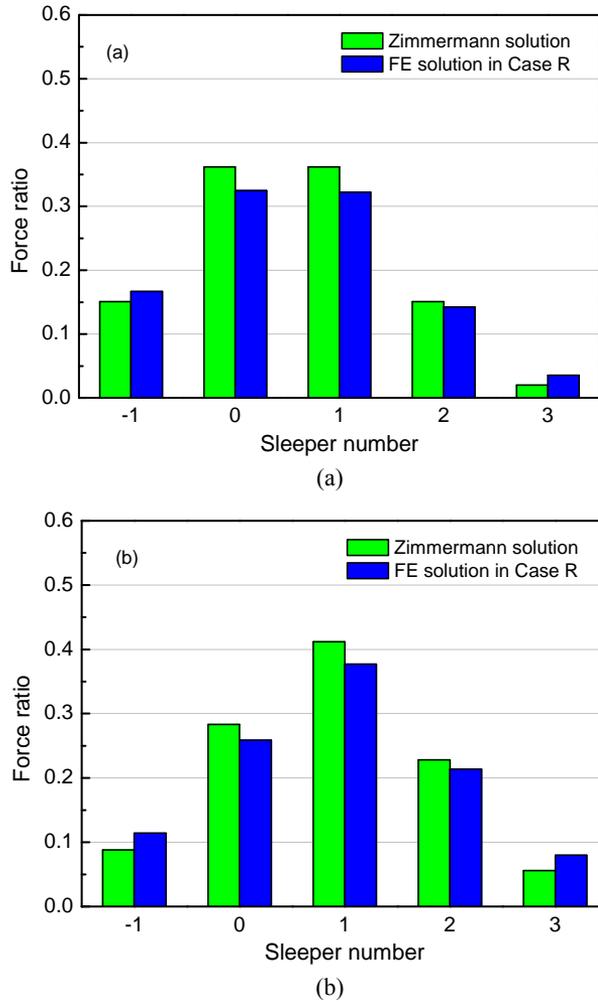
Figure 4.3.6. Comparison of the vertical force distributions at discrete supports for steady state: the Zimmermann and the FE solutions. (a) The wheel is located at $x = -150$ mm, and (b) the wheel is located at $x = 75.5$ mm.

Following the dynamic force at the wheel–rail interface, it can be seen that the vibration at f_2 dies out in 3 waves, which is related to the typical two-lung shape of squats (see Figure 4.3.1) and the possible wave pattern after squats [4.3.1]. Afterwards, the dynamic force variation is dominated by the vibration at f_1, because it attenuates more slowly than the one at f_2 in the simulated system and can last up to 4–5 wavelengths, as shown in Figure 4.3.7. Trial simulations show that the attenuation of the vibration at f_1 is considerably increased when higher damping exists in the track structure. Moreover, the magnitude of the vibration at f_1 is clearly higher than at f_2 for Defect M1 (see Figure

4.3.7(b)). Further simulations show that this phenomenon typically occurs for mature squats with relatively large sizes, and results from the fact that larger defects tend to excite the system more in the relatively low frequency if the rolling speed remains constant.

(a)

(b)
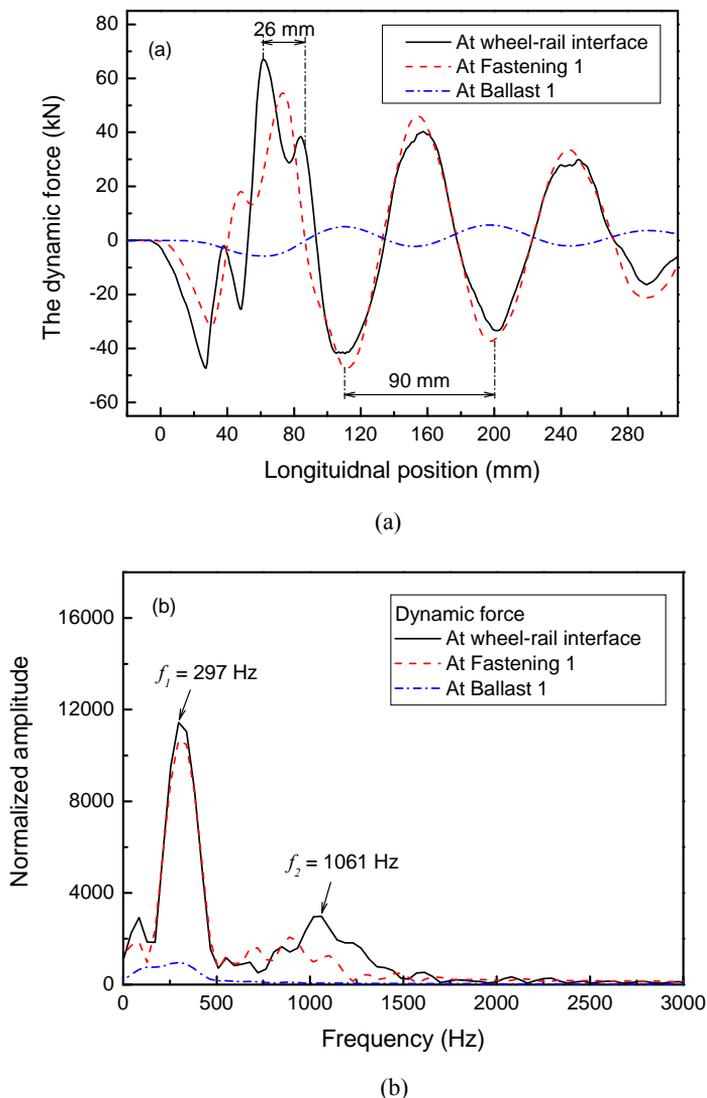
Figure 4.3.7. The dynamic forces at the wheel–rail interface and Support 1 caused by Defect M1. (a) The force variation in the time domain and (b) the amplitude spectrum.
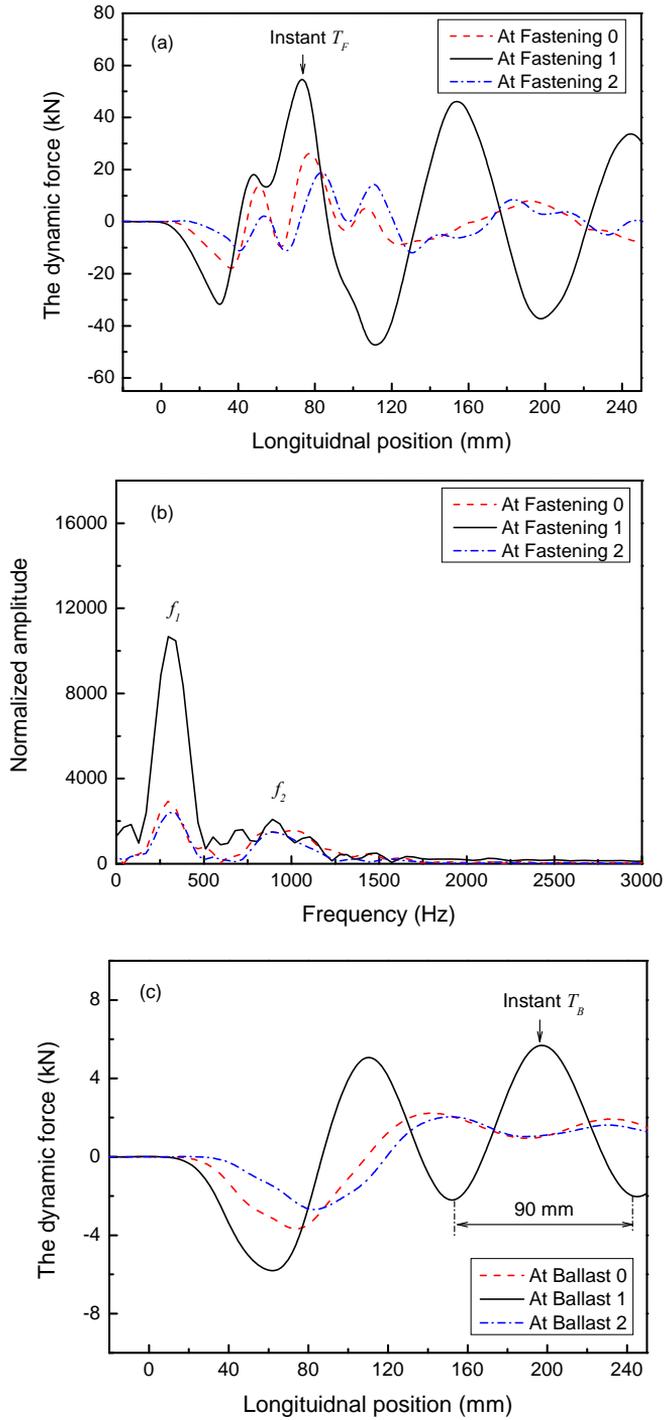
Figure 4.3.8. The dynamic forces caused by Defect M1. (a) At fastenings, (b) the amplitude spectrum at fastenings, and (c) at ballasts.

Through the discrete supports of the rail, the dynamic wheel–rail interaction excited by the defect is transferred downwards. The vibration at f_1 transfers down to the ballast, with a constant frequency at the wheel–rail interface, the fastenings, and the ballast as shown in Figure 4.3.7(b). In contrast, f_2 differs for the dynamic forces at the wheel–rail interface and the fastenings, and is almost absent at the ballast, suggesting different transmission of the corresponding vibration. Further, the vibration at f_1 is mainly transferred to Support 1 that is just below Defect M1, resulting in higher magnitudes of the dynamic forces at Fastening 1 and Ballast 1 than at the others as shown in Figures 4.3.8(a) and 4.3.8(c). However, the magnitude of the vibration at f_2 is more or less the same at the three fastenings shown in Figure 4.3.8(b). Table 4.3.2 summarizes the characteristics of these two vibration components in transmission.

Table 4.3.2. Transmission characteristics of the vibration components at f_1 and f_2.

Vibration component at	f_1	f_2
Transfer downwards	To fastenings and ballast	To fastenings
Transfer longitudinally	Mainly to the closest support	To more supports, and with similar magnitude
Frequency	Constant at the wheel–rail interface, fastenings, and ballast.	Differ at the wheel–rail interface and fastenings

At relatively low frequency range, the coupling between the wheel and the track through the contact is weak, so that for the track sub-system, the dynamic force at the wheel–rail interface can be considered as the external excitation. The track is in forced vibration. This explains why the oscillations of the dynamic forces related to the vibration at f_1, which transfers down to the ballast, have the same wavelengths at the wheel–rail interface, the fastenings and the ballast. Because of the transmission process of the vibrations, phase shift exists between them, also between the dynamic forces at different supports. However, at relatively high frequency range, the coupling between the wheel and the track becomes significant, leading to the considerable influence of the eigen characteristics of track sub-system on the dynamic contact force between the wheel and the rail. This may be why f_2 is slightly different at the wheel–rail interface and fastenings.

It is interesting to notice that the peak magnitude of the dynamic force at Fastening 1 is close to the one at the wheel–rail interface (see Figure 4.3.7(a)). The dynamic force at Ballast 1, however, has a much lower peak magnitude because the ballast is effectively isolated from the vibration by the inertia of the sleeper.

At the instant when the dynamic force at Fastening 1 reaches the maximum (Instant T_F is used from now on), the wheel is located at the position of $x = 73.3$ mm, as indicated in Figure 4.3.8(a). The vertical force distribution at different fastenings at Instant T_F is plotted in Figure 4.3.9(a) with green columns together with the corresponding results of Case R (blue columns). It can be observed that the vertical force is significantly increased by Defect M1 at the two fastenings of the sleeper bay (Fastenings 0 and 1) where the defect is applied. However, the forces at Fastenings −1 and 3 are zero at that instant. This difference in trend at different fastenings can mainly be attributed to two factors: one is that the peak magnitude of the dynamic force at fastenings reaches the maximum at

Fastening 1, and the other is the phase shift of the dynamic forces at different fastenings (see Figure 4.3.8(a)). For the vertical forces at ballasts, however, they are increased at all the supports shown in Figure 4.3.9(b) at Instant T_B when the dynamic force at Ballast 1 reaches the maximum ($x = 197.2$ mm, see Figure 4.3.8(c)).
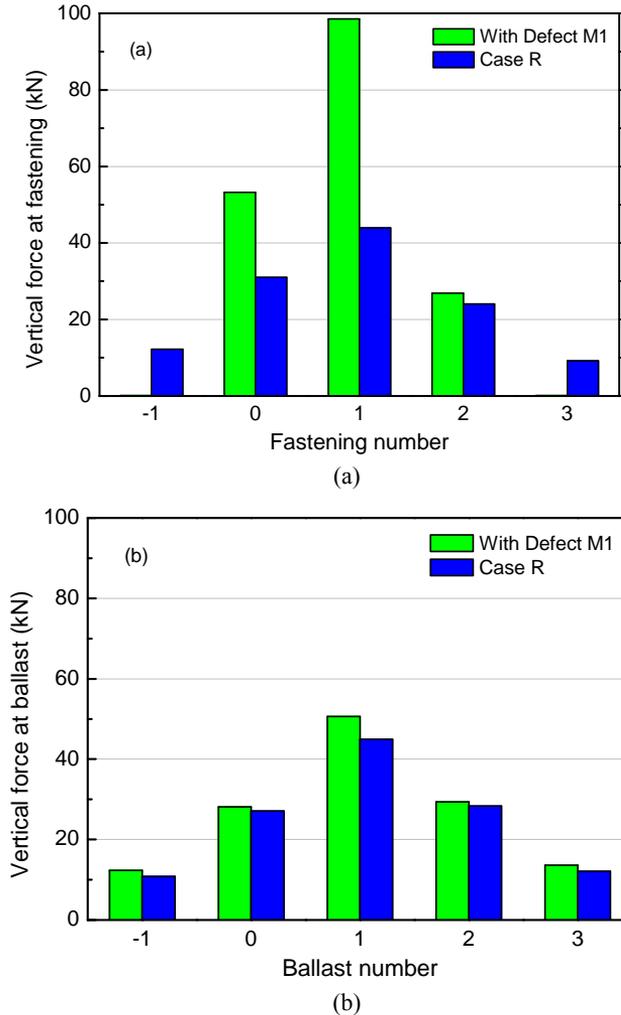
Figure 4.3.9. The vertical force distribution at discrete supports when Defect M1 exists. (a) Fastening forces at Instant T_F, and (b) ballast forces at Instant T_B.

b) The pressure and V-M stress at Sleeper 1

Figure 4.3.10(a) shows the pressure distribution in the rail seat area of Sleeper 1 at Instant T_F when Defect M1 exists, whereas the corresponding result of Case R is plotted in Figure 4.3.10(b). Clearly, the pressure is not uniformly distributed in both cases due to the deflection of the rail. As a result of the dynamic force related to Defect M1, the maximum pressure at Sleeper 1 rises to 4.8 MPa from 2.5 MPa in Case R, and the minimum increases by 2.7 times (from 0.46 to 1.7 MPa).
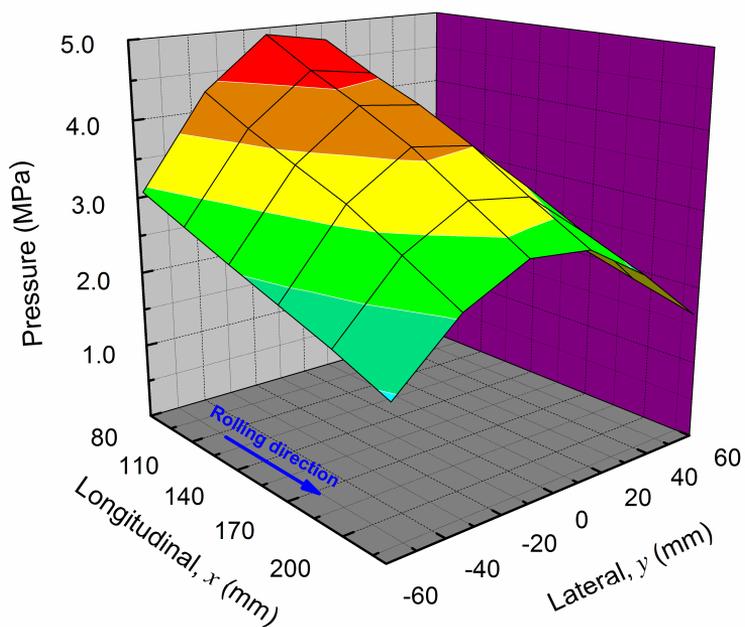
(a)
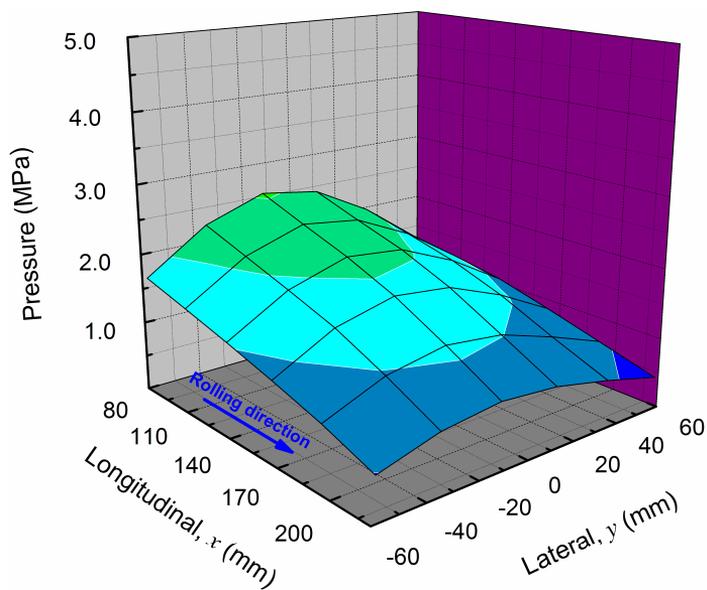
(b)

Figure 4.3.10. The pressure distribution in the rail seat area of Sleeper 1 at Instant T_F. (a) When Defect M1 exists, and (b) Case R.

Applying the pressure distributions shown in Figure 4.3.10 onto the sleeper FE model, detailed stress distributions in the sleeper such as the von-Mises (V-M) stress contours shown in Figure 4.3.11 can be obtained. For clear comparison, the same legend

is applied to the two cases shown in Figure 4.3.11. The stress level increases from blue to red in colour. It is seen that the pattern of the V-M stress distribution is similar in both situations, although the maximum V-M stress is increased from 3.38 MPa to 6.71 MPa by Defect M1.

It is worthy to mention that relatively high stress level only exists in a limited part of a sleeper close to the rail seat area, as can be seen from the stress contours shown in Figure 4.3.11. This confirms that it is sufficient to apply a fixed boundary condition to the sleeper bottom for the stress analysis of a sleeper.

(a)

| 0 | 50 | 0.5E6 | 0.9E6 | 1.5E6 | 2.4E6 | 3.50E6 | 5.0E6 | 6.7E6 | Unit: Pa |

(b)
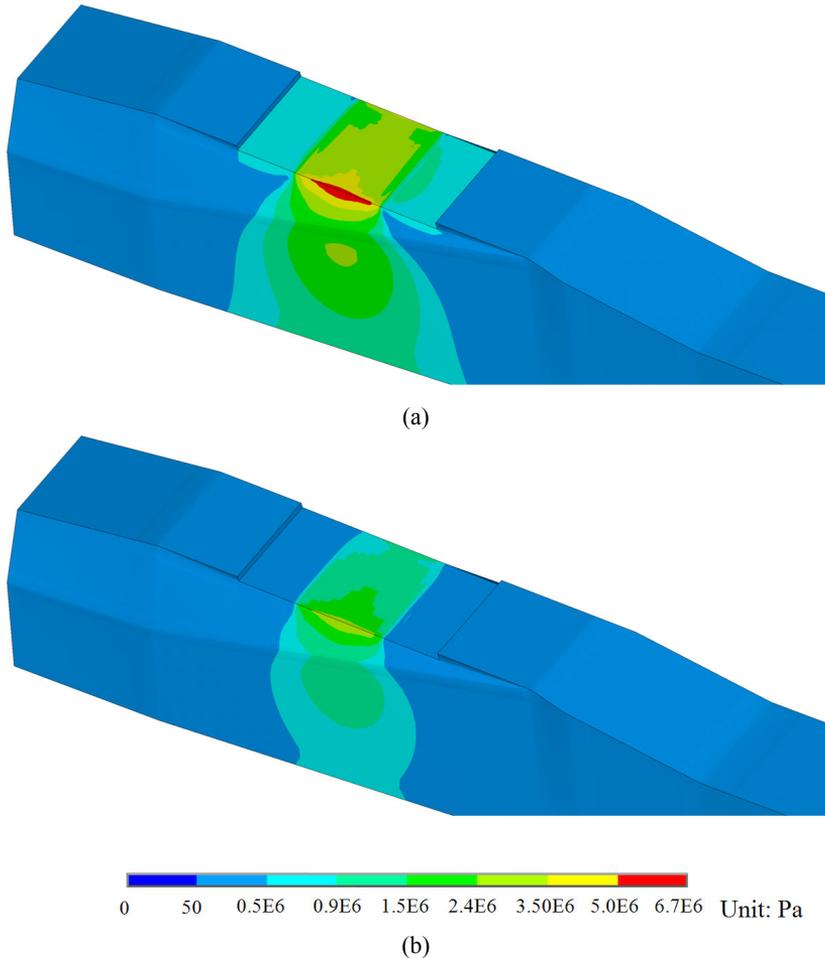
Figure 4.3.11. The V-M stress contour at Sleeper 1 at Instant T_F. (a) When Defect M1 exists (V-M $_{max}$ = 6.71 MPa) and (b) Case R (V-M $_{max}$ = 3.38 MPa).

c) The influence of rolling speed

In the cases discussed above, the default rolling speed of 100 km/h has been used. To examine the influence of the rolling speed, the dynamic responses of the vehicle–track

system to Defect M1 are also calculated under another three different speeds, namely 140, 70 and 40 km/h. Figure 4.3.12 shows the dynamic forces at the wheel–rail interface excited by Defect M1 under different speeds. It can be observed that the vibration at f_2 gradually reduces in both magnitude and frequency with the decrease of the rolling speed, and even disappears in the case of 40 km/h (see Figure 4.3.12(b)). For the vibration at f_1, although its magnitude also reduces with the reduction of the rolling speed, its frequency keeps constant correspondingly, which leads to the proportional increase of its wavelength with the rolling speed. As a consequence, the dynamic force at the wheel–rail interface oscillates differently under different speeds (see Figure 4.3.12(a)). The impact position, where the wheel–rail contact force reaches the maximum, occurs at different locations when the rolling speed varies, as indicated in Figure 4.3.12(a).

(a)

(b)
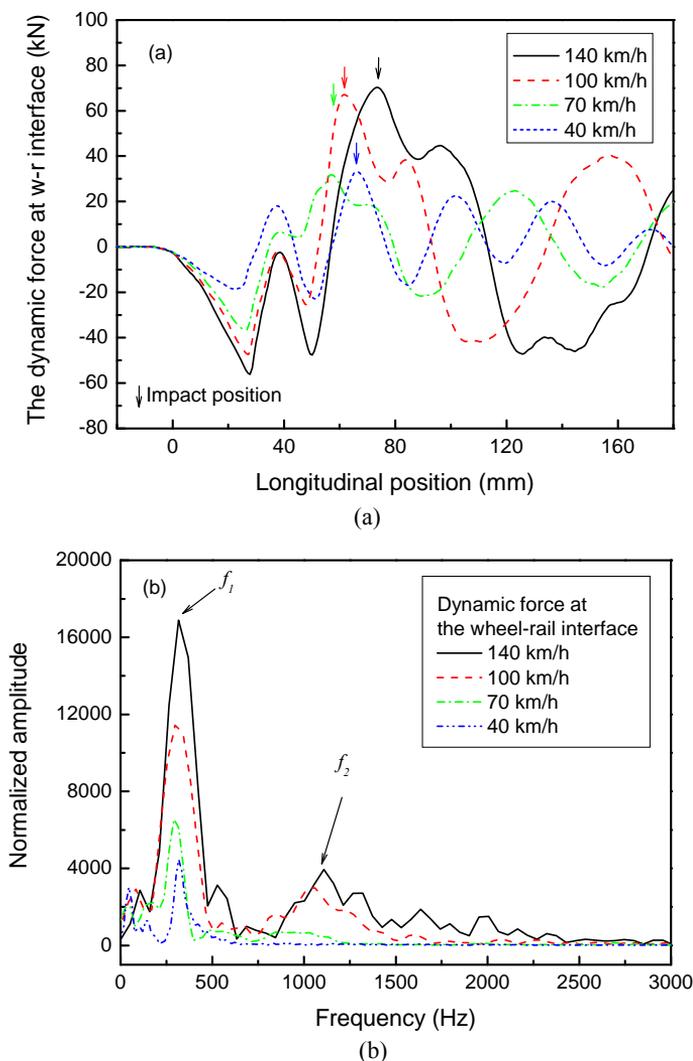
Figure 4.3.12. Defect M1 excited dynamic forces at the wheel–rail interface under different speeds. (a) The force variation in the time domain, and (b) the amplitude spectrum.

Because of the different characteristics in transmissions of the vibrations at f_1 and f_2 into the rail supports, the dynamic forces at Fastenings 0 and 1 vary in a different way when the rolling speed changes, as shown in Figure 4.3.13. Observing into more detail, it can be found that the downward transmission of the vibration at f_2 is also influenced by the rolling speed. For instance, contrary to the wheel–rail contact force, the vibration at f_2 reduces in magnitude at fastenings shown in Figure 4.3.13 when the speed increases from 100 km/h to 140 km/h. In contrast, the peak magnitude of the dynamic force at Ballast 1

(a)

(b)
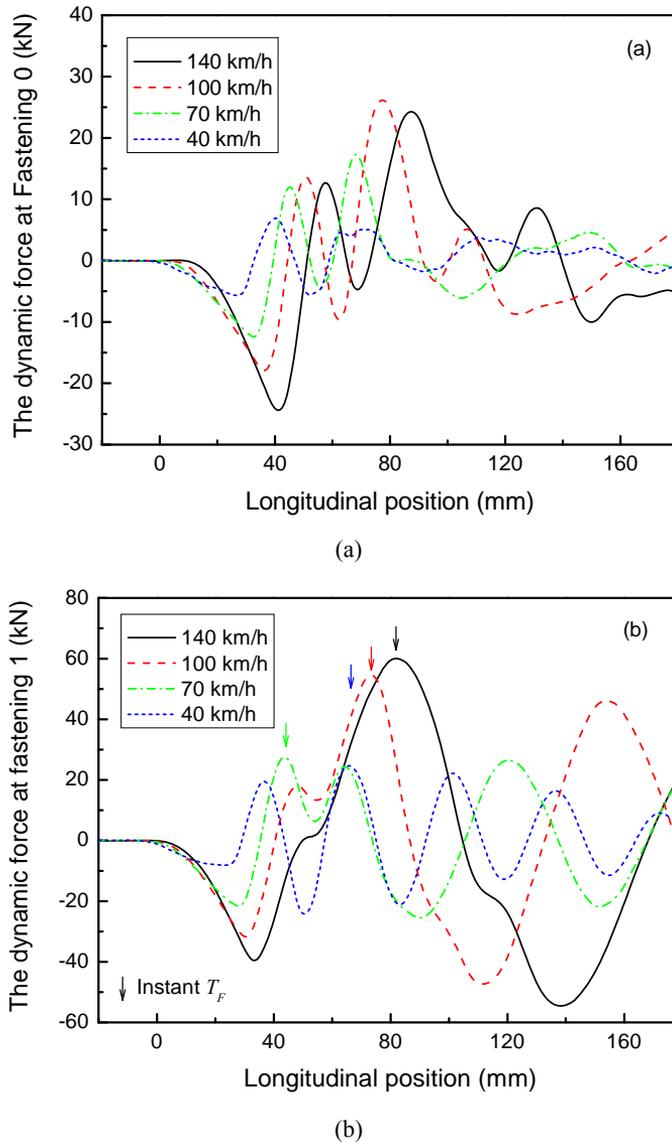
Figure 4.3.13. Defect M1 excited dynamic forces at (a) Fastening 0 and (b) Fastening 1 under different speeds.

is almost independent from the rolling speed (see Figure 4.3.14), which is attributed to the fact that a vibration with a frequency lower than f_1 is excited at ballast under low speeds. Such a vibration can be seen clearly in the result of the case with 40 km/h in Figure 4.3.14, considering that the wavelength of the vibration at f_1 is about 35 mm under the speed. No more efforts will be spent on this vibration in this work because it is only important at ballast under low speed and has relatively low magnitude. In addition, as indicated in Figures 4.3.13(b) and 4.3.14, the positions of the wheel at Instants T_F and T_B change with the rolling speed owing to the variations of the vibration components (such as the ones at f_1 and f_2) in wavelength and in relative magnitude.
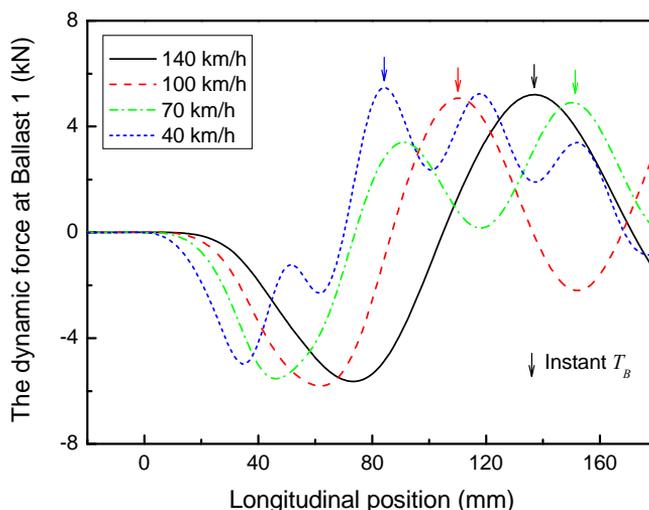
Figure 4.3.14. Defect M1 excited dynamic forces at Ballast 1 under different speeds.

To show the influence of rolling speed more clearly, the peak magnitude of the dynamic forces at Fastenings 0 and 1, and Ballast 1 are plotted versus rolling speed in Figure 4.3.15(a) together with the dynamic force at the wheel–rail interface. It can be seen that the dynamic force at Fastening 1 is close to the one at the wheel–rail interface in magnitude under every rolling speed for the simulated system, whereas the one at Ballast 1 is almost constant in magnitude in the speed range under discussion as mentioned above. Moreover, it is clear that the increase rate with speed is not constant for the dynamic forces shown in Figure 4.3.15(a). The dynamic force at Fastening 0 even reduces when the rolling speed rises from 100 to 140 km/h. These phenomena might be caused by the fact that the simulated squat was taken from a track with a traffic speed of 100 km/h. When a different speed is assumed, the dynamic responses of the system may not be "well-matched" to the existing geometric deviation.

With the variations of the dynamic forces under different rolling speed, the stress level in the sleepers also changes significantly, as shown in Figure 4.3.15(b). The maximum and minimum pressures at Sleeper 1 (at Instant T_F) as well as the resulted maximum V-M stress (obtained from the sleeper FE model) all increase with the rise of rolling speed. Furthermore, the difference between the maximum and the minimum pressures grows from 2.55 MPa under the rolling speed of 40 km/h to 3.37 MPa for the rolling speed of 140 km/h.
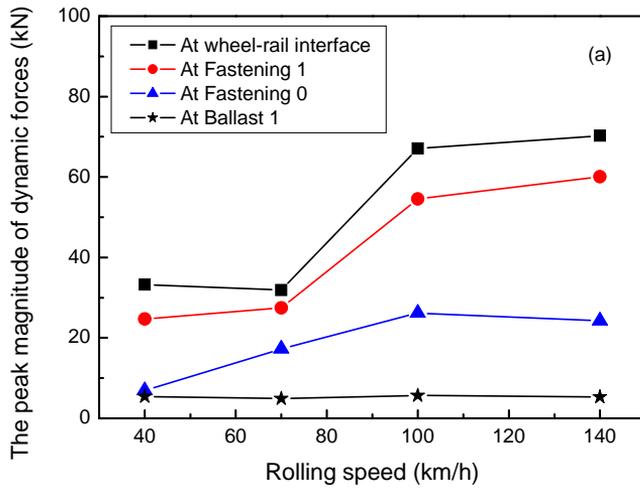
(a)
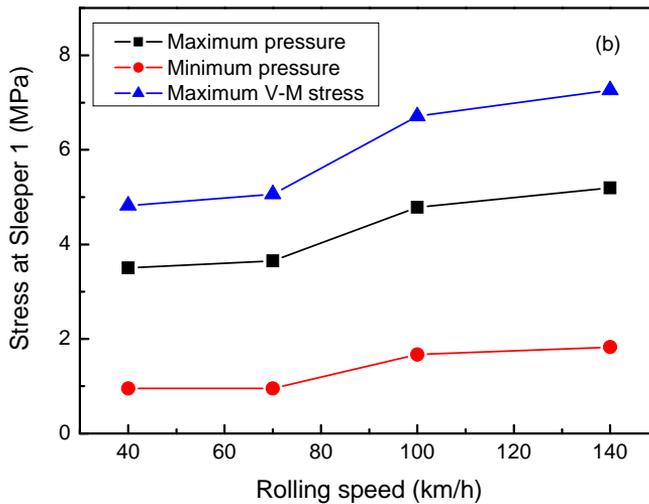
(b)

Figure 4.3.15. The influence of rolling speed when Defect M1 is present. (a) The peak magnitude of the dynamic forces, and (b) The stress level in Sleeper 1 at Instant T_F.

4.3.3.3 Defect M0

Figure 4.3.16 shows the dynamic forces at the wheel–rail interface and at Support 1 excited by Defect M0 (see Figure 4.3.3) under the default rolling speed. Comparing with the results of Defect M1 shown in Figure 4.3.7, it is observed that the peak magnitude caused by Defect M0 is clearly lower than the one by Defect M1. From the amplitude spectrums, it is further found that the vibration components at f_1 and f_2 both reduce significantly in magnitude from Defect M1 to M0. These phenomena should mainly be

attributed to two factors: one is the depth reduction from Defect M1 to M0, and the other is the variation of the excitation frequency (f_{ex}) defined by Equation (4.3.6).

$$f_{ex} = v/L_D \qquad\qquad (4.3.6)$$

where, v is the rolling speed and L_D is the characteristic wavelength of the simulated defect.

(a)

(b)

Figure 4.3.16. The dynamic forces at the wheel–rail interface and Support 1 caused by Defect M0 under the default rolling speed. (a) The force variation in the time domain, and (b) the amplitude spectrum.

For Defect M0 L_D is about 80 mm because for the simulated speed its middle part around $x = 40$ mm is not hit by the passing wheel, whereas it reduces into about 40 mm

for Defect M1 due to its two-part characteristics in geometry, as shown in Figure 4.3.3. Substituting these values into Equation (4.3.6), it can be calculated that f_{ex} is about 350 and 700 Hz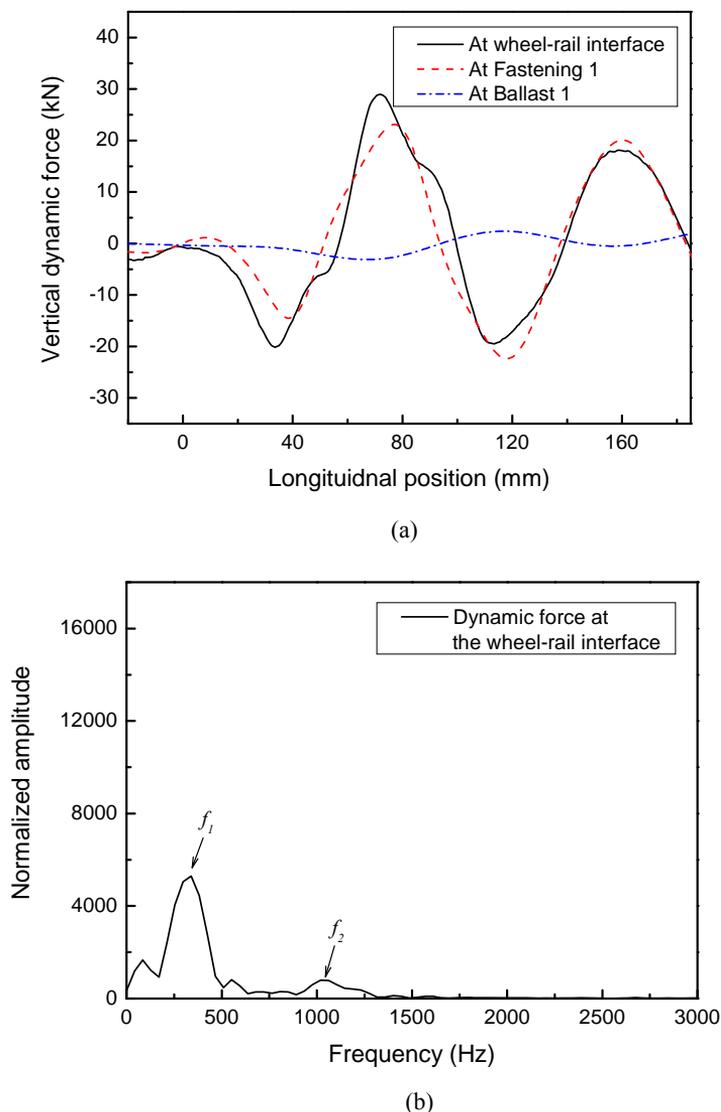 for Defects M0 and M1, respectively, under the default rolling speed of 100 km/h. Clearly, the excitation frequency caused by Defect M1 is closer to f_2 than the one by Defect M0. It can, therefore, be expected that Defect M1 can more efficiently excite the vibration at f_2, as shown in Figures 4.3.7 and 4.3.16. Such a mechanism may also be employed to explain the absence of the vibration at f_2 under the speed of 40 km/h as show in Figure 4.3.12. It should be noted that the excitation frequency defined in Equation (4.3.6) is only used for easy understanding of the influence of defect geometry, and it may not be an accurate expression for the excitation at the simulated SRSD due to the existence of material elasticity and the flexibility of the system.

The dynamic forces excited by Defect M0 at different fastenings and ballasts are plotted in Figure 4.3.17 versus the position of the wheel. In order to explain their differences from the ones shown in Figure 4.3.8, the influences of the defect geometry (see this section) and the transmission of the vibration components in the track structure (see Table 4.3.2) should be both considered. Detailed explanations about this are omitted here to avoid repetition.

4.3.4 Discussions

In order to study the vibration excited by a SRSD more conveniently, its two main vibration components (at f_1 and f_2) are separated from each other by FFT. More simulations carried out show that f_2 does not vary with the size of the defect. Therefore, based on the agreement between the calculated dynamic force variation and the geometric deviation of the simulated squat, it can be concluded that the employed FE model can capture the force variation causing squats [4.3.1]. The calculated dynamic force does not just simply follow the geometric deviation at the simulated defect. Careful examinations in the literature have shown that these squats related force variations and their transmissions in the track structure cannot be captured by the mass–spring–beam models, e.g. in [4.3.2, 4.3.9–4.3.11, 4.3.13]. This confirms the necessity of employing the 3D transient FE model explained above for the investigations on SRSD. Note that further work is required to find out the origins of f_1 and f_2 which might be related to some resonances of the coupled vehicle–track system.

It is mentioned above that the decay rate of the vibration at f_1 considerably increases when higher damping is applied into the fastenings and ballast of the system. Thus, when a track is considered in the simulation, parameters about the system damping in the investigated frequency range should be carefully determined in order to accurately predict the decay rate of the vibrations. In addition, it is expected that f_1 and f_2 vary with the different vehicle–track system.

Considering the fact that the vibrations at f_1 and f_2 are transferred into the rail supports in different ways (see Table 4.3.2), the dynamic wheel–rail contact force may not be directly measured by recording the dynamic forces at the fastenings (such as at the base-plate [4.3.2, 4.3.29]), especially when the vibration at f_2 is relatively large in magnitude compared to at f_1. For instance, the peak magnitude of the dynamic forces at Fastening 1 (see Figure 4.3.13(b)) is 15–25% less than the one at the wheel–rail interface

(in Figure 4.3.12(a)) under the simulated speeds. However, summation of the peak magnitudes at different fastenings (in Figure 4.3.13) is obviously larger than the peak magnitude of the dynamic contact force between the wheel and the rail (in Figure 4.3.12) due to the phase shift.

(a)

(b)
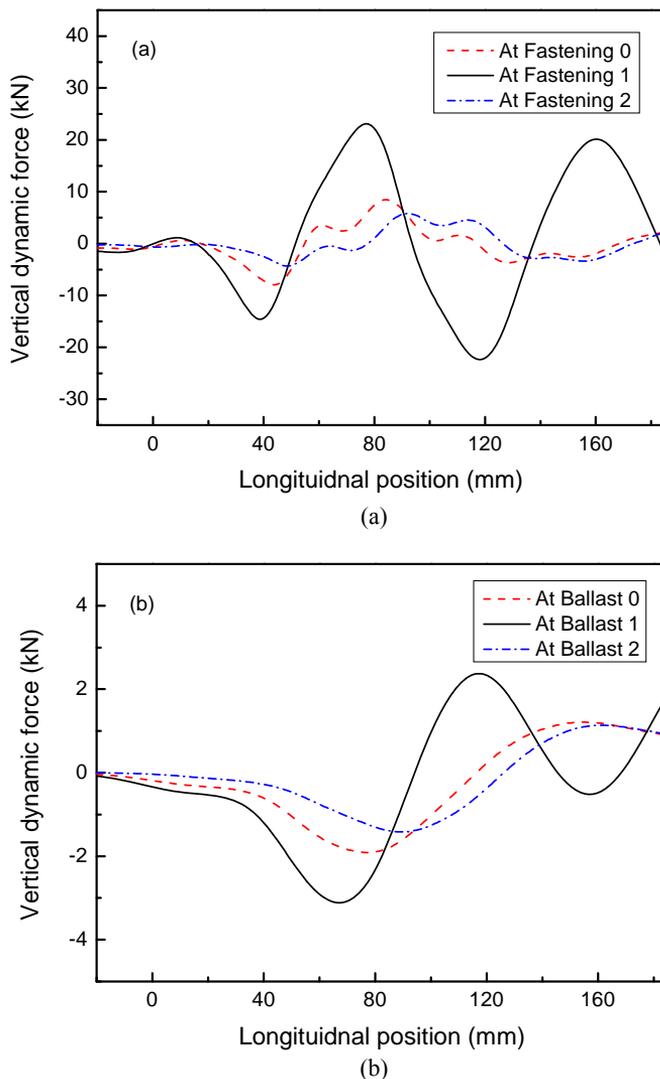
Figure 4.3.17. The dynamic force variations caused by Defect M0 under the default rolling speed. (a) At different fastenings, and (b) at different ballasts.

With the 3D FE modelling described above, the SRSD-excited dynamic forces at both the wheel–rail interface and at discrete supports can be evaluated in the time domain. Hence, based on such a model, a possible method may be developed in the future to optimize the vehicle–track system with the intention of minimizing the dynamic forces caused by SRSD.

This paper also shows that the Zimmermann solution, which is based on a continuous support of rail and has been applied in engineering to calculate the vertical forces at discrete supports, is indeed valid for a quasi-static wheel load. An error of about 10% exists compared to the FE solution of smooth rail. For high frequency contact loads, however, the Zimmermann approach is not applicable due to its basis of statics.

It should be mentioned that the current available measurements at SRSD for authors are only in two dimensions. Some assumptions, therefore, have to be applied in order to extend the measurements into 3D geometry, as explained in Section 4.3.2.2. In the future, detailed information about the 3D geometry is required for more accurate predictions. For certain cases such as with considerable plastic deformation, the tangential interaction between wheel and rail may have to be considered for different friction conditions [4.3.30].

Finally, the V-M stress level at sleepers is calculated based on the pressure results exported from the vehicle–track simulations. However, the V-M yield criterion that applies best to ductile materials may not be suitable for concrete material. Hence, the stress at sleepers may have to be evaluated in another way for damage of concrete sleepers in the further work.

4.3.5   Conclusions

A 3D transient FE model of the vehicle–track interaction system is employed to calculate the dynamic forces excited by SRSD. Based on the obtained results, the following conclusions can be drawn:

(1)   Detailed modelling such as the 3D transient FE model employed in this paper is necessary for the investigations on SRSD. A mass–spring–beam model cannot capture the squats related vibrations and their transmissions in the track structure.
(2)   The Zimmermann solution is suitable to calculate in statics the vertical force distribution at discrete supports, with an error of about 10% at the most heavily loaded support.
(3)   The dynamic force at the wheel–rail interface excited by the simulated SRSD mainly oscillates at two different frequencies. The one with the lower frequency can transfer down to the ballast layer, especially at the support close to the SRSD, whereas the other one has similar magnitude at different fastenings and almost disappears at the ballast layer.
(4)   A mature SRSD mainly increases the vertical forces at the two supports of the sleeper bay where the defect exists. Phase shift occurs between the dynamic forces at the wheel–rail interface, the fastenings and the ballasts due to the transmission process of the vibration.
(5)   At the wheel–rail interface, the magnitude and wavelength of the two main vibration components excited by SRSD both decrease with the reduction of the rolling speed. The vibration at the higher frequency can be absent when it is clearly higher than the excitation frequency.
(6)   For SRSD, the dynamic forces recorded at the fastenings (such as at the base-plate) does not represent the dynamic wheel–rail contact force. Taking the squat simulated

in this work as an example, the peak magnitude at the most heavily loaded fastening is 15–25% less than the one at the wheel–rail interface for the speed of 40–140 km/h.

(7) Because of the detailed modeling of the rail and the sleepers, the non-uniform pressure distribution at the rail seat area of a sleeper is captured by the 3D transient FE model.

(8) The responses of the vehicle–track system greatly vary with geometry of SRSD.

References

[4.3.1] Z. Li, X. Zhao, C. Esveld, R. Dollevoet, and M. Molodova, An investigation into the causes of squats: correlation analysis and numerical modelling. Wear 265 (2008), 1349–1355.

[4.3.2] H.H. Jenkins, J.E. Stephenson, G.A. Clayton, G.W. Morland, and D. Lyon, The effect of track and vehicle parameters on wheel/rail vertical dynamic forces. Railway Engineering Journal 3 (1974), 2–16.

[4.3.3] K.L. Knothe, and S.L. Grassie, Modelling of railway track and vehicle/track interaction at high frequencies. Vehicle System Dynamics 22 (1993), 209–262.

[4.3.4] S.L. Grassie, Models of railway track and vehicle–track interaction at high frequencies: results of benchmark test. Vehicle system dynamics 25(suppl.), (1996), 243–262.

[4.3.5] J.C. Zeman, J.R. Edwards, C.P.L. Barkan, and D.A. Lange, Failure mode and effect analysis of concrete ties in North America. In Ninth International Heavy Haul Conference, Shanghai, China, June 2009, 270–277.

[4.3.6] J. Leong, M. Murray, and D. Steffes, Examination of railway track dynamic models capabilities against measured field data. In Eighth International Heavy Haul Conference, Kiruna, Sweden, June 2007, 257–267.

[4.3.7] K. Knothe, and A. Groß-thebing, Short wavelength rail corrugation and non-steady-state contact mechanics. Vehicle system dynamics 46 (2008), 49–66.

[4.3.8] K. Hempelmann, and K.L. Knothe, An extended linear model for prediction of short pitch corrugation. Wear 191 (1996), 161–169.

[4.3.9] C. Andersson, and T. Dahlberg, Wheel/rail impacts at railway turnout crossing. Proc. IMechE, Part F: J. Rail and Rapid Transit 212 (1998), 123–134.

[4.3.10] W.M. Zhai, C.B. Cai, Q.C. Wang, Z.W. Lu, and X.S. Wu, Dynamic effects of vehicle on tracks in the case of raising train speeds. Proc. IMechE, Part F: J. Rail and Rapid Transit 215 (2001), 125–135.

[4.3.11] T.X. Wu, and D.J. Thompson, The effects of track non-linearity on wheel/rail impact. Proc. IMechE, Part F: J. Rail and Rapid Transit 218 (2004), 1–15.

[4.3.12] M.X.D. Li, E.G. Berggren, and M. Berg, Assessment of vertical track geometry quality based on simulations of dynamic track-vehicle interaction. Proc. IMechE, Part F: J. Rail and Rapid Transit 223 (2009), 131–139.

[4.3.13] M.J.M.M. Steenbergen, Modeling of wheels and rail discontinuities in dynamic wheel–rail contact analysis. Vehicle System Dynamics 44 (2006), 763–787.

[4.3.14] N. Chaar, and M. Berg, Simulation of vehicle–track interaction with flexible wheelsets, moving track models and field tests. Vehicle System Dynamics 44 (2006), 921–931.

[4.3.15] L. Baeza, J. Fayos, A. Roda, and R. Insa, High frequency railway vehicle–track dynamics through flexible rotating wheelsets, Vehicle System Dynamics 46 (2008), 647-662.

[4.3.16]   B. Ripke, and K. Knothe, Simulation of high frequency vehicle–track interactions, Vehicle System Dynamics 24 (1995), 72-85.

[4.3.17]   Z. Wen, X. Jin, and W. Zhang, Contact-impact stress analysis of rail joint region using the dynamic finite element method, Wear 258 (2005), 1301–1309.

[4.3.18]   T. Pang,, and M. Dhanasekar, Dynamic finite element analysis of the wheel–rail interaction adjacent to the insulated joints, Proceedings of 7th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems (CM2006), Brisbane, Australia, September 24–26, 2006, 509–516.

[4.3.19]   P.J. Remington, Wheel/rail noise — Part IV: rolling noise. Journal of Sound and Vibration 46 (1976), 419–436.

[4.3.20]   D.J. Thompson, The influence of the contact zone on the excitation of wheel/rail noise. Journal of Sound and Vibration 267 (2003), 523–535.

[4.3.21]   A. Afshari, and A.A. Shabana, Directions of the tangential creep forces in railroad vehicle dynamics. Journal of Computational and Nonlinear Dynamics, 5 (2010), 021006.

[4.3.22]   S. Iwnicki, Y. Bezin, G. Xie, and E. Kassa, Advances in vehicle–track interaction tools, Railway Gazette International 165 (2009), 47–52.

[4.3.23]   Z. Li, R. Dollevoet, M. Molodova, and X. Zhao, Squat growth–some observations and the validation of numerical predictions, Wear 271 (2011), 148–157.

[4.3.24]   M. Molodova, Z. Li, and R. Dollevoet, Axle box acceleration: measurement and simulation for detection of short track defects, Wear, 271 (2011), 349–356.

[4.3.25]   X. Zhao, and Z. Li, The solution of frictional wheel–rail rolling contact with a 3D transient finite element model: Validation and error analysis, Wear 271 (2011), 444–452.

[4.3.26]   M. Molodova, Z. Li, and R. Dollevoet, Numerical simulation of the axle box acceleration of a railway wheel rolling over a rail geometrical irregularity, Proceedings of the 7th International Conference on Railway Bogies and Running Gears, Budapest, Hungary, 3-6 September, 2007.

[4.3.27]   M. Hiensch, J.C.O. Nielsen, and E. Verheijen, Rail corrugation in The Netherlands —measurements and simulations, Wear 253 (2002), 140–149.

[4.3.28]   D.J. Benson, and J.O. Hallquist, A single surface contact algorithm for the post-buckling analysis of shell structures, Computer Methods in Applied Mechanics and Engineering 78 (1990), 141–163.

[4.3.29]   F. Muller-Boruttau, N. Breitsamter, and S. Pieper, Measuring the effects that flat spots have on the dynamic wheel load and on rail pad forces, European rail Technology Review 49 (2009), 24–29.

[4.3.30]   O. Arias-Cuevas, Z. Li, R. Lewis, and E.A. Gallardo-Hernández, Rolling–sliding laboratory tests of friction modifiers in dry and wet wheel–rail contacts, Wear 268 (2010), 543–551.

Summary and recommendations

Simulations of dynamic vehicle–track interaction at squat type rail surface defects have shown that the vehicle–track system is mainly excited by a squat at two frequencies. For the investigated rolling speed range between 40 and 140 km/h, the two vibration components both increase in magnitude and wavelength with the rolling speed. When the rolling speed is sufficiently low, e.g. at 40 km/h for the simulated system, the vibration component at the higher frequency can be absent.

A growth process of squats has been postulated based on the squat excited dynamic contact forces (obtained from the FE model), and has been validated by track monitoring conducted in the Netherlands. In addition, correlations between squat occurrence and short wavelength irregularities such as discrete supports of rail, corrugation, and welds have been identified.

Studies on the dynamic forces at discrete supports of rail have shown that the squat related vibrations transfer downward in different ways. The vibration with the lower frequency can transfer down to the ballast layer, especially at the support close to the squat, while the other one has similar magnitude at several fastenings near the squat and is negligible at the ballast layer. Consequently, a mature squat mainly increases the forces at the two supports of the sleeper bay where the defect exists.

5 A minimum critical size for squat to initiate

In this chapter, the transient wheel–rail rolling contact at rail surface defects is examined together with the resulted dynamic stress in the rail head. Squat initiation from rail surface defects is further analyzed, for which the tensile strength of rail steel is employed as the fatigue criterion.

Chapter 1: Introduction to squats and their consequences.	Chapter 2: Literature review and outline of this dissertation.

Chapter 3: Solution of frictional rolling contact.

Chapter 4: Squat initiation, growth and wheel–track dynamic forces.

Chapter 5: A minimum critical size for squats to initiate.

Chapter 6: Conclusions and recommendations.
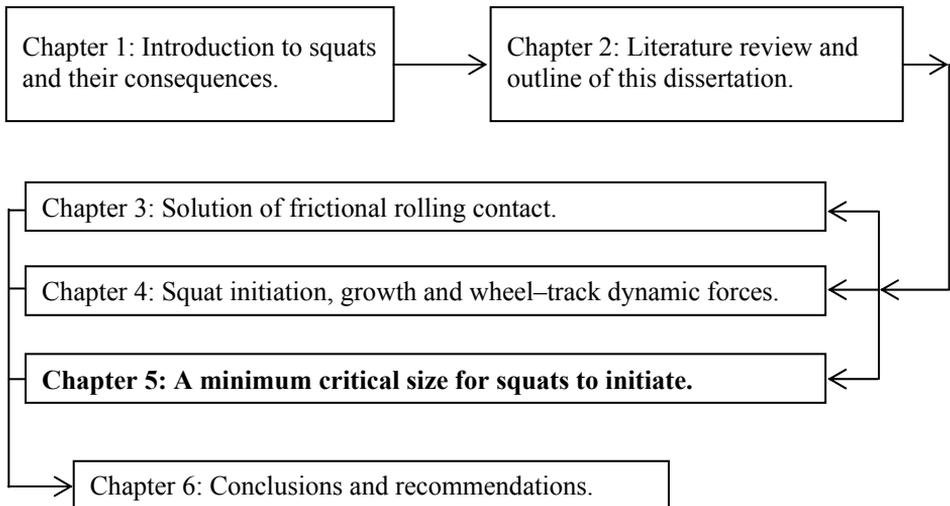
5.1   The determination of a critical size for rail top surface defects to grow into squats[6]

ABSTRACT

Squats as a type of rolling contact fatigue initiate and grow from small rail top geometry defects. But not all such small defects grow into squats. This paper presents a methodology, which is based on material strength and stress evaluation of rolling contact at the defects, for the determination of a critical size for such small defects to grow into squats. To that end a detailed 3Dimensional frictional rolling contact model is integrated in the wheel–track interaction system, and the critical size is determined by comparing the maximal von Mises stress at defects of different sizes with the tensile strength of the material. The methodology is demonstrated by applying it to typical Dutch operational conditions, and the numerically determined critical size is verified with field monitoring. The critical size can be used for distinguishing between light squats and trivial defects, by visual inspection or by automatic image recognition, so that false statistics of squats can be reduced or prevented.

---

[6] This work has been published as: Z. Li, X. Zhao and R. Dollevoet, The determination of a critical size for rail top surface defects to grow into squats, proceedings of 8th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems (CM2009), Firenze, Italy, September 15–18, 2009, 379–388.

5.1.1   Introduction

The rail is designed to provide a smooth surface for the wheels to run on it in order to reduce the rolling resistance. However, many irregularities gradually occur and grow on the running surface of the rail under certain conditions during service. Squats, which are a visual depression in the middle of the running band, are among these irregularities. They have become one of the main concerns in many railway networks with the changes in maintenance practice and with the continuous increase in speed, axle load, etc. In the Netherlands, squats have nowadays become the number one concern for rail damage, since head checks are under control.

Squats usually occur on top of rails in the running band on straight track and shallow curves. Squats have been found on all types of track: ballasted or slab track, with wooden or concrete sleepers, with passenger, freight or mixed traffic. They occur not only on high speed lines, but also on conventional and metro lines. A fundamental characteristic of squats is large local plastic deformation and the absence of shakedown. Many of them are related to rail top surface vertical irregularities such as indentations and wheel burns.

There have been mainly three lines of research on squats: from the point of view of metallurgy, from the point of view of crack growth, and from the point of view of squats initiation, prevention and detection.

Clayton et al. [5.1.1] reviewed a metallurgical research program at British Rail on surface initiated rail problems, with squats among them. Longitudinal-vertical sectioning of rail specimens showed surface initiated cracks that could branch downwards, and when they reached a critical size, brittle fracture could occur and it could result in a broken rail. Bogdański and his colleagues have modeled squat cracks under various conditions. The latest results [5.1.2] treated the effect of entrapped liquid by numerical modeling.

Since 2006, some investigations have been concentrated on the mechanism of initiation and growth of squats, as well as their root causes. The goals of the investigations are early detection and prevention of squats [5.1.3, 5.1.4]. The approach has consisted of correlation and numerical analyses, and their validation by field monitoring. The correlation analyses related squat occurrence to certain parameters in the vehicle–track interaction system and to observations of phenomena around squats in the tracks [5.1.4]. They helped eliminate many of the less influential parameters and simplify effectively the subsequent numerical modeling. The numerical analyses were employed to quantify the relation between the influential parameters identified in the correlation analysis and the dynamic rolling contact forces, stresses and strains. A few of the most influential parameters were further identified. The monitoring of the squats evolution on the tracks has provided realistic inputs to the analyses, and has provided data for validation.

Based on field observations and numerical analysis Li et al. [5.1.3] found that squats may initiate as a result of differential wear and differential plastic deformation.

Such wear and deformation may take place at locations where short track defects excite dynamic peak forces of the wavelength that promotes squat initiation and growth.

Li *et at* [5.1.5] showed that track short wave irregularities, especially those on the rail top surface such as corrugation, indentations and welds of poor quality, were related to squat occurrence. Based on numerical simulations, a squat growth process was postulated. The influence of friction on squat initiation and growth was predicted. The simulations also revealed that the growth of squats was related to some eigen wavelength of the wheel–track interaction system and that high frequency vibration at wheel–rail contact played an important role. The validation of these numerical results is presented in [5.1.5].

In nature, squats are visually a rail top geometry deviation due to large plastic deformation. They must, therefore, grow from some small rail top defects. These can be:

- Indentations by hard alien objects between wheel and rail (see Figure 5.1.1). The objects can be, for instance, a hard ball from an aerosol paint can (Figure 5.1.1(a)), or a ball or roller from a bearing. They can indent into the wheel and be brought forward. Such indentations can be recognized by their periodicity of the wheel circumference.
- Wheel burns. They are usually found at or near stations and traffic signals. Figure 5.1.2 shows such an example.
- Welds of poor quality of continuously welded rail.
- Defects caused by differential wear and differential plastic deformation.
- Short pitch corrugation.
- Vertical misalignments of the rails at switches and crossings (S&C).
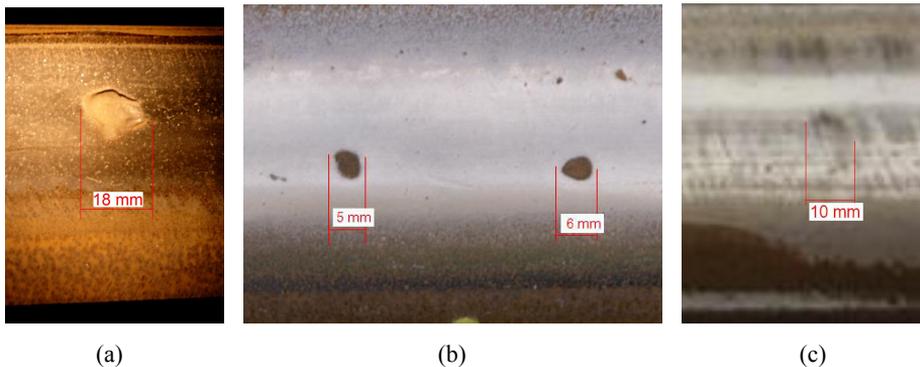

(a) (b) (c)

Figure 5.1.1. Various defects in rail head: (a) an indentation caused by a steel ball from an aerosol paint can, (b) an indentation by unknown object, and (c) a defect of unknown cause.

Figure 5.1.2. A squat developed from a wheel burn.

Among these defects, Li et al. [5.1.4, 5.1.5] show that short pitch corrugation has been responsible for about 33% of the squats, and welds responsible for 10–15% in the Netherlands. The other 52–57 % must be due to other rail top defects, mainly due to indentations, and to less extent due to wheel burns, differential wear and differential plastic deformation. The wheel–rail interaction at S&C is more complicated. The statistics of squats at S&C is not available yet; it is therefore not included in the above statistics.

5.1.2   Motivation of the work

From the viewpoint of rail infrastructure management, prevention and early correction based on timely detection of light squats is more cost effective, and enhances the safety and availability of the network.

According to the squat growth process postulated in [5.1.4] and validated in [5.1.5], a defect causes a dynamic force, which may result in plastic deformation and material hardening. Plastic deformation usually makes the contact geometry more conformal so that the stress will be reduced, if the contact force remains the same. And the hardening will increase the yield stress as long as the tensile strength of the material is not yet reached. Consequently, the stress may be lower at subsequent wheel passages than the increased yield stress, and the material will reach a shakedown state, if the defect is not large enough. The defect will, therefore, not grow into a squat, and it may eventually disappear because of wear. This means that there should be a critical size, only above which a defect can grow into a squat. In other words, a defect can be considered as a squat only when it is larger than the critical size.

Such a critical size has the following significance. Firstly, it can be used as a criterion to distinguish between light squats and trivial defects which may eventually be erased by wear. Currently in the Netherlands squats are classified by visual inspection as being light, moderate and severe [5.1.6]. There is not yet a quantitative criterion to

determine which defect is a light squat, and which is not. Therefore, false statistics and reporting of light squats may occur. Secondly, it can be employed for automated detection and classification of light squats. So far there is not yet an automatic measurement that can effectively detect light squats. The railway infra managers such as ProRail are still relying on visual inspection by human beings along the track, which is inefficient and unsafe. If the critical size is known, it is possible to detect light squats with, for instance, automated image recognition. The light squats may also be automatically detected with instrumented wheelset by establishing a quantitative relation between the size of the defects and the responses of the wheelset to them.

Nevertheless, such a size has not been found yet. This paper presents a methodology for the determination of such a critical size, demonstrated with its application to typical operation conditions of the Dutch railway. The methodology is generally applicable to other operation conditions.

5.1.3   The numerical model

Since squats grow due to plastic deformation, we evaluate the maximal von Mises stress due to rail top geometry defects of various sizes and compare it with the yield stress of the material to assess the growth tendency of the defects. To that end a transient Finite Element (FE) model is employed. The contact stress at the defects is solved for in the vehicle–track interaction system with wheel–rail frictional rolling contact being solved in three dimensions. The integration of the detailed contact modeling in the vehicle–track system guarantees that the mutual influence of the wheel–track system and the wheel–rail contact is fully taken into account. This is important for high frequency dynamic interaction such as in the current case of wheel–rail contact at squats. The model has been employed for previous works [5.1.3, 5.1.4]. In [5.1.7] its validity in the solution of frictional rolling contact in elasticity is established.

The FE model is schematically given in Figure 5.1.3. A half wheelset of a typical Dutch passenger car of the type VIRM and a half straight track are modeled. In the Netherlands a variety of wheel profiles are used. Because squats occur on the rail top, the rail is in contact with the middle part of the wheel tread. An equivalent conicity of 1/40 is taken as the average for all the new and worn wheels. The rail profile is the 54E1 with 1:40 inclination. In the place where the defects are located the elements are 1.11 mm long in the rolling direction. Vehicle primary suspension, rail pad, sleepers and ballast are all taken into account. Comparing to the model in [5.1.4], the sleepers are modeled using solid elements instead of mass elements.

A right-handed Cartesian coordinate system is defined as shown in Figure 5.1.3. Its origin is on the rail surface at the symmetry centre of the rail profile. In the longitudinal (rolling) direction, the origin is at point a in Figure 5.1.3, the leading edge of the defect, see also Figure 5.1.5(a) below. The z axis is normal to the rail surface.
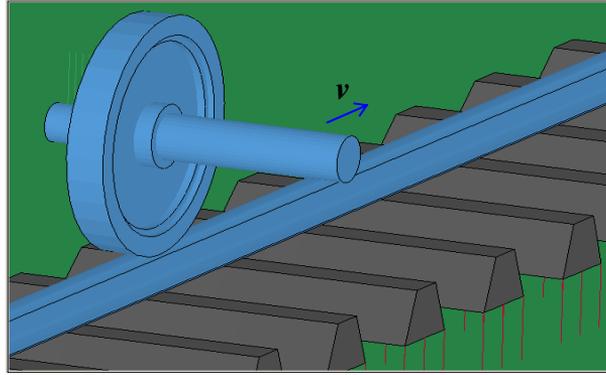
(a)
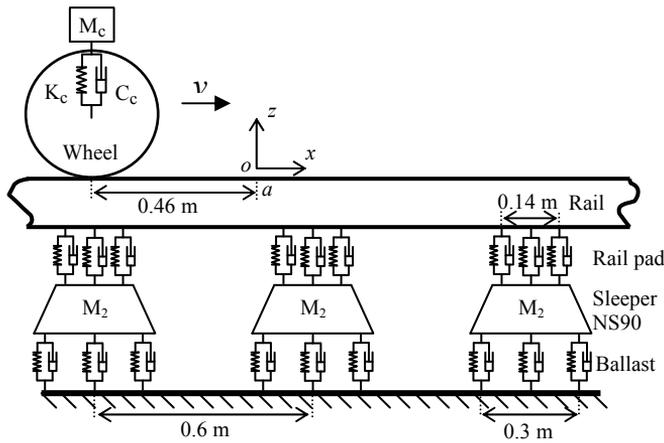
(b)

Figure 5.1.3. The FE vehicle–track interaction model. (a) 3D view and (b) longitudinal-vertical view.

The values of the parameters of the model are listed in Table 5.1.1. To consider the worst scenario, the static vertical contact force N is the half wheel load increased by 30% to take into account the long wave dynamic force of the vehicle. The un-sprung mass of the half wheelset is distributed as it is in reality because the real geometry of the wheelset is meshed. The sprung mass is lumped and supported by the primary suspension. Elastic material is used for both the wheel and rail to calculate the maximal von Mises stress, so that it can be compared with the yield stress to assess the tendency for plastic deformation. The Young's modulus is $E = 210$ GPa, and Poisson's ratio is 0.3.

It is noticed that the exploitation of available coefficient of friction (COF, f) is limited by the traction and braking capability of the rolling stock. For instance the design value of the maximal traction coefficient is 0.22 for the VIRM, whereas the design friction exploitation for braking is usually lower than for traction. Here we distinguish between COF f and the traction coefficient μ,

$$\mu = F_L / N \tag{5.1.1}$$

Where N is the vertical contact load and F_L is the longitudinal friction force, $F_L \leq fN$. No lateral force is considered. Considering the variation in the contact force during rolling due to vibration, the traction coefficient used in this paper are nominal values calculated with the static load.

Table 5.1.1. The values of the parameters in the model.

Parameters		Values
Wheel diameter		0.92 m
Static wheel load		116.8 kN
Speed of the wheel		140 km/h
Primary suspension	Stiffness	880 kN/m
	Damping	4000 Ns/m
Railpad	Stiffness	1300 MN/m
	Damping	45000 Ns/m
Ballast	Stiffness	45 MN/m
	Damping	32000 Ns/m
Wheel &rail	Young's modulus	210 GPa
	Poisson's ratio	0.3
	Density	7800 kg/m^3
Sleeper (NS 90)	Young's modulus	38.4 GPa
	Poisson's ratio	0.2
	Density	2520 kg/m^3
	Mass	280 kg

5.1.4 The derivation of a critical size

The longitudinal-vertical rail profiles can be measured with the device RailProf at the centre of the 54E1 profile. Figure 5.1.4 shows measurements of two light squats. Note that the interval of the measurement data point is 5 mm; the curves are therefore not very smooth.

5.1.4.1   Maximal V-M stress of Defects with Sharp Edge

a) The defect models

A series of idealized defects with increasing size are applied to the rail head of the FE model to investigate their maximal V-M stress. The shape of the defects is generalized from the observations of many measurements at small defects and light squats, such as those shown in Figure 5.1.4.
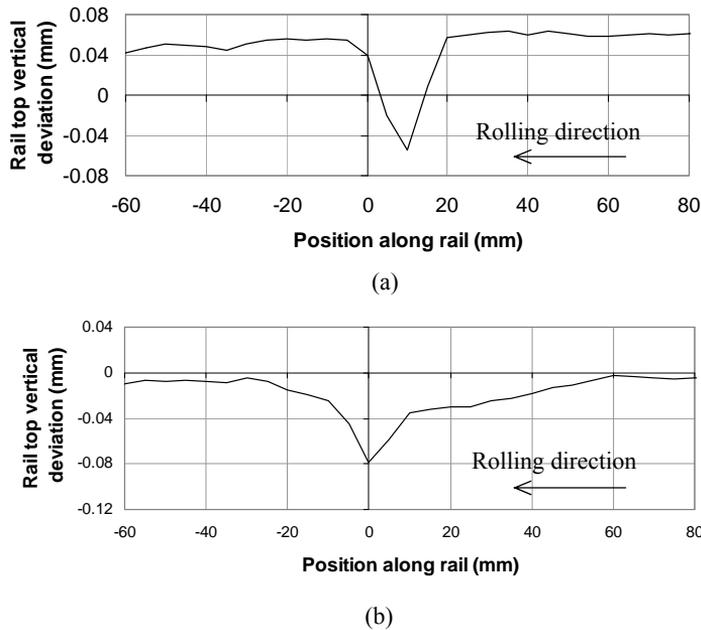
(a)

(b)

Figure 5.1.4. Rail top longitudinal-vertical profiles measured along the centre of rail head at two small defects. The measurement data point interval is 5mm. The curves are not smoothed.

When an indentation is formed its edge is usually sharp in the beginning. Defects with sharp edge are, therefore, first investigated. Figure 5.1.5 shows the longitudinal-vertical and lateral-vertical profiles of the simulated defects at their deepest cross-sections. Their appearance in 3 dimensions is given for Defect_10 in Figure 5.1.6. The naming convention of the defects is such that Defect_4.4 represents a defect of 4.4 mm long in the rolling direction, see Figure 5.1.5(a).

To avoid the influence of the defect geometry, the profiles of the defects are kept as similar with each other as the FE meshing allows. Particularly, the shapes of the defects are kept the same at the trailing edges (i.e. the part of the defect which is the last to come into contact with the wheel, namely the edge on the right side in Figure 5.1.5(a)). The depth of Defect_10 is first determined based on the fact that the dimension of a typical wheel–rail contact is in the order of 10 mm, and the corresponding typical compression of the rail is in the order of 0.05 mm. Field observations show that wheels have usually no or very little contact with the bottom of small defects, see the black rusty bottom of the defects in Figures 5.1.1 and 5.1.2. The depth of Defect_10 is therefore chosen to be 0.06mm. The depth of the other defects is determined by keeping them geometrically similar to Defect_10.

The generalized defects are applied to the rail with abscissa 0.0 mm of Figure 5.1.5(a) being at the location *a* in Figure 5.1.3(b), the origin of the coordinate system. This location is chosen because statistics [5.1.4] showed that it was in this region of a sleeper span where about three quarter of the squats occurred, and because simulations show that the dynamic force is the highest there, if other conditions are the same.
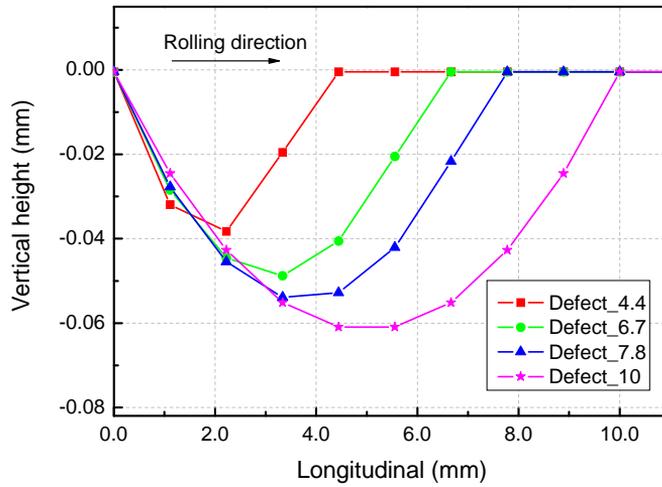
(a)
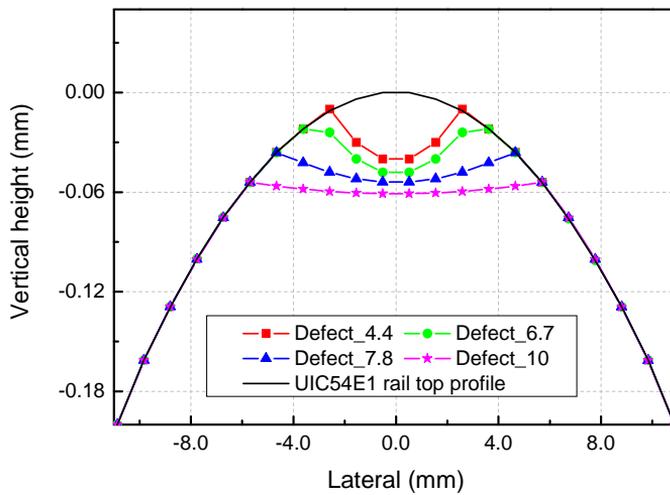
(b)

Figure 5.1.5. The shape and dimension of the generalized defects. (a) Profiles at the deepest longitudinal-vertical cross-section and (b) profiles at the deepest lateral-vertical cross-section.
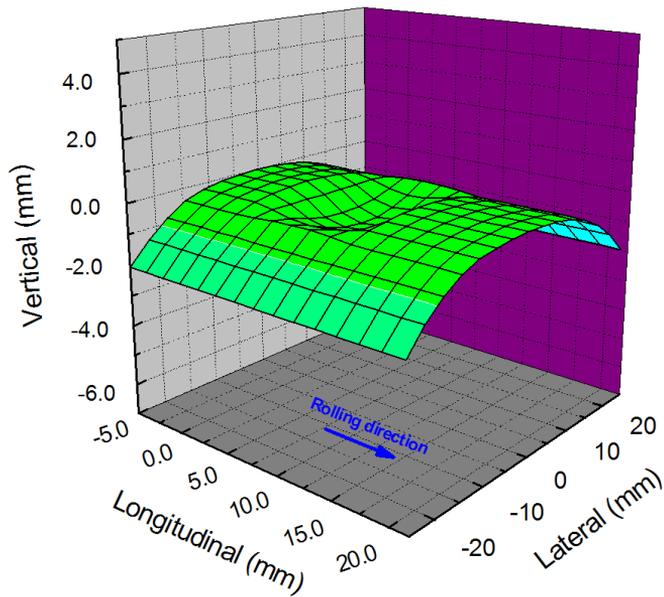
(a)
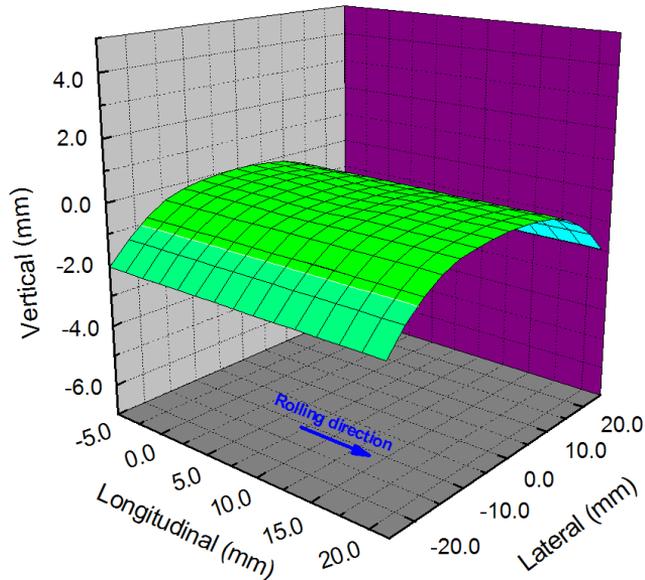
(b)

Figure 5.1.6. The 3D appearance of Defect_10 when it is applied to the rail head. (a) Defect with 10x magnification in vertical direction and (b) defect without magnification. Note in (a) the magnification is applied only to the defect, not to the rest of the rail top surface, the scale of the ordinate remains therefore unchanged.

b) The transient contact at defect

Still taking Defect_10 as an example, the pressure and surface shear stress distributions along the longitudinal axis of the contact patch are plotted for different instants (different time steps) in Figure 5.1.7, from which the transition process of the contact patch at the defect can be observed. Apparently, peaks of both the pressure and the surface shear stress occur at the leading and trailing edges of the defect due to the effect of stress concentration. Comparing in Figure 5.1.7(a) the sharp peaks of the pressure at steps 642 and 654 when contact is at the defect's edges, with the maximal pressure of step 665 when the wheel has passed the defect and is in contact with the smooth rail head, it can be seen that the stress concentration has significantly increase the maximal pressure. The same is true for the shear stress in Figure 5.1.7(b), where the maximal values at the defect's edges are at steps 651 and 663, while the result of the smooth contact is at step 667. As a result, the V-M stress at the defect is much higher than on the smooth surface.

It is noted that in Figure 5.1.7 the pressure and the shear stress are not given for the same time steps. The reason is that the pressure and the surface shear stress reach their respective maximums at different instants. To explain, the elliptic contact of two smooth bodies is first looked at: When the friction is not fully exploited the pressure and surface shear stress may reaches their maximums at different locations in the contact patch. The maximum pressure occurs always at the center of the contact patch, whereas the maximum surface shear stress occurs on the border between the areas of adhesion and slip. On the other hand in the presence of a defect both the maximal pressure and the shear stress may reach their respective maximums at the same location, i.e. the defect's edges, so that the instants when the respective maximums reach the edges will be different.

Figure 5.1.8 shows stress distributions when contact is at the trailing edge. As can be seen the defect geometry has greatly changed the pressure distribution from the ellipsoid shape that contact between smooth wheel and smooth rail should have to that of Figure 5.1.8(a). Due to the dependency of the surface shear stress on the pressure, the shear stress distribution is also changed accordingly.

One may observe in Figure 5.1.7(b) that the largest shear stress is at the leading edge ($x=0.0$), instead of being at the trailing one ($x=10.0$). This is not necessarily true, and should be caused by the large time interval for outputting data for the plotting: It is too costly to output results of every time step for the plotting, and therefore for plotting purpose results were output at pre-selected instants only, so that an output instant may not be precisely at the time of the maximum, but closely next to it. It is emphasized that this error is only for this plotting. For the analyses and for the determination of the critical size the true calculated maximums are used.

Figure 5.1.9 shows as an example the contact patch and its division into the areas of slip and adhesion at step 654. Compared to wheel–rail rolling contact without surface defect, the contact patch, and the areas of slip and adhesion are all changed by the presence of the defects. Although symmetric defect is applied to the symmetric rail head, the contact patch is not completely symmetric about the $y = 0$ axis. This is owing to the rail inclination, which causes the rail to bend toward the gage side under the vertical

wheel load, so that the contact geometry is not symmetric any more. Detailed analysis of the transient rolling across the defect is beyond the scope of this paper.
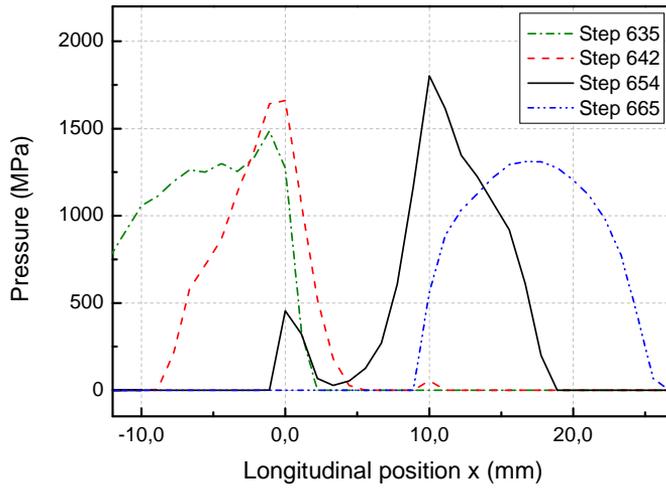
(a)
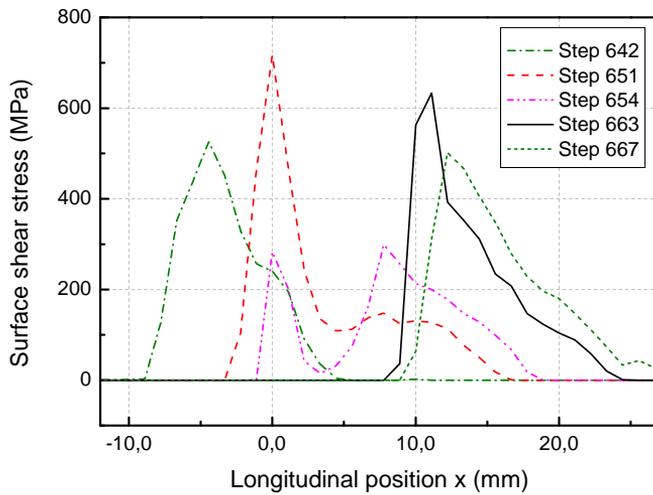
(b)

Figure 5.1.7. The transition process along the longitudinal axis of the contact patch (y=0) at Defect_10. (a) The pressure and (b) the surface shear stress.

151

(a)

(b)

Figure 5.1.8. The 3D stress distribution at Defect_10. (a) The pressure distribution at step 654 and (b) the surface shear stress distribution at step 663.
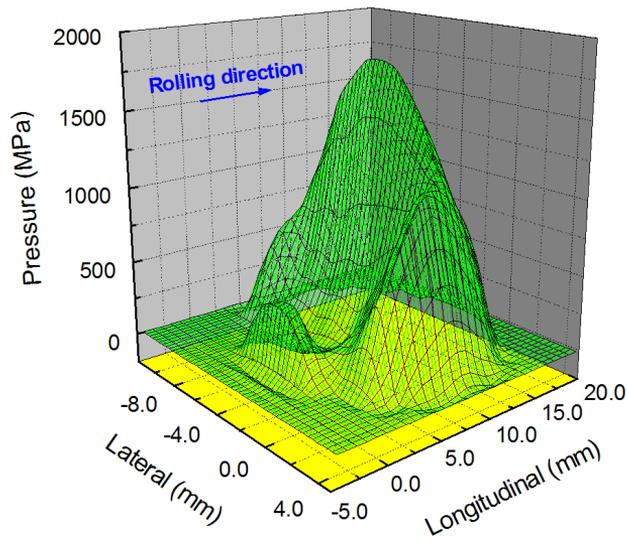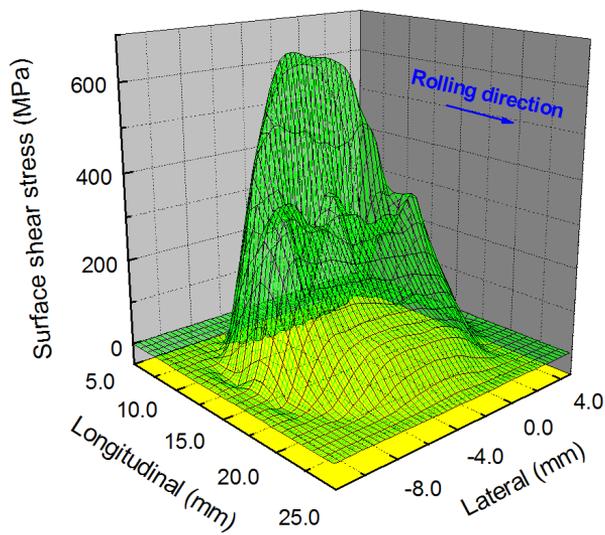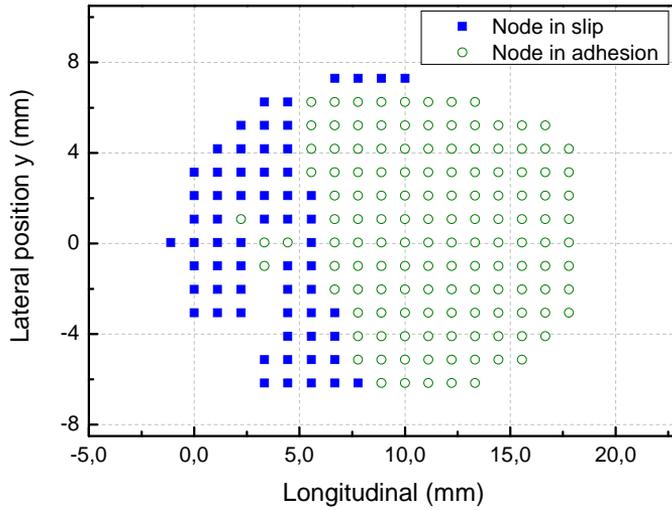
Figure 5.1.9. The slip and adhesion areas at step 654.

c) The V-M stress distribution along depth

The maximal V-M stress on the surface and along the depth in the rail at the defects is shown in Figure 5.1.10. Note that the maximal V-M stress shown at each depth in Figure 5.1.10 was found from all the V-M stress in the layer of material at that depth in the area of the trailing edge at all the time during the simulation. All the maximal V-M stress given in this paper was determined in this way. It guarantees that the true maximal is found. From the results it can be seen that with the increase in the defect size, the V-M stress increases, irrespective of the COF and the traction coefficient. For the rail material grade R260Mn, the tensile strength is around 900MPa. It can be seen in Figure 5.1.10 that for some of the cases the tensile strength of R260Mn is exceeded at the surface, while for some other not. This confirms the expectation that there should be a critical size below which the maximal V-M stress at the surface can be lower than the tensile strength, and shakedown may finally set in so that the defect cannot grow further. Here the effect of plastic deformation on the reduction of the stress by make the contact geometry more conforming has not been taken into account. Here it is also emphasized that it is the surface V-M stress which is used to compare with the material strength, because in the sub-surface the material will be under high hydro-static pressure and the material behaviors differently. The material parameters for such condition are not yet available.
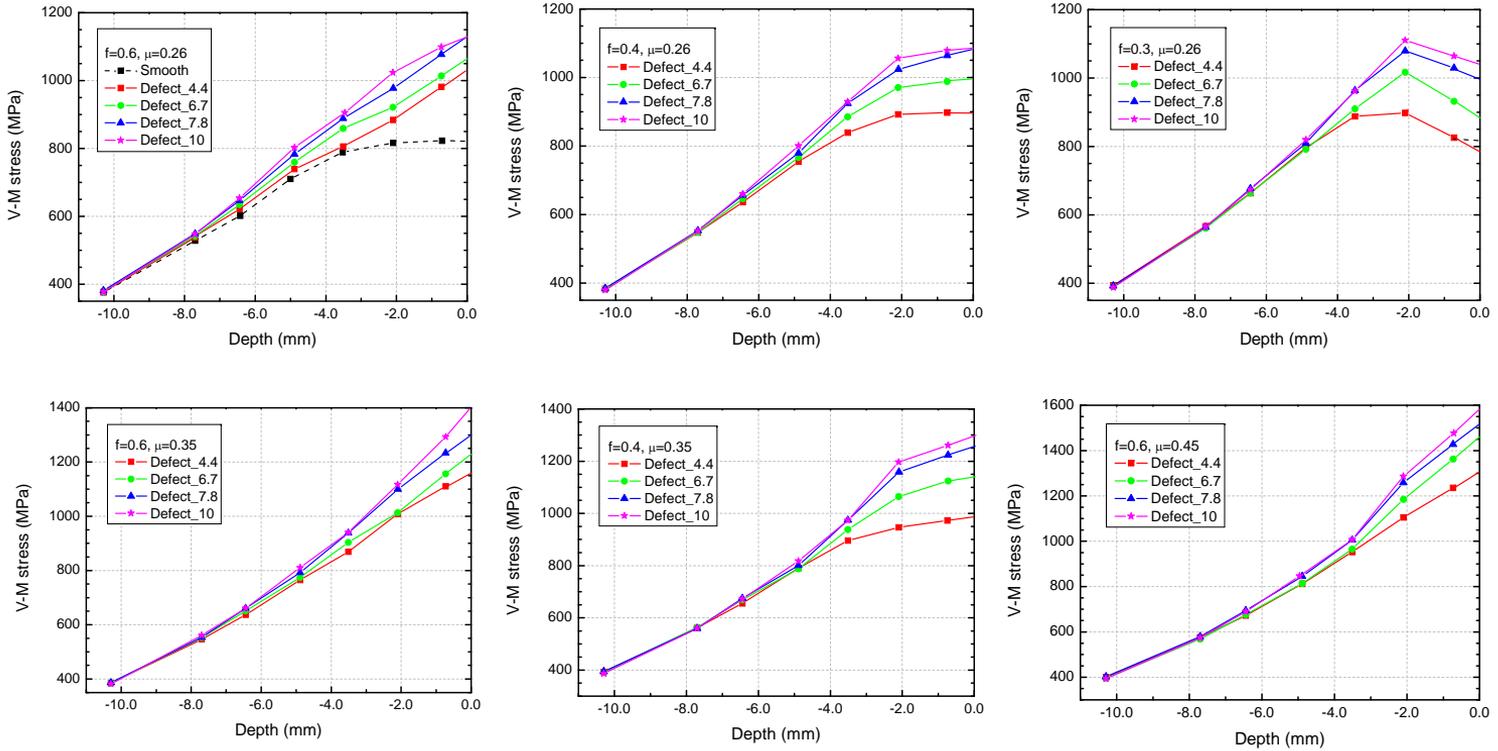
Figure 5.1.10. The Maximal V-M stress distributions along the depth in the rail at the defects with different COF (f) and traction coefficient (μ).Origin of the abscissa is at the rail top surface. The results for 'smooth' are obtained from the rail without defects.

5.1.4.2   The Influence of COF and Traction Coefficient

Putting the results of Defect_6.7 with different *f* and *μ* in the same figure (Figure 5.1.11) the influence of COF and traction coefficient on the V-M stress can be seen. Figure 5.1.11(a) shows that the stress level significantly increases with the rise of the traction level when the COF is kept constant. On the other hand, when the traction level keeps the same, the maximal stress on the rail surface also rises with the increase of the friction coefficient, whereas the stress in the subsurface changes conversely, as shown in Figure 5.1.11 (b). The cause of this phenomenon is explained as follows.
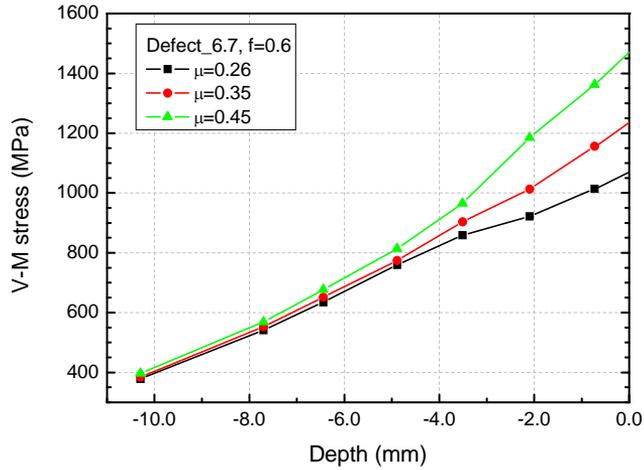
(a)
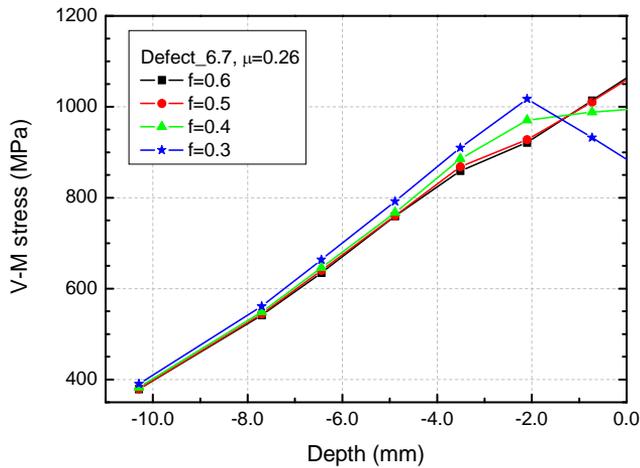
(b)

Figure 5.1.11. Influence of friction coefficient and traction coefficient on the maximal V-M stress for Defect_6.7. (a) The influence of traction coefficient and (b) the influence of friction coefficient.

Assuming a constant traction coefficient of 0.26 and for the same normal traction distribution, the tangential traction distribution in the contact patch is plotted in Figure 5.1.12 with two different friction coefficients. When the friction coefficient rises from 0.3 to 0.6, it is seen that the maximal tangential traction increases from *P2* into *P1*. Correspondingly, the slip area also shrinks greatly. Because the surface V-M stress level is mainly determined by the tangential traction in the contact patch, the surface stress in Figure 5.1.11(b) increases with the friction coefficient.

For the stress level in the subsurface, it is another situation because the V-M stress is determined by both the normal pressure and the tangential stresses. The maximal pressure occurs always at the center of the contact patch, while the highest surface shear stress changes with COF *f*, i.e it shifts to the left handed side with increasing *f* (Figure 5.1.11(b)). Their combined effect changes the maximal V-M stress from increase with increasing *f* on the surface to decrease with increasing *f* when it is deep in the subsurface, for instance at the depth of −2 mm, see Figure 5.1.11(b).
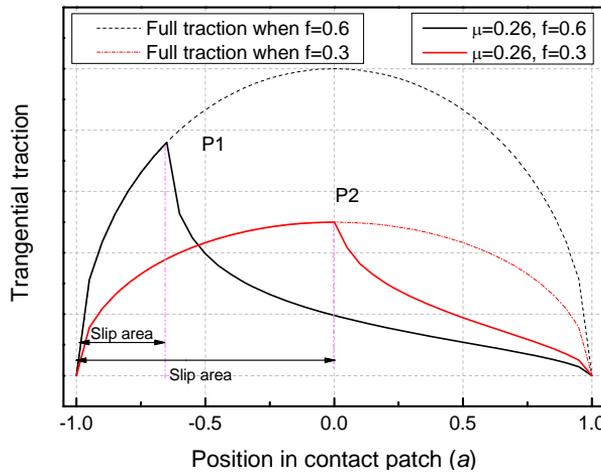
Figure 5.1.12. The tangential surface stress distribution in a contact patch under the same pressure when friction coefficient changes.

5.1.4.3   The Influence of Local Curvature at the Edge of Defects

For the above analysis the edges of the defects have been sharp, i.e. with an infinite curvature (see Figure 5.1.5). This may be the case when an indentation is just formed. The edges will, however, be run flatter by passing wheels. In order to find out the influence of the local curvature at the defects, the longitudinal profile of Defect_6.7 is adjusted at the trailing edge as shown in Figure 5.1.13, named as Defect_6.7_tran. The lateral profile is kept the same as in Figure 5.1.5(b).

As expected the influence of local curvature change on the total resultant contact forces is negligible. However, the surface V-M stress decreases significantly with the decrease in the local curvature, no matter what the friction coefficient is, as shown in Figure 5.1.14. The influence of the local curvature decreases quickly with increasing depth into the rail and disappears gradually when the depth is over more than 3 mm.

This is due to the stress concentration at the sharp edge. The higher is the local curvature, the more concentrated the stress is and the more plastic deformation would occur locally for the same load. The plastic deformation will make the contact geometry smoother and more conformal, as illustrated in Figure 5.1.13 with the change of the defect from Defect_6.7 to Defect_6.7_Tran. Consequently, the stress will be lower at the deformed defect for the same loading conditions (see Figure 5.1.14). If the load is not high enough the material may shakedown and no further plastic deformation will occur, i.e. the defect will not grow further.
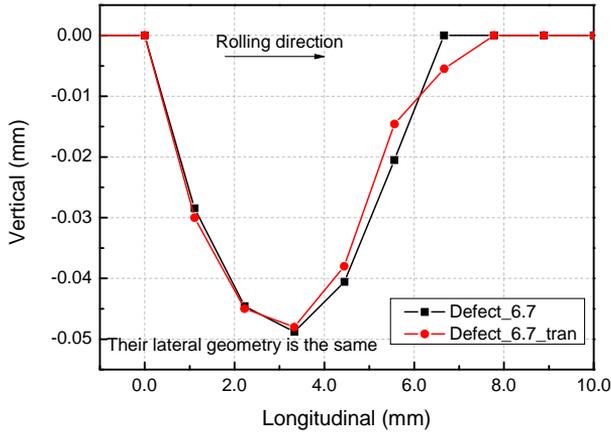
Figure 5.1.13. The local curvature variation at the trailing edge of Defect_6.7.
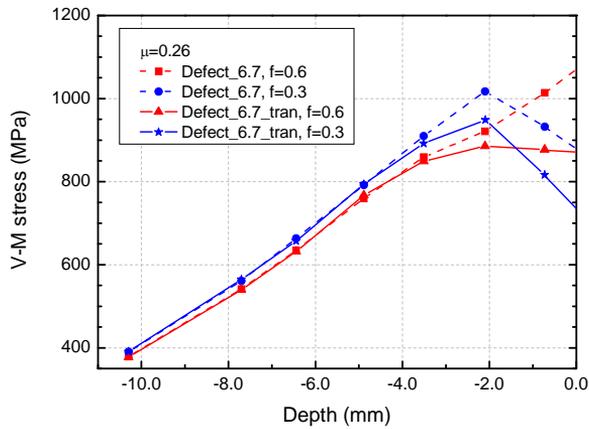
Figure 5.1.14. The maximal stress along depth at different defects. Depth = 0.0 is at rail surface.

5.1.4.4   The Maximal V-M Stress at Smoothed Defects

For the defect profiles shown in Figure 5.1.5, most of the resulting surface V-M stress level is higher than the tensile stress, see Figure 5.1.10. Therefore, most of the defect profiles will be changed due to plastic deformation by wheel passages, becoming

smoother and more conformal, as discussed in section 5.1.4.3. Assuming that after some wheel passages the defects take the shapes shown in Figure 5.1.15 (designated as *_smooth), their V-M stress level is again calculated. In defining the shape of the defects their local curvature is kept for the trailing edges as similar to each other as it is allowed by the FE meshing, as shown in Figure 5.1.15. But the curvature cannot be identical due to the limited number of discrete nodal points of the FE model in the concerned area. The local curvature is chosen based on observations of measured longitudinal-vertical rail profiles at defects, see Figures 5.1.4 (a) and (b) for examples. The lateral profiles are kept the same as shown in Figure 5.1.5.
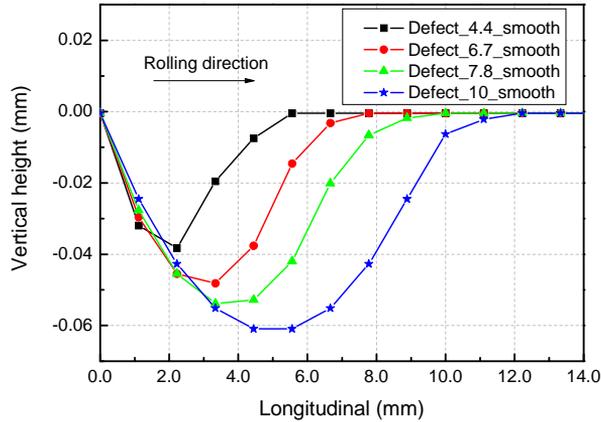
Figure 5.1.15. The generalized profiles of defects in the longitudinal-vertical section, the trailing edges of which are assumed to be smoothed by plastic deformation due to wheel passages. The leading edges are the same as those of Figure 5.1.5(a).

The maximum V-M stress on the surface and at different depth in the rail is shown in Figure 5.1.16. It can be seen that for $\mu = 0.26$ some of the maximal V-M stress already exceeds 900MPa (the tensile strength of the R260Mn rail grade).

Under normal operational conditions the traction coefficient of most of the wheels will not be higher than 0.26 for the electrical multiple units in the Netherlands. For instance, as mentioned earlier, the design value of the maximal traction coefficient is 0.22 for the mainstream VIRM, whereas the adhesion coefficient for braking is usually smaller than for traction. Hence, the results with the traction coefficient of 0.26 are employed for further analysis on defect growth, taking the worst scenario into account.

In Table 5.1.2 the maximal V-M stress is listed for all the simulated defects with $\mu=0.26$, with sharp or smoothed trailing edge and different COF. For the smoothed cases, the maximal V-M stress, which is close to the material strength of 900MPa, is given in bold letters. The maximal COF available in steel-on-steel wheel–rail contact under dry friction is widely accepted as being 0.6. In reality such a high value is usually rare [5.1.8]. From this we may conclude that 6.7 mm is the lower bound of the critical size, and 7.8mm is the upper bound for a defect to grow into a squat, i.e. when a defect is short in the rolling direction than 6.7 mm, its chance to grow into a light squat is very small, while if it is longer than 7.8 mm, the chance is high.
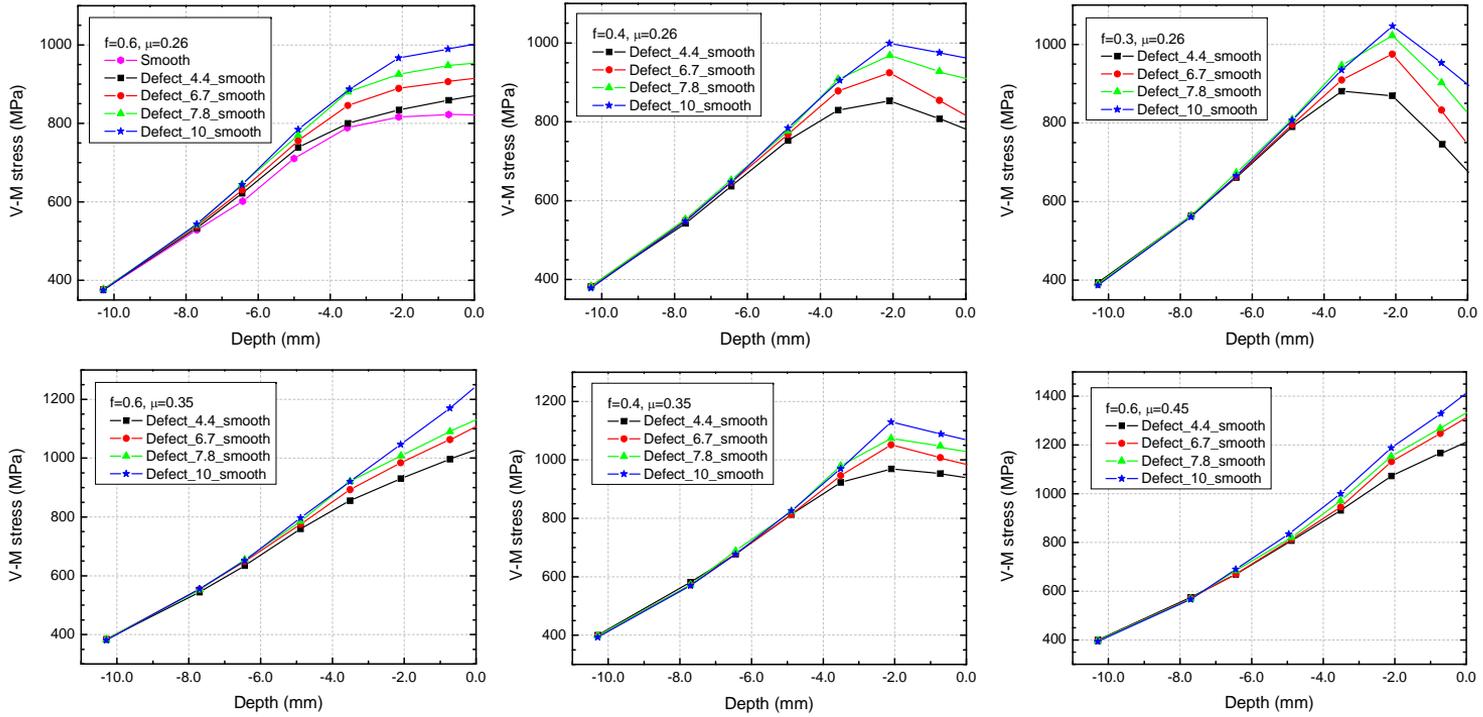
Figure 5.1.16. The Maximal V-M stress distribution at defects under different friction coefficient (f) and traction coefficient (μ). Origin of the abscissa is at the rail top surface.

Table 5.1.2 the maximal surface V-M stress (unit: GPa) μ=0.26

f	0.6	0.5	0.4	0.3
Defect_4.4	1.02	0.98	0.9	0.79
Defect_6.7	1.06	1.04	1.0	0.89
Defect_7.8	1.12	1.11	1.07	1.0
Defect_10	1.12	1.12	1.07	1.05
Defect_4.4_Smooth	**0.87**	0.85	0.78	0.68
Defect_6.7_Smooth	**0.91**	**0.87**	0.82	0.76
Defect_7.8_Smooth	0.95	**0.92**	**0.91**	0.84
Defect_10_Smooth	1.0	0.98	0.96	**0.9**

As it is mentioned above, the defect profiles were assumed based on observations of measured defect profiles. In Figure 5.1.17 two profiles used above, Defect_6.7_Tran (see Figure 5.1.13) and Defect_6.7_Smooth (see Figure 5.1.15), are put together. The corresponding maximal surface V-M stress is listed in Table 5.1.3. The small difference in the smoothed profiles also causes some small differences in the stress. In reality, the profiles of the defects and squats vary largely and change continuously due to wear and plastic deformation. The generalize profile is only a representative of their average.

Not only there are diversities in the COF and in the defect profile, there are also other unpredictable factors in the vehicle–track interactions system, like irregularities in the track, variations in speed, in the thickness of rail head and wheel tire, etc. In view of these and for easy practical use, we may round the lower bound of 6.7 mm of the critical size off downwards to 6 mm, and the upper limit 7.8 mm upper wards to 8 mm, and conclude that a defect less than 6mm long has little chance to grow, while a defect longer than 8mm will almost certain to grow into a squat. Since large defects will cause large contact stress, see Figures 5.1.10 and 5.1.16, it can be expected that larger defects will grow faster.
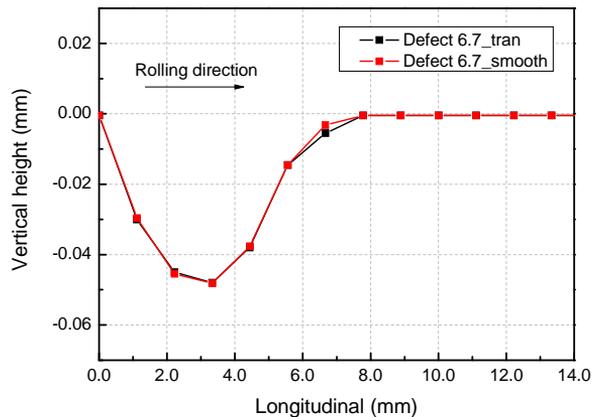
Figure 5.1.17. Comparison of two smoothed defect profiles

Table 5.1.3 influence of different smoothed profiles
(μ = 0.26, unit: GPa)

f	0.6	0.3
Defect_6.7_Tran	0.89	0.75
Defect_6.7_Smooth	0.91	0.76

Comparing lateral size of each defect shown in Figure 5.1.5(b) with its counterpart in the rolling direction (Figure 5.1.5(a)), they are approximately equal to each other. This was designed to reflect observations: the longitudinal and lateral dimensions of a matured squat are usually similar, except when the squat is severe so that it is much extended in the longitudinal direction. This means that we can generalize the derived critical size to both the longitudinal and lateral directions, namely when a defect is smaller than 6mm in both directions, its chance to grow into a squat is very small, and when it is larger than 8mm, the chance is large.

5.1.5   Verification of the critical size

Field monitoring confirmed that the numerically derived critical size is valid. Figure 5.1.18 shows that two rail defects of 5 and 6 mm long in the rolling direction disappeared due to wear, and a defect of 10mm grew into a moderate squat with the typical V form. Figure 5.1.19 shows the growth of another defect, possibly caused by indentation, of about 11 mm in its initial length in the rolling direction. The evidence of its growth can best be seen in the widening of the running band at the defect: in June 2007, it was about 35mm, and in May 2009, it became 42 mm, while elsewhere the width of the running band remained more or less the same. Running band widening is a major characteristics of squats [5.1.9–5.1.11]. According to the squat growth process postulated in [5.1.4] and validated in [5.1.5], the defect of Figure 5.1.19 will grow into a squat.

It may be worth noting that for safety reason the monitoring was carried out on a track section with only 3 MGT traffic a year. It is because of the light traffic and the small size, the defects in Figures 5.1.18(b) and 5.1.19 grew slowly. For the same reasons, no defects of the size between 8 and 10 mm have been found by monitoring which show clear growth.

5.1.6   Discussions

The FE model has been applied to typical Dutch railway loading conditions for the derivation of a critical size for squats to initiate and grow from small defects. The generalized defect is typically applicable to indentations and wheel burns. They are a major part of the squat initiation sources, especially indentations. For corrugation its wavelength is already larger than the derived critical size. But the magnitude of corrugation is in its early stage smaller than the depth of the defects assumed in Figure 5.1.5; the derived critical size may therefore not directly be applicable to corrugation. To determine the relation between the magnitude of the corrugation and the resulting contact force, stress and depth of the affected material layer by the dynamic interaction, the presented methodology may be employed.

The parameters considered in this work are the longitudinal and lateral dimensions of the defects, the coefficient of friction and the traction coefficient. The tensile strength of the material is used as the limiting parameter. It implies that if the strength is enhanced, the critical size can be increased.

The track parameters have not been varied for short wave irregularities in the track though severe squats will often cause local track quality deterioration. It is mot considered here because the critical size is intended for squat initiation from small rail surface defects, the resulting dynamic force in the track is assumed to be not yet large enough to have caused significant deterioration of the track quality. On the other hand, long wave dynamic force caused by vehicle–track interaction has been taken into account by increasing the static wheel load by 30%.

At high frequency wheel–rail interaction at defects such as squats, not only the unsprung mass is important, but also its distribution. The unsprung mass and its distribution, name those of the wheelset and the rail, together with the traffic speed and other vehicle and track parameters used in this paper are typical of the Dutch operational conditions. For other railway the conditions may be somewhat different.
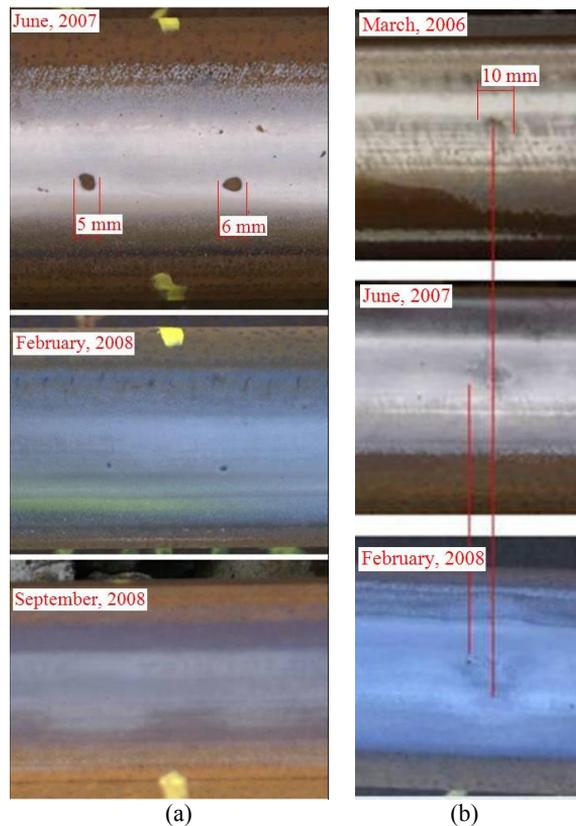
Figure 5.1.18. Verification of the critical size. (a) two defects of 5 and 6mm were erased by wear, and (b) a defect of 10mm had grown into a moderate squat with the typical V form [5.1.4].
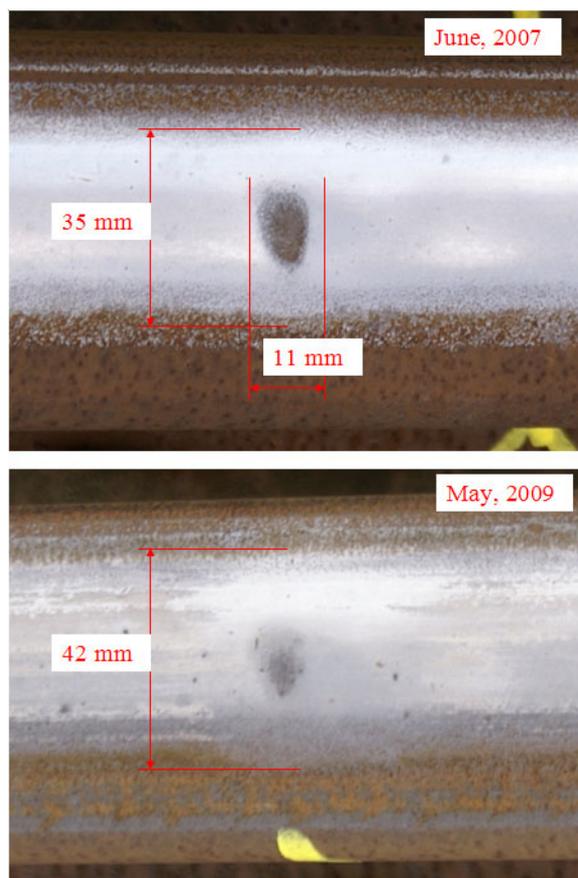
Figure 5.1.19. Verification of the critical size. A defect of original length 11mm or so grew.

5.1.7   Conclusions

In this paper an FE wheel–track interaction model is integrated with detailed frictional rolling contact solution. It is applied to derive a critical size for small rail top geometry defects to grow into squats. It is found that when a defect is smaller than 6mm in both the rolling and lateral directions, its chance to grow into a squat is very small, and when it is larger than 8mm, the chance is large. This result has been verified by field monitoring.

The critical size has been determined for the typical Dutch railway operational conditions, but the method is generally applicable to other railways.

The critical size can be used as a criterion for visual inspection to distinguish between light squats and other trivial defects, and consequently for maintenance action, such as rail grinding. It may also be used for automatic image recognition of light squats.

References

[5.1.1] P. Clayton, M.B.P. Allery, and P.J. Bolton, Surface damage phenomena in rails, in J. Kalousek R.V. Dukkipati and G.M.L. Gladwell Ed.: Proceedings of the conference on Contact Mechanics and Wear of Rail/wheel Systems, Vancouver, British Columbia, July 6-9, 1982, 419-443.

[5.1.2] S. Bogdanski and R. Lewicki, 3D model of entrapped liquid in rolling contact fatigue cracks in rails, Wear 265 (2008), 1356 – 1362.

[5.1.3] Z. Li, X. Zhao, R. Dollevoet, and M. Molodova, Differential wear and plastic deformation as a cause of squat at track local stiffness change combined with other track short defects, Vehicle System Dynamics 46(Supp) (2008), 237-246. DOI: 10.1080/00423110801935855.

[5.1.4] Z. Li, X. Zhao, C. Esveld, R. Dollevoet and M. Molodova, An investigation into the causes of squats: correlation analysis and numerical modeling, Wear 265 (2008), 1349-1355.

[5.1.5] Z. Li, X. Zhao, M. Molodova and R. Dollevoet, The validation of some numerical predictions on squats growth, Proceedings 8th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems, Florence, Italy, 15 – 18 September, 2009, 369 - 378, ISBN: 978-88-904370-0-7.

[5.1.6] J. Smulders, Management and research tackle rolling contact fatigue, Railway Gazette International, June 2003, 439 – 442.

[5.1.7] X. Zhao, Z. Li and R. Dollevoet, Solution of the wheel–rail rolling contact in elasticity and elasto-plasticity using a transient finite element model, Proceedings 8th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems, Florence, Italy, 15 – 18 September, 2009, 871 – 879, ISBN: 978-88-904370-0-7.

[5.1.8] Z. Li, O. Arias-Cuevas, R. Lewis and E. A. Gallardo-Hernandez, Rolling–sliding laboratory tests of friction modifiers in leaf contaminated wheel–rail contacts, Tribol Lett, 33 (2009), 97-109. DOI 10.1007/s11249-008-9393-3.

[5.1.9] UIC Code 712, Rail defects, 4th edition, January 2002

[5.1.10] Rail Damages, the Blue Book of Rail Track, February, 2001, UK

[5.1.11] Classification of Squats explained with examples, Squat handout, ProRail, version 2, March 2004 (in Dutch)

Summary and recommendations

Simulations have shown that transient contact solutions are greatly affected by the rail surface defects. As a consequence, higher dynamic stress is achieved at the defected area. Based on stress analyses of defects of different sizes, a methodology has been developed in this chapter to determine a critical size for rail surface defects to grow into squats. For the typical Dutch railway system, it is found that the critical defect size is between 6 and 8 mm. Defects smaller than such a size will gradually be worn away. Monitoring tests conducted in the Netherlands have verified such a critical size.

6. Conclusions and recommendations

Conclusions of this research are presented in Section 6.1 from three aspects: the contact solutions, the modeling of high frequency vehicle–track interaction, and the squatting. In Section 6.2, recommendations for future work are proposed to better understand rail damages such as squats and to further develop the three-dimensional FE model.

Chapter 1: Introduction to squats and their consequences.	Chapter 2: Literature review and outline of this dissertation.

Chapter 3: Solution of frictional rolling contact.

Chapter 4: Squat initiation, growth and wheel–track dynamic forces.

Chapter 5: A minimum critical size for squats to initiate.

Chapter 6: Conclusions and recommendations.
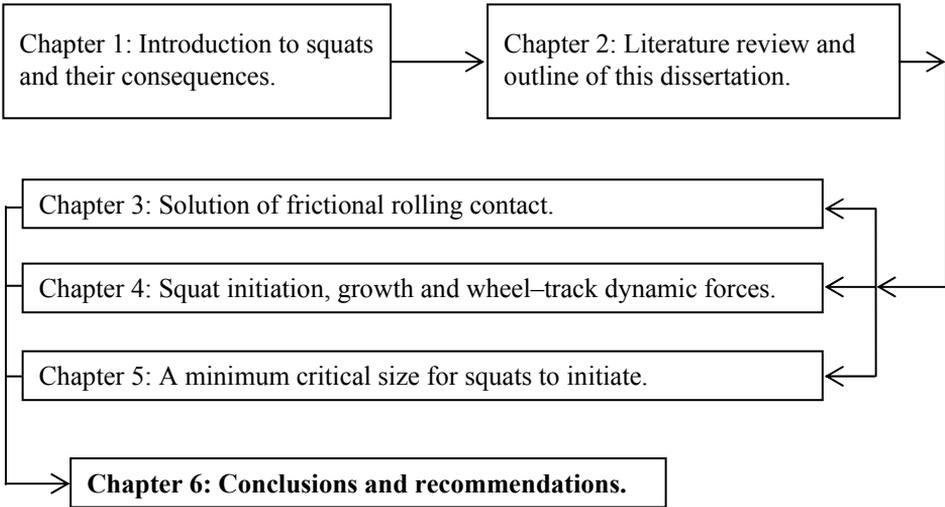
6.1 Conclusions

6.1.1 Contact solutions

A dynamic FE model has been developed to solve the frictional rolling contact between a wheel and a rail in the time domain. Both normal and tangential solutions are obtained and have been validated in statics. Different friction and traction conditions can be simulated for both elasticity and elasto-plasticity. The following conclusions are drawn.

1. Normal and tangential contact solutions obtained from the FE model are sufficiently accurate for investigations into rail failure. Different creepages and spin can be considered.

2. Quasi-static or steady-state rolling can also be simulated with the FE model by applying a sufficient dynamic relaxation.

3. Small fluctuation exists in the contact solutions. This should be due to the continuum and structural vibrations included by the modeling approach. Numerical errors of the FE method may also have played a role. This needs further assessment.

4. Compared to solutions of Hertz's theory and Kalker's CONTACT, the elastic solution of the FE model confirms that the half-space approximation is valid for contact between smooth wheel tread and rail top.

5. The elastic FE solution shows that the tangential contact has negligible influence on the normal solution when the same elastic material is used for both bodies in contact and when the contact area is flat.

6. When plastic deformation occurs, the normal problem becomes non-linearly related to the tangential problem. This is because plastic deformation significantly modifies the contact geometry and the magnitude of the plastic deformation increases with the friction force transmitted in the contact patch.

7. In terms of the steady-state rolling contact between wheel tread and rail top, the influences of plastic deformation on contact are as follows. Compared to the elastic solution:
   - The contact patch is increased in size, shifted forward in the rolling direction, and changed from an ellipse to an egg shape.
   - In the rear part of the contact patch, the pressure is reduced in the middle but increased at both sides. In contrast, it is increased in the whole range in the front part of the contact patch. As a result, the maximum pressure is reduced in magnitude and moved forward in its longitudinal position.
   - The magnitude of the surface shear stress varies following the trend of pressure in the slip area, whereas it decreases along the longitudinal axis with a lower descending rate in the adhesion area.

- The micro-slip redistributes due to plastic flow and modifications of the contact patch and the slip area.
- The maximum V-M stress in the contact patch is reduced, and residual stress is left behind if there is no pre-existing residual stress.

8. When the friction force transmitted in the contact patch is considerably less than its limit, the influence of friction and traction conditions is as follows.
   - When the friction coefficient is constant:
     - o Under elastic condition, the maximum surface shear stress in the contact patch increases with the traction force, while its location moves toward the contact center. In contrast, the same pressure distribution remains. Consequently, a larger traction force leads to a higher V-M stress in the contact patch and along the depth.
     - o Under elasto-plastic condition, the pressure distribution becomes related to the tangential problem as mentioned above, but the V-M stress in the contact patch and along the depth still increases with the traction force due to material hardening. Furthermore, a larger traction force leaves a higher residual stress behind.
   - When the traction force is kept constant and the friction coefficient changes, the surface shear stress in the contact patch is redistributed because its limit is modified. As a result, under elastic condition, the maximum surface shear stress increases with friction coefficient, but its location moves away from the contact center or towards the trailing edge. Furthermore, because of the same pressure distribution, the maximum V-M stress in the contact patch increases with friction coefficient, and along depth the increase gradually reduces in magnitude and turns into a decrease at sufficient depths.

6.1.2   Modeling of high frequency vehicle–track interaction

A 3D dynamic FE model of vehicle–track interaction has been developed to investigate the high frequency dynamics in the presence of short wavelength defects such as squats. The FE contact model is integrated into the vehicle–track model to solve the frictional rolling contact between wheel and rail. Employment of the explicit time integration ensures that the transient effect is sufficiently considered. Therefore, continuum vibrations are included in the solution together with the structural ones. Nonlinearities involved in contact, structure, and materials can all be simulated. Detailed stress distributions in the contact can be calculated. Conclusions drawn from the work of this dissertation are as follows.

1. The FE model has been proved to be a valid and promising tool to treat the high frequency dynamics of a system containing rolling contact and to solve the problem of rolling contact in the presence of Coulomb friction. Plasticity, complicated contact geometry and dynamics can all be considered.

2. The predicted high frequency response of the vehicle–track system to squats by the FE model is in agreement with field observations and train borne measurements.

3. Detailed modeling such as the 3D FE model employed in this work is necessary for the investigation of squats. A mass–spring–beam dynamic model does not allow a proper stress analysis due to the lack of consideration of the continua.

6.1.3 Squatting

1. Squats, a kind of rail RCF, are related to high frequency dynamics. The wavelength of squats is typically between 20 mm and 40 mm, which corresponds to 950–1900 Hz for a rolling speed of 140 km/h and bears some resemblance to corrugation.

2. These wavelength and frequency characteristics of squats can be employed for early detection of squats by measurement of dynamic responses excited by the rail surface defects. This has been proven by the axle box acceleration detection system developed by TU Delft.

3. A growth process of squats has been postulated to explain how a light squat grows into a mature one.
   - It is believed that the growth of squats is dominated by the related dynamic contact force. The damage mechanism behind this process may be differential wear and/or differential plastic deformation.
   - Because the dynamic contact force is determined by certain resonances of the coupled vehicle–track system, a light squat always grows to the typical "squatting" shape, no matter what its origin is.
   - Measurement of dynamic responses and field monitoring, conducted in the Netherlands, have verified this growth process.

4. Squats initiate at indentations, wheel burns, welds, corrugation, etc. Light squats do not normally have a typical "squatting" shape. However, the majority of mature ones typically bear the "squatting" shape.

5. Squats can only initiate from rail surface defects (such as indentations) that are large enough.
   - The stress level at a rail surface defect is higher than at smooth rails due to the contact geometry variation and the related dynamic contact forces.
   - For the Dutch case, the critical size is found to be 6–8 mm. Defects smaller than such a size will be worn away gradually. This has been verified by track monitoring.
   - Taking into account the critical defect size in classifying light squats, false statistics of squats can be reduced. Trivial defects can be left out unnecessary maintenance work can be avoided.

6. The vehicle–track system is mainly excited by squats in the vertical direction at two frequencies. In other words, the vertical vibration caused by a squat is composed mainly of two vibration components.
   - The vibration component at the higher frequency approximately corresponds to the typical wavelength of squats observed in the field, i.e. 20–40 mm, as mentioned above.

- The frequencies of the two components do not change with the dimensions of the squat. This suggests that the two vibration components might be related to some resonances of the coupled vehicle–track system.
- With the enlargement of a squat, the vibration component at the lower frequency increases faster in magnitude than the other. Consequently, the force oscillation at a relatively large squat is dominated by the vibration component at the lower frequency.
- For the investigated rolling speed range of 40–140 km/h, both the magnitude and wavelength of the two vibration components increase with the rolling speed. The vibration component of the higher frequency can be absent when the rolling speed is sufficiently low, e.g. at 40 km/h for the simulated system, because the excitation frequency of a squat decreases as rolling speed reduces.

7. The two vibration components are transferred into the discrete supports of the rail in different ways.
   - The lower frequency component can transfer down to the ballast layer, especially at the supports close to the squat, while the other one has similar magnitude at fastenings around the squat and is negligible at the ballast layer.
   - A mature squat mainly increases the forces at the two supports of the sleeper bay where the squat exists.
   - For short wavelength defects, the dynamic forces at the fastenings do not represent the dynamic wheel–rail contact force. For a typical mature squat, the peak magnitude at the most heavily loaded fastening is 15–25% less than at the wheel–rail interface for the speed of 40–140 km/h.

8. Simulations have shown that the plastic deformation of the surface material, as the physical essence of squatting, is highly related to rolling contact with high friction exploitation.

The work of this dissertation has been the basis for a guideline to best practice of squat treatment, written upon invitation by UIC [6.1].

6.2   Recommendations on future research

6.2.1   To understand and treat the physics of rail damage

6.2.1.1   Squatting and its counter measures

It can be imagined that if wheels steadily roll over a smooth rail of homogeneous material, squats as isolated damages will not occur because contact forces and stresses are constant along the rail. Therefore, an initial track defect, which can cause oscillation of contact stresses at the fixed location, is required to trigger the squatting phenomenon. Such a process is shown in Figure 6.1 with a flow chart.
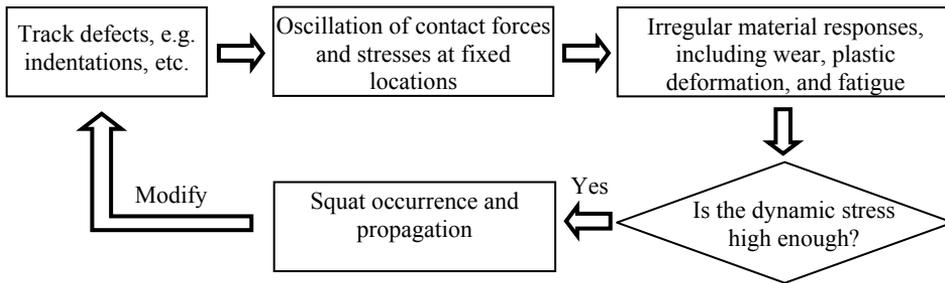
Figure 6.1. The occurrence of squats.

In reality, oscillation of contact stresses can result from different track defects. Among them, singular rail top defects analyzed in Section 5.1 are considered as one of the most important origins of squats. However, not every singular rail top defect can lead to a squat because a rail material can sustain a certain oscillation of contact stresses on the existing networks, which leads to the concept of critical defect size (see Section 5.1 for more information). Except singular rail top defects, other origins such as corrugations, crossings and switches, etc. have not been examined yet. Their critical states for squat initiation, which are probably not defined by sizes of the defects, could be analyzed in the future.

It should be emphasized that initial track defects also include inhomogeneity of the rail material, e.g. at welds. Squats may initiate from the weak parts of an inhomogeneous rail when the strength of the material is low enough. The critical state for squat initiation under this situation could also be studied in future work.

To prevent the occurrence of squats, i.e. to break the growth cycle shown in Figure 6.1, counter measures may be designed from the following aspects: to remove/reduce the origins, to reduce the dynamic forces/stresses at an origin, and to increase the tolerance of the rail material to oscillation of contact stresses. The last two strategies, which increase the tolerance of a railway system to initial track defects, probably lead to a series of counter measures maintaining the technical states of the track system, especially the parameters influencing the high frequency dynamics. In the meantime, better understanding of the material damage mechanisms is also needed. All these require more investigations into the vehicle–track interaction at squats.

This dissertation has found that squats mainly excite the vehicle–track system at two frequencies. However, the origins of the two vibration components, which might be related to some resonances of the coupled vehicle–track system, have not been identified yet. This should be further studied so that better counter measures against squats may be designed.

6.2.1.2   Material damage mechanisms

The three damage mechanisms of rail surface material, namely wear, plastic deformation, and RCF, should be studied separately, and then integrated to understand rail damages such as the initiation of squats.

a) Wear

Currently, frictional work based wear models are normally employed to predict the wear behavior of rail surface material [6.2, 6.3]. Because the transient results of surface shear stress and micro-slip are available from the FE solution for deriving the frictional work, the wear behavior of the material can further be examined in detail with a suitable wear coefficient or a series of wear coefficients. This provides a means for understanding the development of some irregular wear in which the dynamic effects need be considered.

b) Cyclic plasticity

To understand the initiation mechanism of rail surface defects such as squats, cyclic plasticity may have to be taken into account, by which work hardening and pre-existing residual stress are considered inherently. Today, the main difficulties of studying cyclic plasticity lie in:

- The computation costs of simulating cyclic loading are still too high for PC workstations, especially when the 3D transient problems are treated and the profile changes with time need to be considered. To overcome this obstacle, computers with higher performance have to be employed.
- To the best knowledge of the author, the plastic behavior of rail steels under cyclic loading has not yet been well known for a multi-axial stress field. This is particularly true when the high hydrostatic stress in the rail is considered.

Bearing in mind the rapid development of computer hardware and software in the past decades, 3D simulations of cyclic plasticity including the contact calculation should become available soon. Simplified material models may first be employed. In the future, well-established plastic material models for cyclic loading, if necessary, may be employed for more accurate predictions.

c) RCF

By a suitable fatigue model, initiation of fatigue cracks may be assessed based on the stresses and strains obtained from simulations of cyclic loading. The competition between fatigue initiation and wear can further be examined to estimate the occurrence of RCF. Subsequently, the dominant damage mechanism of a rail, among wear, plastic deformation, and RCF, can be determined, and the remaining service life of the rail can be evaluated. The development of squats can also be better understood.

6.2.1.3 Corrugations

Corrugations, a problem observed over a century, have been found to be an important origin of squats on the one hand. On the other hand, studies have also shown that squats and corrugations are similar in wavelength. The latter may suggest certain unidentified relationships between squats and corrugations. Hence, it may lead to a better understanding of corrugation to apply the approaches of this dissertation to the investigation of corrugations.

6.2.1.4   Continuum vibrations

Small fluctuations of stress have been observed in contact solutions of the FE model. These are inherent and realistic for dynamic problems, and should be largely caused by some high frequency vibrations of the simulated system, mainly the continuum vibrations. Further studies are needed to verify this, which may lead to a better understanding of the transient rolling contact phenomenon. In addition, those continuum vibrations occurring in the vicinity of the contact patch may be of importance for rolling noise generation and damage related to high frequency wheel–rail interaction.

6.2.2   Further developments and applications of the 3D FE model

The FE model can further be improved from other aspects, some of which are discussed below.

6.2.2.1   Rolling contact

In this dissertation, the FE model is only employed to simulate the contact between wheel tread and rail top on a tangent track. The contact patch is located in the middle of the rail top. However, to treat problems in curves where the wheelset significantly moves in the lateral direction, the model has to be improved to include the lateral creepage and the related curving behavior of wheels.

It should be noted that the spin, unlike the longitudinal and lateral creepages, is automatically taken into account in a simulation of the 3D FE model. Conformal and multi-point contact problems can also be solved if a proper mesh is applied. These are attributed to the fact that the transient rolling of a wheel over a rail is simulated in the model with the consideration of their real 3D geometry.

Although the wheel–rail rolling contact is examined in this dissertation, the approach may be adapted to many other rolling contact problems such as of gears and bearings. To such an end, geometry and materials of the contact bodies, friction and traction conditions, rolling speeds, loads, and the structure of the simulated system have to be defined accordingly.

6.2.2.2   Friction models

In this dissertation, the friction between wheel and rail has been modeled with Coulomb's law of dry friction. A contact interface in reality can be wet, lubricated, or contaminated; the contacting surfaces have various degrees of asperities. For such situations, more advanced friction models may be needed. In [6.4], for instant, a friction model was developed for lubricated contact by greases.

References

[6.1]   Z. Li, A guideline to best practice of squat treatment, UIC, Oct. 2011.
[6.2]   Z. Li, Wheel–rail rolling contact and its application to wear simulation, PhD dissertation, TU Delft, the Netherlands, 2002.

[6.3]   K. Knothe, A Groß-thebing, Short wavelength rail corrugation and non-steady-state contact mechanics, Vehicle system dynamics 46 (2008), 49–66.

[6.4]   R.I. Popovici, Friction in the wheel–rail contacts, Ph.D. thesis, University of Twente, Enschede, The Netherlands, 2010.

Relevant publications

JOURNAL PUBLICATIONS

1. X. Zhao, Z. Li, A 3-D finite element solution of frictional wheel–rail rolling contact in elasto-plasticity, Wear (Submitted).
2. X. Zhao, Z. Li, J. Liu, Wheel–rail impact and the dynamic forces at discrete supports of rails in the presence of singular rail surface defects, Journal of Rail and Rapid Transit, 226 (2012) 124–139.
3. X. Zhao, Z. Li, The solution of frictional wheel-rail rolling contact with a 3-D transient finite element model: validation and error analysis, Wear 271 (2011) 444–452.
4. X. Zhao, Z. Li, C. Esveld and R. Dollevoet, The effects of the tangential traction on wheel-rail dynamic contact. WSEAS Transactions on Applied and Theoretical Mechanics (ISSN: 1991-8747) 2 (2007) 52 –59.
5. Z. Li, X. Zhao, C. Esveld, R. Dollevoet and M. Molodova, An investigation into the causes of Squats: Correlation Analysis and Numerical Modeling, Wear 265 (2008) 1349–1355.
6. Z. Li, X. Zhao, R. Dollevoet, and M. Molodova, Differential wear and plastic deformation as causes of squat at track local stiffness change combined with other track short defects, Vehicle System Dynamics 46 (Supp) (2008) 237–246.
7. Z. Li, R. Dollevoet, M. Molodova and X. Zhao, Squat growth—some observations and the validation of numerical predictions, Wear 271 (2011) 148–157.

CONFERENCE PUBLICATIONS

1. X. Zhao, Z. Li, and R. Dollevoet, The dynamic vertical force at the discrete supports of the rail in the presence of singular rail surface defects. Proceedings of RASD2010 (10th international conference), Southampton, UK, 12–14 July 2010.
2. X. Zhao, Z. Li and R. Dollevoet, Solution of the wheel-rail rolling contact in elasticity and elasto-plasticity using a transient finite element model. Proceedings of the 8th international conference on contact mechanics and wear of rail/wheel system (CM2009), Florence, Italy, 15–18 September 2009, 871–879.
3. X. Zhao, Z. Li, and R. Dollevoet, An investigation on squats growth using a transient finite element model, Proceedings of CROW Infradagen 2008, Delft, 25–26 June 2008.
4. X. Zhao, Z. Li, and R. Dollevoet, An investigation on elastic-plastic rolling contact over rough surfaces using a 3-D dynamic finite element model, Proceedings of the STLE/ASME International Joint Tribology conference 2008, IJTC2008-71091, Miami, Florida, USA, 20–22 October 2008.
5. X. Zhao, Z. Li, C. Esveld and R. Dollevoet, The dynamic stress sate of the wheel-rail contact. Proceedings of 2nd IASME/WSEAS international conference on continuum mechanics, Portoroz, Slovenia, 15-17 May 2007, 127–133. (Best student paper award)
6. Z. Li, X. Zhao and R. Dollevoet, The determination of a critical size for rail top surface defects to grow into squats, Proceedings of the 8th international conference on contact

mechanics and wear of rail/wheel system (CM2009), Florence, Italy, 15–18 September 2009, 379–388.

7. Z. Li, X. Zhao, and R. Dollevoet, Rolling contact at rail top geometry defects: reality and simulation, Advances in Contact Mechanics: a tribute to Prof. J.J. Kalker, Delft, the Netherlands, 22–24 October 2008, 57–58.

8. Z. Li, X. Zhao, C. Esveld and R. Dollevoet, Rail stresses, strains and fatigue under dynamic wheel-rail interaction. Proceedings of international heavy haul conference 2007, Kiruna, Sweden, 11–13 June 2007, 389–396.

9. Z. Li, X. Zhao, R. Dollevoet, and M. Molodova, Differential wear and plastic deformation as a cause of squat at track local stiffness change combined with other track short defects. IAVSD 20th Symposium, Berkeley, USA, 13–17 August, 2007.

10. Z. Li, X. Zhao, C. Esveld and R. Dollevoet, Causes to squats: correlation analysis and numerical modeling. Proceedings of the 7th international conference on contact mechanics and wear of rail/wheel system (CM2006), Brisbane, Australia, 24-27 September 2006, 439–446.

11. Z. Li, M. Molodova, X. Zhao and R. Dollevoet, Squat treatment by way of minimum action based on early detection to reduce life cycle costs. Proceedings of the 2010 Joint Rail Conference, JRC2010, Urbana-Champaign, Illinois, USA, 27–29 April 2010.

12. Z. Li, R. Dollevoet, M. Molodova and X. Zhao, The validation of some numerical predictions on squats growth, Proceedings of the 8th international conference on contact mechanics and wear of rail/wheel system (CM2009), Florence, Italy, 15–18 September 2009, 369–377.

Curriculum vitae

Name:	Xin Zhao （赵鑫）
Date of Birth:	13-12-1981
Place of Birth:	Taian, Shandong, P.R. China

Jan. 2006 – June 2012	PhD candidate in Section of Road and Railway Engineering, Delft University of Technology, the Netherlands
Sep. 2003 – Dec. 2005	Master student in Vehicle Application Engineering, Southwest Jiaotong University, Chengdu, P.R. China, cum laude, excellent master thesis award of Sichuan province.
Sep. 1999 – Jul. 2003	Bachelor student in Heat Energy & Power Engineering, Southwest Jiaotong University, Chengdu, P.R. China