

**SINTEF Teknologi og samfunn**

Postadresse: 7465 Trondheim
Besøksadresse: S P Andersens veg 5
7031 Trondheim
Telefon: 73 59 36 13
Telefaks: 73 59 02 60

Foretaksregisteret: NO 948 007 029 MVA

SINTEF RAPPORT

TITTEL

Pemro aktivitet 2.1:**Analyser av infrastruktur-relaterte forsinkelser på strekningen
Asker – Oslo S**

FORFATTER(E)

Inger-Anne F. Sætermo, Nils Olsson og Mads Veiseth

OPPDRAKSGIVER(E)

RAPPORTNR.	GRADERING	OPPDRAKSGIVERS REF.	
GRADER. DENNE SIDE	ISBN	PROSJEKTNR.	ANTALL SIDER OG BILAG
ELEKTRONISK ARKIVKODE	PROSJEKTLEDER (NAVN, SIGN.)		VERIFISERT AV (NAVN, SIGN.)
SINTEF RAPPORT aktivitet 21 v2.doc	Nils Olsson		Mads Veiseth
ARKIVKODE	DATO	GODKJENT AV (NAVN, STILLING, SIGN.)	
	2006-10-06	Bjørn Andersen	

SAMMENDRAG

For å kunne iverksette tiltak som forbedrer punktlighet må man ha eksakt og konkret informasjon om *årsaken* til punktlighetsbristene.

I analysene som er omtalt i denne rapporten ble informasjon om registrerte forsinkelser i TIOS sammenstilt med informasjon om registrerte feilhendelser i systemene Banedata og Banemeldings-sentralen.

Slik systemene er i dag er det ikke mulig å få sammenstilt informasjonen slik at man får et godt og representativt bilde av punktlighetsbrister og tilhørende årsak. Systemene må kobles via *tid* og *sted*, og tid og sted for punktlighetsbrist korresponderer ikke med tid og sted for årsakshendelse. Analysene viser så langt at:

- Informasjon om hendelser kan ikke sammenstilles automatisk.
- Manuelt kan man sammenstille informasjon om et fåtall hendelser. Resultatet som framkommer er mangelfullt og kan gi et feil bilde av situasjon og årsaksforhold.
- Det er avdekket manglende registreringer i Banedata og BMS.

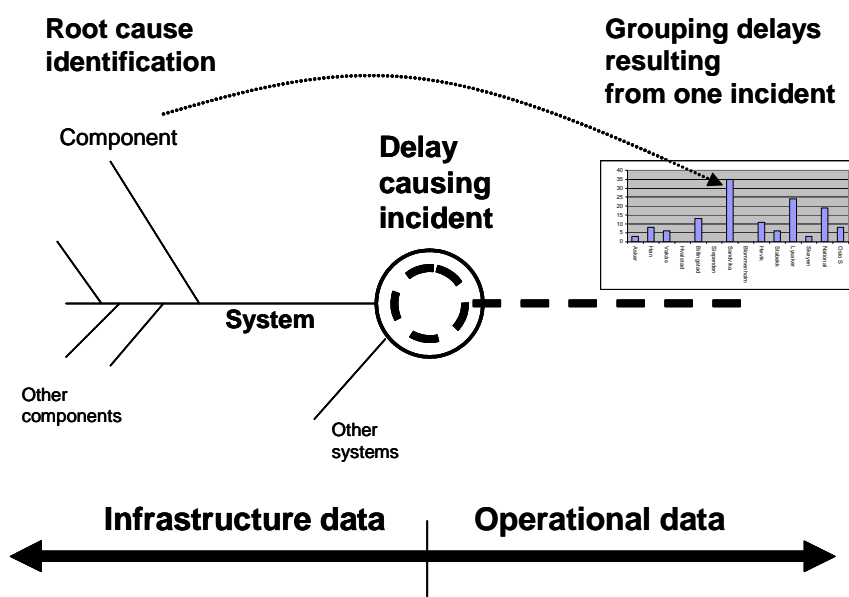
STIKKORD	NORSK	ENGELSK
GRUPPE 1		
GRUPPE 2		
EGENVALGTE		

INNHOLDSFORTEGNELSE

1	Bakgrunn og formål.....	3
2	Metode	3
3	Beskrivelse av case	3
4	Beskrivelse av datasystemer.....	4
	4.1 TIOS	4
	4.2 Banedata	4
	4.3 Banemeldingssentralen (BMS).....	4
5	Analyser	5
	5.1 Innledende analyser med TIOS	5
	5.2 Sammenstilling av TIOS og Banedata	7
	5.3 Sammenstilling av TIOS og Banemeldingssentralen (BMS).....	10
6	Oppsummering.....	12
	Vedlegg A Infrastructure's influence on rail punctuality.....	13

1 Bakgrunn og formål

Målet med forskningsrådsprosjektet Pemro er å utvikle metoder og system for identifisering, bruk og oppfølging av prestasjonsindikatorer for jernbanedrift i Norge. Systemet TIOS har vært sentralt i mye av arbeidet som er gjort i Pemro-prosjektet. TIOS forteller hvor en forsinkelse er *registrert*, men ikke hvor den *grunnleggende årsaken* til forsinkelsen er. For å kunne iverksette tiltak som forbedrer punktligheten må man ha eksakt og konkret informasjon om *årsaken* til punktlighetsbristen.



Figur 1.1 Årsaksforhold og registrering

Som vist i Figur 1.1 kan en årsak være lokalisert et helt annet sted i tid og rom, enn tilhørende registrering av punktlighetsbrist. Ideen bak aktivitet 2.1 i Pemro var å sammenstille driftsinformasjonen i TIOS med informasjon i andre systemer for å se om det er mulig å få tak i de sanne årsakene til punktlighetsbrister. Ettersom det finnes flere andre systemer som inneholder informasjon om infrastrukturrelaterte feilhendelser ble det besluttet å se spesielt på det.

2 Metode

Det ble besluttet å definere et case (dvs en strekning og en periode) og bruke systemene TIOS, Banedata og Banemeldingssentralen (BMS) for å få informasjon om avvik og punktlighetsbrister.

Først analyseres caset ved hjelp av TIOS, deretter sammenstilles relevant informasjon i TIOS med tilsvarende informasjon i Banedata og BMS.

3 Beskrivelse av case

Caset som ble valgt er strekningen Asker – Oslo S i perioden 1/6-2005 til 31/5-2006. All trafikk på strekningen i denne perioden er tatt med i analysene.

For strekningen og perioden som helhet er det 35 341 registrerte forsinkelser totalt i TIOS¹, og i sum utgjør dette 3 773 forsinkelsestimer.

4 Beskrivelse av datasystemer

4.1 TIOS

TIOS (Trafikkinformasjon og oppfølgingssystem) er en landsdekkende database som skal inneholde informasjon om toghendelser, togene (materiell, farlig gods osv.), kapasitetsbegrensninger og avviksinformasjon. Systemet er utviklet og eid av Jernbaneverket, men operatørene kan få tilgang til systemet for å se registreringer på sine tog.

På strekninger som dekkes av fjernstyringssystemene VICOS og EBICOS innhentes data om togposisjoner automatisk, på øvrige strekninger må slike data overføres til TIOS manuelt. Registrering av årsaker til avvik gjøres manuelt. Alle tog som er mer enn 4 minutter forsinket mellom to målepunkter skal registreres med en årsakskode.

Med utgangspunkt i datagrunnlaget til TIOS kan man lage ulike former for rapporter over punktlighet for regioner, togprodukter og strekninger.

4.2 Banedata

Et av systemene som gir informasjon om mulige feilårsaker innenfor strekningen og perioden vi har definert som case er Banedata. Banedata inneholder informasjon om alt utstyr og komponenter som infrastrukturen består av, samt informasjon om hendelser relatert til infrastrukturen som f.eks feilhendelser og svikt. For at en hendelse skal kunne registreres i Banedata må hendelsen kunne relateres til en bestemt komponent. Infrastrukturen består av svært mange komponenter, og når disse skal registreres i Banedata må tekniske egenskaper ved komponenten også legges inn. For enkelte strekninger mangler det komponenter i registeret. Kravet om at hendelsene skal relateres til en komponent gjør dermed at det kan være etterslep av feilregistreringer med tilhørende mangelfull rapportering.

Banedata inneholder informasjon om hendelse, komponent, sted, tid, tidsperiode hvor hendelsen påvirket trafikken, samt aksjoner som er iverksatt for å rette feil.

4.3 Banemeldingssentralen (BMS)

Banemeldingssentralen inneholder data om alle rapporterte hendelser relatert til infrastrukturen. Informasjonen er derfor lik den som finnes i Banedata, men i Banemeldingssentralen er ikke de rapporterte hendelsene linket opp mot noen spesiell komponent. Det er flere hendelser registrert i Banemeldingssentralen enn i Banedata.

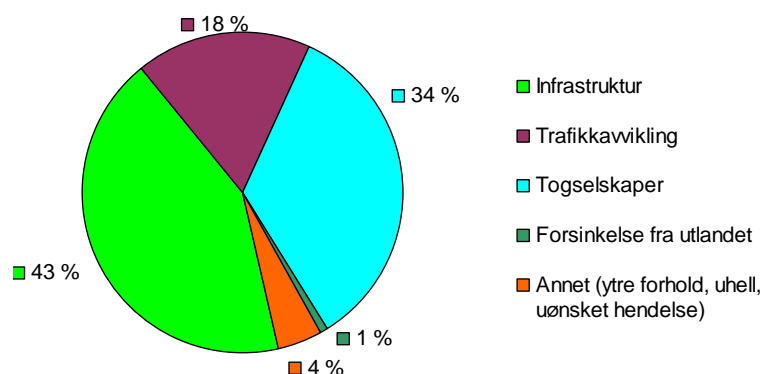
¹ Dvs registrert merforsinkelse (utover forrige registrering) på 4 min eller mer

5 Analyser

5.1 Innledende analyser med TIOS

TIOS er brukt til å analysere forsinkelsesomfang, antall registrerte merforsinkelser samt årsaksangivelser for det utvalgte caset. Vi har sett på ulike dimensjoner innenfor dette og dekomponert informasjonen per operatør, over tid på året osv. Det er også gjort en analyse av tognummer og frekvens for å se om det kan avdekkes enkelt-tog som har spesielle punktlighetsproblemer.

Forsinkelsestimerne fordeler seg på ulike forhold som vist i Figur 1.1.



Figur 5.1 Forsinkelsestimer fordelt på årsakskoder. (Asker – Oslo S 1/6-2005 til 31/5-2006)

De to årsakskodene som i sum står for mest forsinkelse er infrastruktur med 43 % av forsinkelsestimerne, og det som kan relateres til togselskapene som står for 34 % av forsinkelsestimerne. Når det gjelder infrastruktur er det på denne strekningen mest forsinkelse knyttet til sikringsanlegg, signalanlegg og fjernstyring. Når det gjelder årsakskoder som relateres til togselskapene er det i all hovedsak ”stasjonsopphold” som er oppgitt som hovedårsak. Dette skal da være stasjonsopphold som ikke er forårsaket av trafikkstyringen, men av andre forhold som togselskapene er ansvarlig for. En detaljering av årsakskodene i Figur 5.1 er vist i Tabell 5.1.

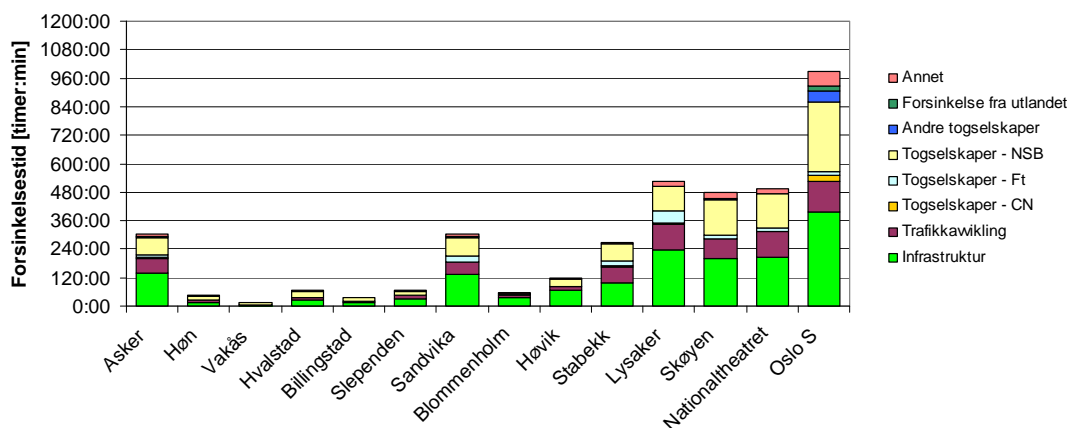
Tabell 5.1 Forsinkelsestimer fordelt på årsakskoder. (Asker – Oslo S 1/6-2005 til 31/5-2006)

Forsinkelsesårsak		Timer:min
<i>Infrastruktur</i>	1. Bane	286:05
	2. Sikringsanlegg, signalanlegg, fjernstyring	888:13
	3. Elkraft/Kontaktledningsanlegg	158:55
	4. Tatt ut og lagt til kode 2	00:55
	5. Planlagt vedlikeholdsarbeid infrastruktur	164:52
	6. Materiell med feil sperrer sporet/blokkstrekning	111:26
	<i>Sum</i>	<i>1610:26</i>
<i>Trafikkavvikling</i>	<i>Sum</i>	<i>670:06</i>
<i>Togselskaper</i>	81. Feil ved materiell	198:47
	82. Materiell sent satt opp i togspor	219:25
	83. Manglende personell	106:11
	84. Stasjonsopphold	748:59
	85. Planforutsetninger ikke oppfylt	25:17
	<i>Sum</i>	<i>1298:39</i>
<i>Forsinkelse fra utlandet</i>	<i>Sum</i>	<i>24:54</i>

<i>Annet</i>	92. Ytre forhold (ekstreme værforhold og lignende)	63:18
	93. Uhell, påkjørsel	50:48
	94. Uønsket hendelse	55:03
	<i>Sum</i>	<i>169:09</i>

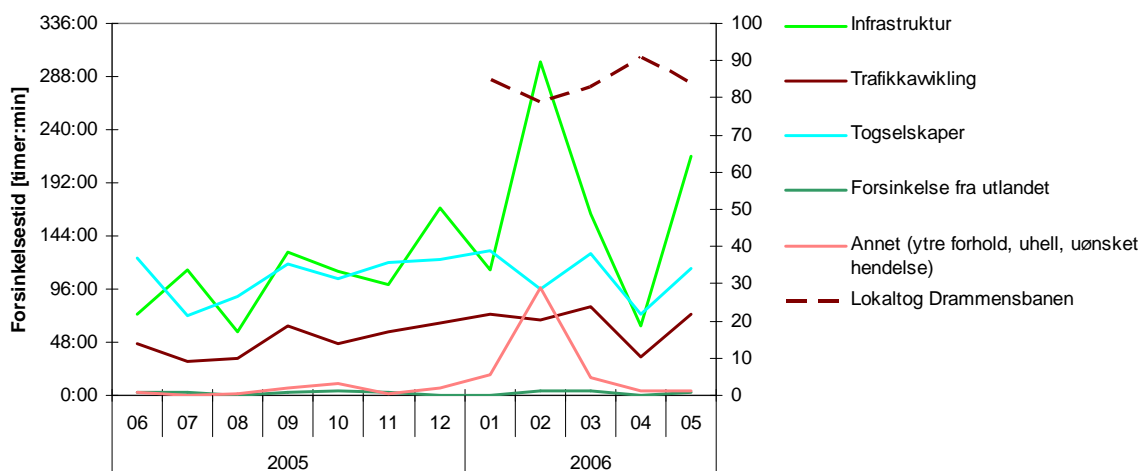
Som tabellen viser er det feil knyttet til sikringsanlegg/signalanlegg/fjernstyring og stasjonsopphold som står for de største bidragene til forsinkelsesminutter. Uten disse ville totalt registrerte forsinkelsesminutter i caset reduseres med 43%. Kode 2: sikringsanlegg, signalanlegg og fjernstyring står alene for nesten 25% av de registrerte forsinkelsesminuttene.

Vi har sett hvordan forsinkelsesminuttene fordeler seg på stasjonene på strekningen. Det er vist i Figur 5.2.



Figur 5.2 Fordeling av registrerte forsinkelsesminutter på stasjoner.

Stasjonene med mest trafikk har naturlig nok flest registrerte forsinkelsesminutter. Oslo S er stasjonen med mest forsinkelsestid – nesten dobbel så mye som Lysaker. Infrastruktur er oppført som årsak til de fleste forsinkelsesminuttene på *alle* stasjoner. Det er også laget en oversikt over hvordan forsinkelsesminutter av ulike årsaker fordeler seg i tid, se neste figur.



Figur 5.3 Fordeling av forsinkelsesminutter over tid

Det mest påfallende i mønsteret er en klar topp for infrastrukturrelaterte feil samt kategorien ”annet” i februar 2006. Dette henger sammen med et uvanlig stort snøfall på Østlandet, samt en

avsporing på Asker stasjon. (Den stiplede kurven til høyre relaterer seg til prosenttallene på høyre akse og viser punktlighet for Drammensbanen de første 6 månedene i 2006.)

Det ble videre utført en analyse som viser hvilke tog som hadde flest registreringer av hver av feilkodene. En oversikt over dette er vist i Tabell 5.2. Det er svært mange tog som har mange registreringer av samme type feil, for eksempel har tog 2143 totalt 137 registreringer av forsinkelser knyttet til stasjonsopphold, mens tog 5302 har 80 registreringer av forsinkelser knyttet til trafikkavvikling. Et høyt antall gjentatte registreringer indikerer at man bør se nærmere på disse togene, vurdere ruteleie osv.

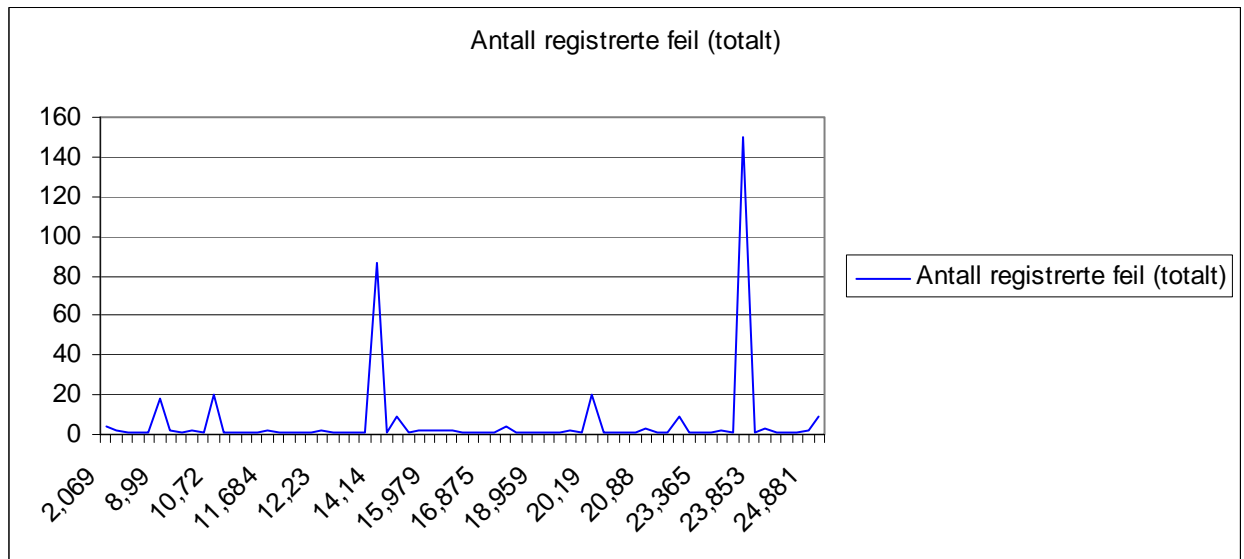
Tabell 5.2 "Gjengangere" på listen over registrerte forsinkelser

Forsinkelsesårsak		Tognr	Antall
<i>Infrastruktur</i>	1. Bane	2142	23
	2. Sikringsanlegg, signalanlegg, fjernstyring	1608	48
	3. Elkraft/Kontaktledningsanlegg	1121	11
	5. Planlagt vedlikeholdsarbeid infrastruktur	2181	73
	6. Materiell med feil sperrer sporet/blokkstrekning	1110	13
		<i>5302</i>	<i>80</i>
<i>Trafikkavvikling</i>			
<i>Togselskaper</i>	81. Feil ved materiell	1027	25
	82. Materiell sent satt opp i togspor	41881	57
	83. Manglende personell	1051	10
	84. Stasjonsopphold	2143	137
	85. Planforutsetninger ikke oppfylt	1509	16
<i>Forsinkelse fra utlandet</i>	<i>Sum</i>	<i>411881</i>	<i>34</i>
<i>Annet</i>	92. Ytre forhold (ekstreme værforhold og lignende)	2114	10
	93. Uhell, påkjørsel	1657	8
	94. Uønsket hendelse	875	5

5.2 Sammenstilling av TIOS og Banedata

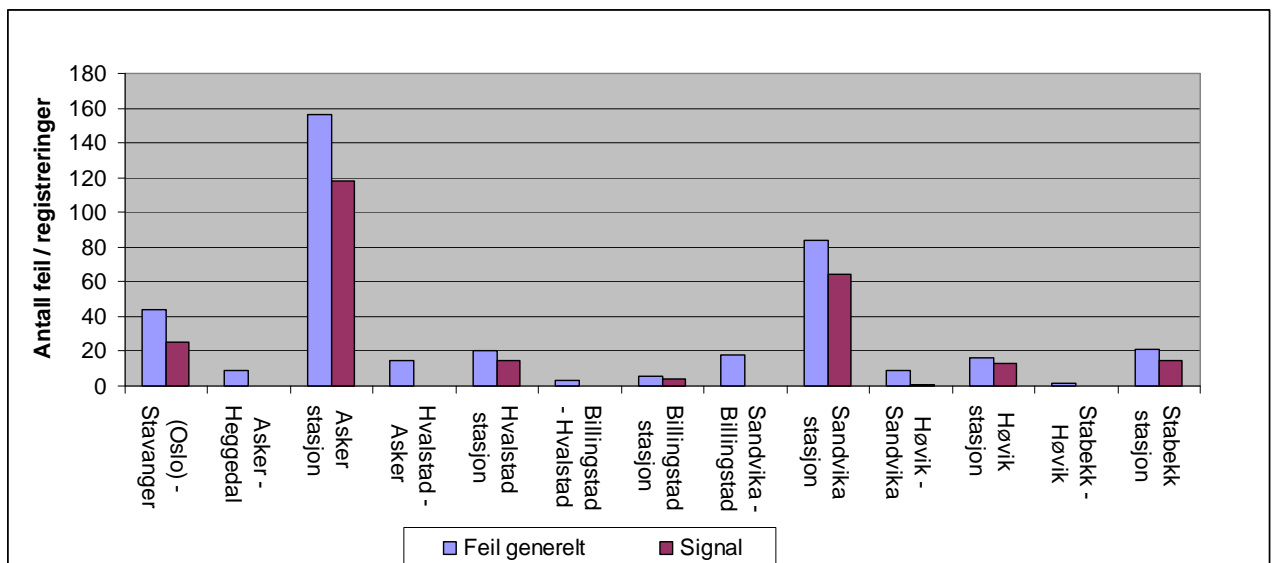
Banedata inneholder informasjon om infrastrukturen: Systemet gir en oversikt over utstyr og komponenter, samt informasjon om hendelser relatert til infrastrukturen som f.eks feilhendelser og svikt. En oversikt over hvordan registrerte feil i banedata sprer seg over strekningen viser at de aller fleste feilene registreres på stasjonsområder. Banedata-inndeling gjør at datasettet vi fikk tilsendt stopper før Oslo S. Følgende stasjoner er vist i Figur 5.4:

- Stabekk stasjon: km 8.99
- Høvik stasjon: km 10.72
- Sandvika stasjon: km 14.14
- Hvalstad stasjon: km 20.19
- Asker stasjon: km 23.85



Figur 5.4 Fordeling av feil langs banestrekningen

Hvis man går inn i det underliggende materialet ser man at det er flest registrerte feil knyttet til signalanleggene. Andelen signalfeil av totalt antall registrerte feil kommer fram av figur 5.5.



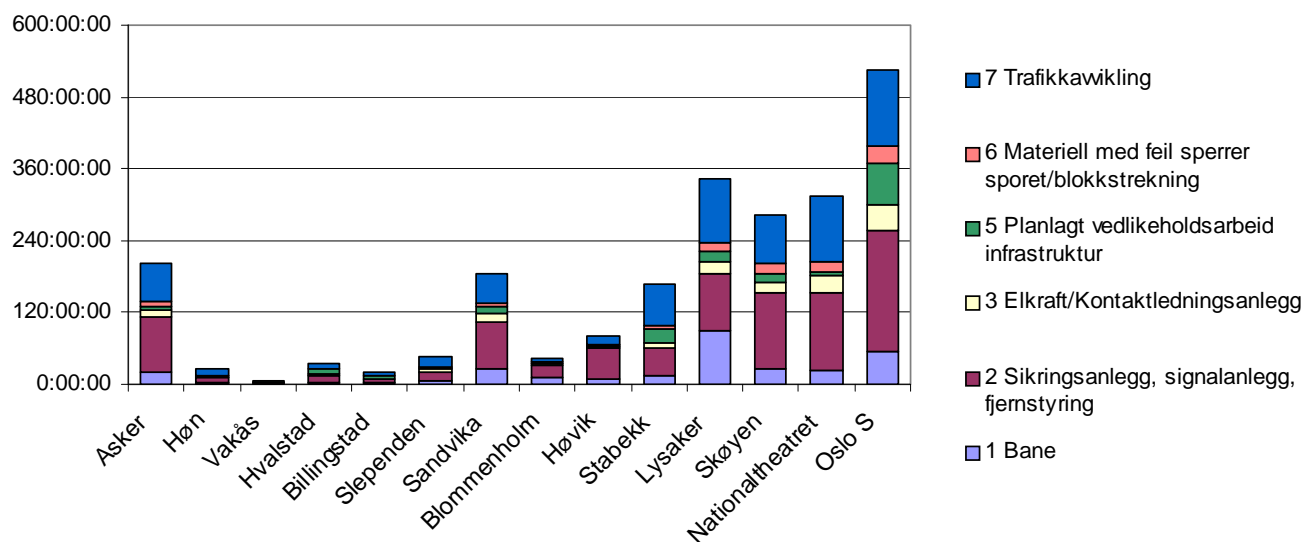
5.5 Fordeling av feil generelt og signalfeil langs banestrekningen

På strekningen som dekkes er Asker stasjon stedet med flest registrerte feil. De fleste feilene er relatert til signal, og en tredjedel av signalfeilene er knyttet til at sporveksel/sporsperre ikke er i kontroll.

Også Sandvika har et høyt antall registrerte feil i Banedata, sikringsanlegget er en vesentlig kilde til feilregistreringer. Årsaken til dette er også her at sporveksel/sporsperre ikke er i kontroll, årsaken er manglende smøring/renhold, Snø og is samt manglende snørydding.

Tilsvarende tall kan hentes ut fra TIOS, se Figur 5.6. Her er det kun sett på feil relatert til infrastruktur og trafikkavvikling, da det er disse som er sammenlignbare med registreringene i

BaneData. Også TIOS har registrert store forsinkelser på Asker stasjon, og sikringsanlegg/signalanlegg/fjernstyring er den årsakskategorien som bidrar til flest forsinkelsesminutter.



Figur 5.6 Oversikt over forsinkelsesminutter per stasjon (Kilde: TIOS)

På et overordnet nivå stemmer det godt overens – stasjoner og årsakskoder som er registrert i TIOS og Banedata gir samme bilde av driftssituasjonen. Men i tråd med intensjonene i denne aktiviteten prøvde vi å gå bakenfor dette, for å se om vi kan finne konsekvensene i form av registrerte forsinkelsesminutter i TIOS av *hver enkelt feil* som er registrert i Banedata.

I Banedata registreres tid og sted hvor *feil* oppstår. I TIOS registreres tid og sted for hvor *forsinkelse* oppstår. Feilene som er registrert i Banedata kan gi opphav til forsinkelser andre steder og til andre tider – med andre ord kan man ikke bruke informasjon om tid og sted i TIOS til å finne hvilke feil som forårsaket de ulike forsinkelsene. På samme måte kan man ikke bruke tid og sted for feil i Banedata til å avdekke hvilke konsekvenser i form av forsinkelsesminutter hver feil medførte.

Vi har satt opp noen sammenligninger som illustrerer dette poenget. I tabellen under har vi sett på antall forsinkelser registrert ved tre ulike stasjoner. Det framkommer tydelig at i en betydelig andel av forsinkelsene er årsaken lokalisert et helt annet sted.

5.3 Lokalisering av årsaker

<i>Årsakskategori</i>	<i>Stasjon</i>	<i>Antall registrerte forsinkelser (number)</i>	<i>Registrerte forsinkelser hvor årsaken er et annet sted (number)</i>
<i>Sikringsanlegg, signal, fjernstyring</i>	<i>Høyen</i>	96	45
<i>Bane</i>	<i>Oslo</i>	319	83
<i>Elkraft/kontaktledningsanlegg</i>	<i>Sandvika</i>	92	47

Fordi tid og sted for forsinkelse og feil ikke nødvendigvis korresponderer må sammenstilling av informasjon i TIOS og Banedata gjøres manuelt. Koblingen må gjøres på bakgrunn av

informasjon som framkommer i kommentar-feltet i TIOS. Det er tilfeldig hvorvidt dette feltet er fylt ut eller ikke, og det er tilfeldig hvilken informasjon som legges inn i feltet. Det gjør at det er et fåtall hendelser og forsinkelser som kan sammenstilles. Det er kun mulig å koble et lite utvalg av hendelser. Resultatet av koblingen er dermed vanskelig å tolke og det er ikke usannsynlig at det er skjevheter i utvalget: Det vil sannsynligvis vil være noen store, kjente feilhendelser (avsporing på Asker f.eks) som alle vet om og som man får koblet. Mindre hendelser som kan være hyppig forekommende er kanskje ikke så kjent for alle togledere og blir dermed ikke knyttet til forsinkelsen. Dette innebærer at verdien av den manuelle linkingen av feilhendelser i Banedata og registrerte forsinkelser i TIOS ikke er stor, ettersom det som kommer fram neppe gir et godt bilde av det som skjer. Det kan derfor ikke si oss noe om hvor vi bør sette inn ressursene for å forbedre punktligheten.

Koblingen av TIOS og Banedata er en tidkrevende øvelse som er nærmere omtalt i paperet i vedlegg A. En av konklusjonene her er at ut fra omfanget på infrastrukturregistreringer i TIOS og registrering av feilhendelser i BaneData mangler det høyst sannsynlig en del hendelser i BaneData.

BaneData i seg selv gir selvfølgelig nyttig informasjon om tilstanden til infrastrukturen, men koblingen mot TIOS er per i dag svak. Resultatet av koblingen er ikke komplett, og er vanskelig å tolke. Det gir ikke et situasjonsbilde som man kan bruke til å definere tiltak for forbedring av punktligheten.

Oppsummering:

- Kobling av systemene er arbeidskrevende (må gjøres manuelt)
- Hvorvidt man får matchet hendelser eller ikke er nokså tilfeldig
- Det ser ut til at det mangler registreringer i BaneData

5.3 Sammenstilling av TIOS og Banemeldingssentralen (BMS)

Banemeldingssentralen inneholder rapporterte hendelser relatert til infrastrukturen. Hendelsene er ikke knyttet opp mot noen spesiell komponent. Banemeldingssentralen er ikke tilrettelagt for analyser – formatet er vanskelig å jobbe med, og er ikke egnet for automatisk linking til andre systemer. Eksempel på rapportering er vist i 5.7.

Banemeldingssentralen rapporterer tid og sted for når en *hendelse* ble rapportert. TIOS rapporterer tid og sted for hvor en forsinkelse oppstod.

Nr.	Feiltid	Status	Faggruppe	Ansvarlig	Beskrivelse	Sted	Objekttype
53646	2005-07-01 17:10 5 dager 14 timer	utbedret (2. Pri.)	Signal og sikring	Signalvakt Asker	(Innmelder: TXP Asker) Blunkbelegg i felt 3 og 26 samt UA feltert. Oppfølging 07-07-2005 07:40 (Larsen, signal Oslo) Glødeskall fjernet	Asker (St) 23.83 km [1410] Lysaker - Asker	
53629	2005-07-01 07:36 0 dager 6 timer	utbedret (2. Pri.)	Signal og sikring	Signalvakt Asker	(Innmelder: Togleder) Innkjørhovedsignal 4332UD viser fast rødt lys i steden for blinkende. Oppfølging 01-07-2005 13:45 (Bren) Resatt	Sandvika (St) 14.14 km [1410] Lysaker - Asker	
53615	2005-07-01 00:42	utbedret	Signal og sikring	Signalvakt Asker	(Innmelder: togleder)	Sandvika (St)	

5.7 Eksempel på rapport i Banemeldingssentralen

Banemeldingssentralen fokuserer på *hendelser*. Informasjonen er derfor på mange måter lik den som finnes i Banedata, men i Banemeldingssentralen er ikke de rapporterte hendelsene linket opp mot noen spesiell komponent. Det er flere rapporteringer om feil registrert i Banemeldingssentralen enn i Banedata.

De to oppsettene i tabell 5.4 og tabell 5.5 viser en registrert hendelser i hhv TIOS og Banemeldingssentralen som korresponderer. Dette eksemplet illustrerer forskjellen på innhold og fokus i de to systemene.

5.4 Eksempel på informasjon fra Banemeldingssentralen:

Felt	Informasjon
Nummer (ID)	72379
Tid	2006-09-09 14:28, 0 dager og 3 timer
Status	Reparert, 1. prioritet
Klassifisering	Elkraft/kontaktledningsanlegg
Ansvarlig	Elkraft, Asker
Beskrivelse	Tre over konktaktledning ved Hvalstad – Billingstad.
Sted	Hvalstad (Station), 20.190 km on [1410] (Lysaker) - Asker
Type	-

5.5 Eksempel på informasjon fra TIOS:

Årsaks- kode	Dato	Tog nummer	Stasjon	Innstilt? (Ja/Nei)	Forsinkelse	Kommentar
3	2006-09-09	2151	Hvalstad	Nei	00:04	Tre falt over ledning ved Hvalstad

Også sammenstillingen av informasjon mellom TIOS og Banemeldingssentralen må gjøres manuelt. Det er for så vidt enklere å sammenstille data fra Banemeldingssentralen med TIOS enn data fra Banedata med TIOS. Årsaken er at Banemeldingssentralen også har driftsforstyrrelser som utgangspunkt.

Det er gjort en slik manuell sammenstilling (se paper i vedlegg A), og det viser seg at enkelte hendelser lar seg matche. Sammenstillingen lar seg imidlertid kun gjøre for et fåtall hendelser, og det er stor usikkerhet knyttet til om disse hendelsene gir et godt og riktig bilde av årsaksforholdene.

Oppsummering

- Formatet i Banemeldingssentralen gjør analyser og kobling mot andre systemer vanskelig (mye manuelt tidkrevende arbeid)
- Hvorvidt man får matchet hendelser eller ikke er nokså tilfeldig

6 Oppsummering

Systemene må kobles via *tid* og *sted*. Tidsrommet for når man har en feil og tidsrommet for når denne feilen påvirker trafikken korresponderer ikke og er svært vanskelig å avgrense. Analysene viser så langt at:

- Det er i dag ikke mulig å få en god kobling mellom TIOS og hhv Banedata eller Banemeldingssentralen:
 - Informasjon om hendelser kan ikke sammenstilles automatisk.
 - Kun informasjon om et fåtall hendelser lar seg sammenstille, resultatet som framkommer er mangelfullt og kan gi feil bilde av situasjon og årsaksforhold.
- Det kan se ut til at TIOS er mest omfattende, det er avdekket manglende registreringer i Banedata og BMS.

Hvert av systemene TIOS, Banedata og Banemeldingssentralen gir nyttig informasjon om trafikkavvikling og tilstanden til infrastrukturen. Koblingen mellom systemene er vanskelig og tidkrevende, og resultatet så svakt at det ikke gir noe godt grunnlag for å prioritere ressursinnsats på tiltak for forbedring av punktligheten.

Det jobber kontinuerlig med disse systemene. Innenfor Pemro lar vi dette arbeidet ligge på is inntil det skjer endringer i systemene som gjør at det er mer å hente på kobling av dem.

Vedlegg A Paper: Infrastructure's influence on rail punctuality

Infrastructure's influence on rail punctuality

Authors:

Mads Veiseth

PhD Student

Norwegian University of Science and Technology

Inger Anne F. Saetermo

Senior Scientist

SINTEF Technology and Society

Nils Olsson

Senior Scientist

SINTEF Technology and Society

Infrastructure's influence on rail punctuality

Abstract

This paper focuses on the possibilities for root cause analyzes of delays and estimation of effects of delay causing incidents by linking infrastructure data and operational data. Such information could be used to prioritize improvement activities and to quantify the effects of improvement measures. We have analysed and compared data from two infrastructure databases and one punctuality database in Norway. Linking of data between the databases should theoretically and technically be possible but our analyses have shown that it is not possible. To make linking of information possible we recommend the railway industry to develop a general code-system for all three databases that directly tie the information together.

Introduction and background

Punctuality is highly valued by railway customers and is claimed to be the number one factor determining railway service quality in most countries (NEA, 2003). Most railway companies set punctuality targets for their services but punctuality statistics tell that the actual level of punctuality lies below the targets for many rail services² (see for example Jernbaneverket, 2006). Veiseth and Bititici (2006) have compared best practice of performance measurement with measurement of punctuality in Norway, Sweden and Scotland. They conclude there is need for more focus on processes and factors that influence punctuality in the improvement work. They also recommend the railway industry to link measurements of punctuality with other management and data systems to make the measurement system more balanced and complete.

From punctuality statistics in Norway, it can be found that about 20- 30% of the total delay minutes are caused by infrastructure faults, which are distributed between three main categories: the track, the signal system, and the power supply system (Jernbaneverket, 2005). For most lines in the Norwegian railway network, all train movements are automatically registered by the signal system at different places along the lines and stored in a database. The database does also hold information about the actual schedule. How many minutes the trains are delayed can therefore easy be calculated. Manual registrations of what has caused the delays can also be found in the database, but this data is not detailed enough to track how many minutes of delay a selected infrastructure fault has generated. Such information had been valuable for the improvement work to investigate cause and effects relationships, to prioritize improvement activities and to quantify the effects of improvement measures.

This paper discusses the possibilities of using additional information sources to estimate how infrastructure faults affect punctuality and to analyze root causes of delays. The paper starts with a general discussion about measurement of punctuality, quality of punctuality data, improvement of punctuality and what causes delays. Based on this we propose a way to improve quality of punctuality data by linking operational data and infrastructure data. To investigate the possibilities

¹ The punctuality target varies between different train categories and countries as well as the way to calculate punctuality can vary between train categories and countries. A typical target is 90 % of all trains should be on time at final destination.

² The Norwegian railway administration

of this kind of linking we have analysed and compared information from three databases in the Norwegian railways: one punctuality database and two infrastructure databases. We have also discussed how the databases' design needs to be developed to achieve automatic linking of information between the data sources.

Measurement of punctuality and quality of punctuality data

In the railway industry, punctuality is measurements of the operations' reliability and performance. Rudnicki (1997) defines punctuality as "that a predefined vehicle arrives, departs, or passes a predefined point at a predefined time". This means that punctuality is related to delays which are the deviations between the actual and predefined departure or arrival time for a train (Olsson and Haugland, 2004).

Punctuality is normally measured as the percentage of trains that arrive on time at their final destinations. Trains are defined as not-punctual if they are delayed more than a predefined time limit. The size of the time limit varies typically between 1-30 minutes and depends on the type of the trains (e.g. local trains, Inter City trains and freight trains). 'Delay minutes/hours' is another indicator related to punctuality which tells how many minutes/hours trains are delayed. In addition to measurement of delays (punctuality), the delays' causes are normally registered manually by a set of codes (Veiseth and Bititci, 2006).

Data quality refers often to in which extent data satisfy the users' requirement or are suitable for a given process (Capiello et al., 2004). Redman (2005) defines data-quality as:

"Data are of high quality if they are fit for their indented uses in operations, decision making and planning. Data are fit for use if they are free of defects and possess desired features"

This definition underlines that data quality is about data free of defects and that they should be suitable for its use. Based on this definition we conclude that punctuality data of good quality can be characterised by:

- Data with as little as possible error
- Data that satisfies the users needs

Improvement of punctuality and use of punctuality data

Because customers value punctuality highly and actual punctuality level is below the target for many rail services, there is a need to work with improvement of punctuality. The probably most well-known model for quality improvements is the "Deming- circle" which consists of four steps: Plan, do, check and act (Shiba et. al., 1993). To use the model, at least two types of information are required: information about what causes the situation one want to improve and information about if the chosen improvement measures actually give the expected and wanted effects. The first requirement demand knowledge about cause and effect relationships while the latter can be solved by developing indicators that measure the effect of the measures carried out.

Improvement work in the railway industry has traditionally been based on practical experience (Veiseth, 2002). Practical experience is valuable and necessary, but it is also important to base the improvement work on data analyses and facts (Aune, 1997). When it comes to improvement of punctuality there are especially two sets of data which is used for this purpose: registrations of delays and registrations of what causes the delays. In most countries there has been a development the last decades when it comes to the first set of data. More and more of the registrations are now

collected automatically through the infrastructure signal systems and stored in databases. This has resulted in a larger volume and improved reliability of data.

When it comes to the second set of data, the delay cause registrations, they are still registered manually. In several countries there is disagreement about the accuracy and completeness of these registrations (Lundin et. al., 2005, Veiseth and Bititci, 2006). An important aspect here is that the personnel that do the registration neither have full access to information about the root causes of the delays nor have they enough time to investigate the root causes closer, especially not in hectic periods. One result is that the cause category “miscellaneous” often is one of the largest in many punctuality statistics (Jernbaneverket, 2006 and Veiseth, 2002).

Delay causes and infrastructure faults

Punctuality statistics tell that infrastructure faults are a major source for delayed trains. Table 1 shows how the total amount of delay hours registered in Norway in 2005 are distributed between the three main cause categories: infrastructure, operators and miscellaneous and accidents:

Cause category	Delay hours
Infrastructure	4118 hours
Operators	5280 hours
Miscellaneous and accidents	4728 hours

Table 1: Delay hours registered in Norway in 2005 distributed on main cause categories (Source: Jernbanverket, 2005)

The table shows that more than 4000 delay hours was caused by infrastructure conditions, which is about 30 % of the total amount of delay hours. This number can be split into five subcategories: track, signal (includes safety and communications systems), power supply, planned work and blocked tracks (Table 2)

Cause category	Delay hours
Track	552 hours
Signal	1911 hours
Power supply	624 hours
Planned work	753 hours
Blocked track	278 hours

Table 2: Infrastructure related delay hours registered in Norway in 2005 distributed to cause categories (Source: Jernbanverket, 2005)

This paper focuses on infrastructure faults, and we define infrastructure faults as “faults related to the track, the signal, safety and communication system and the power supply system”.

Increase the quality of punctuality data by linking operational and infrastructure data

Different strategies can be followed to improve the quality of punctuality data. One could be to improve the accuracy of the delay causes registrations. Veiseth and Bititci (2006) suggest that one way to achieve more accurate registrations could be to make the registration-system more transparent, to motivate and train the people that record and register the information and to develop indicators that measure the quality of these data.

Another way to improve the quality of punctuality data could be to use additional data sources, e.g. put together information from different databases. The definition of data quality state that an important element is that the data should be suitable for its use (Redman, 2005). Important users of punctuality data are personnel and projects that tries to improve punctuality. We therefore argue that quality of punctuality data will increase if we manage to increase the possibilities of root-cause analyses and make estimation of the operational effects of these causes possible.

Infrastructure databases include detailed information about reported delay causing incidents and information about components and faults history. We therefore think it should be possible to link this information with operational data from punctuality databases to identify root causes of delays and to estimate effects of delay causing incidents (Figure 1).

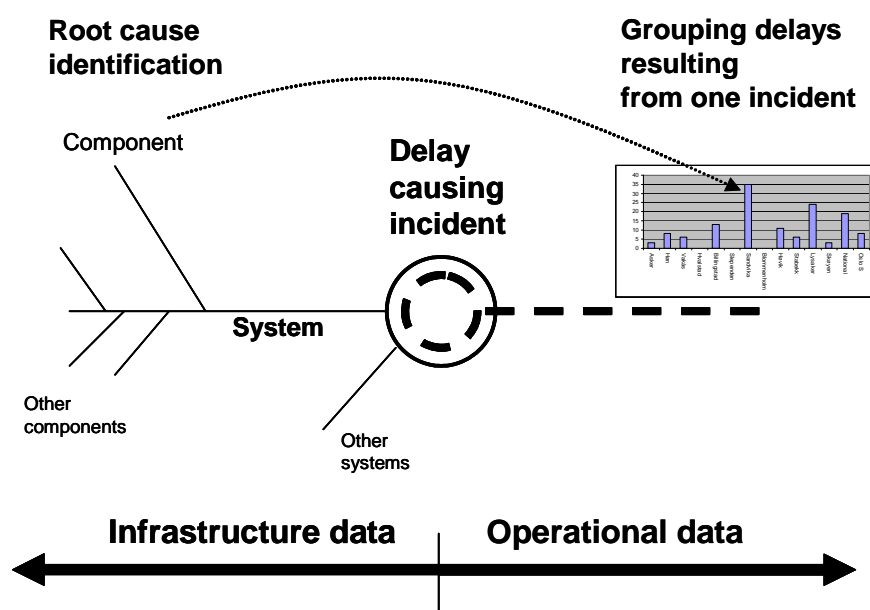


Figure 1: Root cause identification of delay causing incidents and estimation of effects by linking infrastructure data and operational data

Methodology

In our study we have investigated the possibility of root cause analyses and estimation of effects of selected infrastructure faults by linking information from several data sources. For this purpose we have analysed and compared data from two infrastructure databases and one punctuality database in Norway:

- TIOS: the punctuality data base which include registered delays and delay causes.
- BMS: Infrastructure database which contains information of all reported incidents regarding the infrastructure.
- Banedata: Infrastructure database that contains information of all infrastructure components and fault history.

We have first attributed delay minutes/hours to the lines and parts of lines by using information from the TIOS database. We have then investigated the possibilities of linking information from the three data sources to estimate how many minutes of delay a selected infrastructure fault has generated. The linking should theoretically be possible because all the three databases are

supposed to hold information about the same delay causing incidents. It should also be technically possible because they all contain information about time and location. To analyse the possibilities we have carried out the following tasks:

1. Investigated the possibility of attributing delay minutes/hours to different lines and part of lines by using TIOS data.
2. Compared number of registrations in the three databases for specific lines and period of times.
3. Analysed match of data from the three databases.

Based on the results we have discussed the possibility for automatic linking of data between the databases. The three databases and the type of information they contain are described below in addition to a short introduction to the Norwegian railway network. How the analyses are carried out, together with the results, are then presented.

Punctuality database (TIOS)

TIOS means: “traffic-information and follow-up system” and is developed and owned by Jernbaneverket³. TIOS collects information regarding delays at different places along the lines and stores this in a database. In addition, the delays causes are manual registered directly into the database. The intention is that for all trains that become more than four minutes delayed between to registration points (e.g. between two station), the delays causes are registered by a set of codes. From the database it is possible to extract different sorts of reports, for example punctuality reports for different regions, train services and lines. The following information can be found in the TIOS databases regarding delay causes: Cause code, date, train number, station, cancelled (yes/no), delay minutes and comments.

Rail infrastructure data (Banedata)

Banedata is an extensive system for information regarding Jernbaneverkets infrastructure. It includes information about all equipments and components to support activities such as maintenance planning. It also includes information about incidents related to the infrastructure, for example infrastructure faults. All registered incidents are connected to specific equipment or components. Key-data from Banedata includes all registered faults distributed to the different lines. You also find information about if the incident affected the train traffic or not. The following information can be found in Banedata related to delay causes: Description and classification of incident, object, location, time, period the incident affected the train traffic and action carried out.

Rail messages centre (BMS)

BMS contains systemized data of all reported messages of incidents regarding the infrastructure. The type of information in BMS is therefore similar to the data in Banedata, but because BMS does not have to attach all incidents to specific equipment or components, more incidents are registered here compared to Banedata. BMS also holds information about if the reported incidents affected the train traffic or not. The following information can be found in BMS which is related to delay causes: Description and classification of incident reported, type of object and responsible unit, location, time, affected the train traffic (yes/no), status and priority and action carried out.

The Norwegian railway network

The Norwegian railway network consists of 4.077 km railway line, of which 38% is not electrified (based on information on www.jernbaneverket.no). A special characteristic of the Norwegian rail network is the relatively low share of double tracks. Only 5% of the network consists of double tracks and all double tracks are located close to Oslo. All long distance, regional and freight trains mainly run on single tracks. Many commuter trains in the Oslo are run on both single and double tracks. Figure 2 shows an overview of the Norwegian railway network.

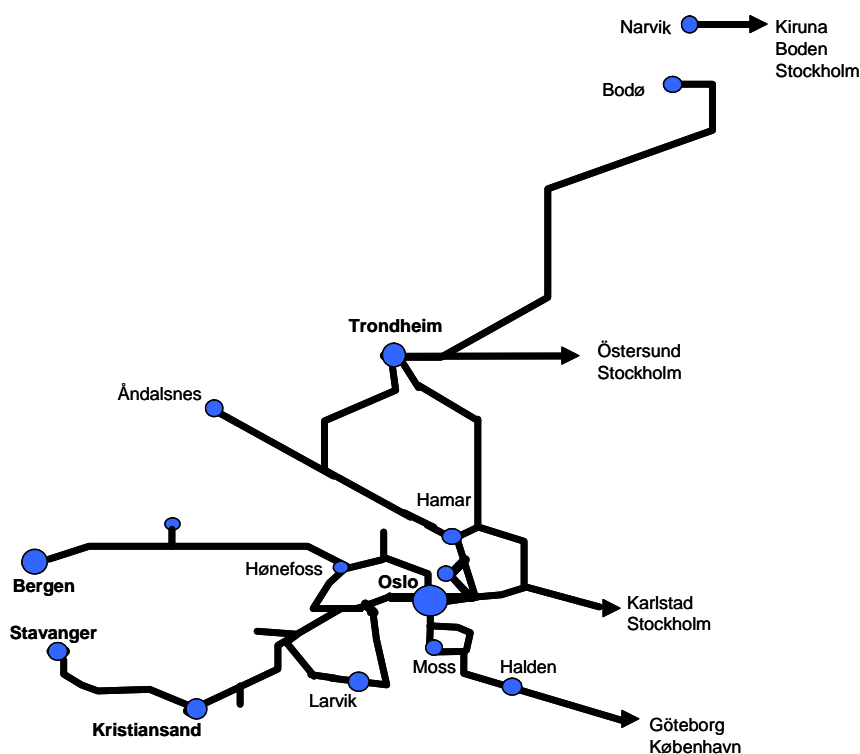


Figure 2: Norwegian railway network

Attributing delay minutes/hours to different lines and part of lines

From punctuality statistics (which is based on TIOS data), we find that 3086 hours and 22 minutes of delays was caused by infrastructure faults in Norway in 2005. Table 3 shows how these are distributed on the three Norwegian railway regions: East, West and North.

Region	Delay hours (hours: minutes)	Percent (%)
East	2263:24	73
North	239:28	8
West	583:30	19

Table 3: Delay hours caused by infrastructure faults registered in Norway in 2005 distributed to regions (Source: Jernbaneverket, 2005)

The table tells that 73% of all registered delay hours in 2005, caused by infrastructure faults, occurred in the East region. This might not be surprising because a large share of the train

production in Norway is carried out in this region. How these delay hours are distributed on different lines in the eastern region is shown in Figure 3

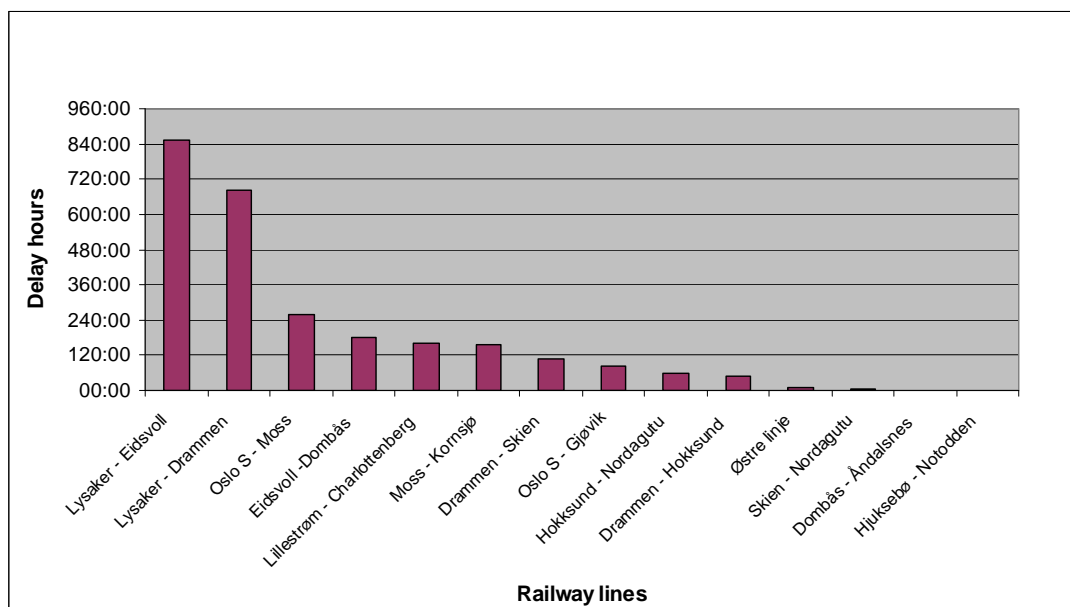


Figure 3: Delay hours registered in region East caused by infrastructure faults distributed to lines. Data from 01.01.2005 to 31.12.2005 (Source: TIOS database)

The figure tells that a large amount of the delay hours are registered on two lines: the Eidsvoll – Lysaker line and the Lysaker – Drammen line. Total 1538 delay hours are registered at these lines, which is about 60 percent of all registered delay hours caused by infrastructure faults in region East in 2005.

Because TIOS contains information about where the delay has occurred, represented by names of station, it is possible to distribute the delay hours for a line to different parts of the line. The delays registered at a station means that the delay has occurred at the station or between the station and the previous measure point. To investigate how the delay hours are distributed along a line we have analysed data from one of the busiest railway sections in Norway; the Oslo – Asker line. (The line goes from Oslo Central Station to the Asker station and is included in the two lines that have most delay hours in the figure above). We have used data from 01.06.2005 to 31.05.2006 for this purpose, and the result is presented in Figure 4.

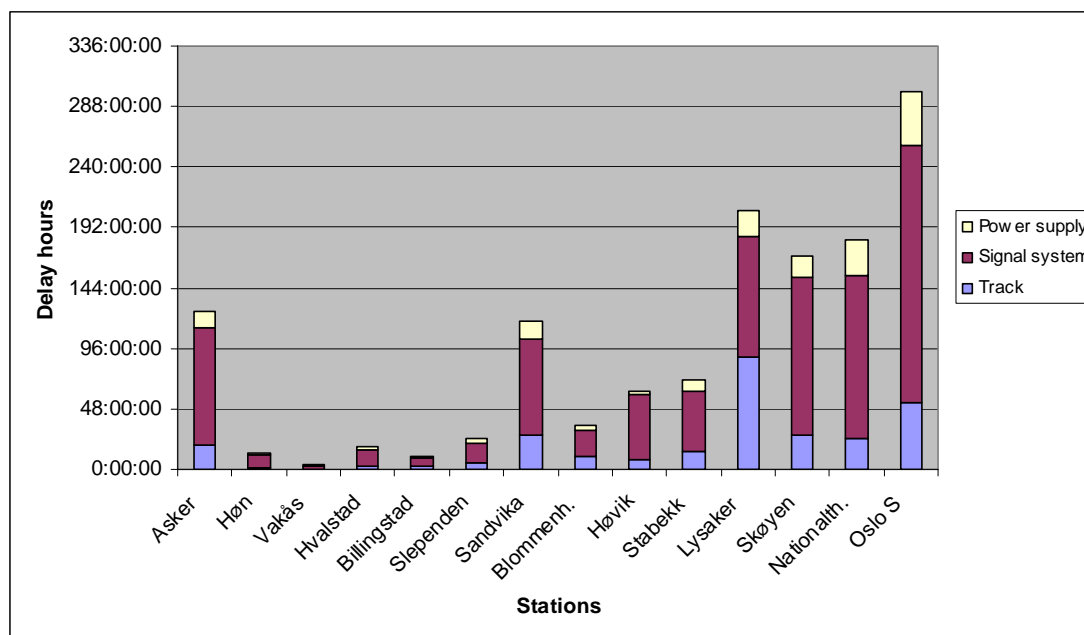


Figure 4: Delay hours caused by infrastructure faults for the Asker – Oslo line, distributed to different parts of the line and cause categories. Data from 01.06.2005 to 31.05.2006 (Source: TIOS database)

To check the quality of these results we have analysed delay causes registered at three different stations in detail. For some of the registered delay causes it is possible, through the comment field in the TIOS data-base, to check if a delay cause registered at a station actually occurred at this station (or between the station and the previous measure point), or if it occurred somewhere else. For the selected stations we have compared:

- The number of delay causes registered at the given station
- The number of these delay causes that for sure did not occurred at this station or between the station and the previous measure point.

The result is presented in Table 4.

Cause category	Station	Delays causes registered (number)	Delay causes actually occurred other places (number)
Signal	Høn	96	45
Track	Oslo	319	83
Power supply	Sandvika	92	47

Table 4: Comparison between the number of delay causes registered at given stations and delay causes that for sure occurred other places. Data from 01.06.2005 to 31.05.2006 (Source: TIOS database)

Based on the results we conclude that several of the delay causes registered at a specific station has not actually occurred at this station or between the station and the previously measure point. It is therefore not possible to only use data from the TIOS database to distribute delay causes to different sections of a line. For some of the registrations it is possible to find out where the delay cause occurred through the information in the comment field, but not for all. This is because a comment is missing for many of the registered delays causes and the comments that do exist do

not always contain information about location. Furthermore, the comment-filed is text-based which means one has to go through the data by hand and make subjective judgments.

Comparison of number of registrations in the three data bases

To compare the number of registrations in the three databases we have analysed two cases:

- The Asker- Oslo line (local trains, Inter City and long distant trains).
- The Bergen line from Hoenefoss to Bergen (long distant trains).

For both cases we have compared the number of registrations related to infrastructure faults for week 36 in 2006. Table 5 and Table 6 summarize the findings. All data is number of registered delay causes (TIOS) or registered infrastructure faults (BMS and Banedata). The number in the parentheses indicates the number of registered faults in BMS and Banedata which is said to affect the operations.

	TIOS	///	BMS	Banedata
Track	0	///	2(1)	0
Signal	22	///	29(3)	2(0)
Power supply	9	///	2(1)	0
Total	31	///	33(5)	2(0)

Table 5: number of registered delay causes related to infrastructure faults (TIOS) and number of registered infrastructure faults (BMS and Banedata) for the Oslo- Asker line, week 36, 2006. The numbers in parentheses indicate the number of registered faults in BMS and Banedata which is said to affect the operations.

	TIOS	///	BMS	Banedata
Track	7	///	6(0)	0(0)
Signal	0	///	8(1)	4(0)
Power supply	3	///	4(1)	1(0)
Total	10	///	18(2)	5(0)

Table 6: number of registered delay causes related to infrastructure faults (TIOS) and number of registered infrastructure faults (BMS and Banedata) for the Bergen line between Hoenefoss and Bergen, week 36, 2006. The numbers in parentheses indicate the number of registered faults in BMS and Banedata which is said to affect the operations.

It is important to notice that it is not possible to directly compare all the numbers in the table above. TIOS gives information about the number of times trains have become delayed because of infrastructure faults, while BMS and Banedata give information about number of (reported) infrastructure faults. This means that one registered fault in Banedata or BMS can correspond to more than one registered delay cause in the TIOS database. But still, the analyses give useful information about how the registrations in the three data bases correspond. It tells that:

- There is a mismatch between registered infrastructure faults in Banedata and BMS
- More infrastructure faults are registered in BMS compared to Banedata
- There are probably more faults which have affected the operations than indicated in BMS and especially in Banedata.

Match of data from the three databases

To check the possibility of linking information between the three databases we have analysed how many of the registered delay causes in TIOS which match registrations in Banedata and BMS. For this purpose we have used the two sets of data described in the previous chapter: Registrations for the Bergen line from Hoenefoss to Bergen for week 36, 2006 and registrations for the Asker – Oslo line for the same period of time. Below is an example of how the analysis is carried out presented.

For the Asker –Oslo line there are 9 delay causes registered as “power supply” in week 36, 2006. One of them is regarding train number 2151 (Table 7).

Cause-code	Date	Train number	Station	Cancelled? (Yes/No)	Delay minutes	Comments
3 (Power supply)	2006-09-09	2151	Hvalstad	No	00:04	Tree fell over cable at Hvalstad

Table 7: Data from TIOS regarding delay cause for train number 2151, 09.09.2006.

In BMS we find the following information that seems to match the information in TIOS (Table 8):

Field	Information
Number (ID)	72379
Time	2006-09-09 14:28, 0 days and 3 hours
Status	Repaired, 1. priority
Classification	Power supply
Responsible	Power supply, Asker
Description	Tree over cable at Hvalstad – Billingstad in the Drammen-Oslo track. Actions: 09-09-2006 17:47 by NN, Bear-cable was off and some of the power bridges, wires was off in both main tracks.
Location	Hvalstad (Station), 20.190 km on [1410] (Lysaker) - Asker
Type of object	-
Did affect the operations?	Yes

Table 8: Data from BMS that seems to match the data from TIOS regarding delay cause for train number 2151, 09.09.2006

In BaneData we do not find any information for week 36 that correspond with the information found in TIOS and BMS. The conclusion is then: match found in BMS but not in Banedata. Table 9 and Table 10 summarize the results from the analyses:

Category	TIOS (number of registered delay causes)	BMS (number of matches with TIOS)	Banedata (number of matches with TIOS)
Track	0	-	-
Signal	22	3	0
Power supply	9	9	0
Total	31	12	0

Table 9: Match of data from the three databases for the Asker – Oslo line. Data from week 36, 2006

Category	TIOS (number of registered delay causes)	BMS (number of matches with TIOS)	Banedata (number of matches with TIOS)
Track	7	0	0
Signal	0	-	-
Power supply	3	2	0
Total	10	2	0

Table 10: Match of data from the three databases for the Bergen line between Hoenefoss and Bergen. Data from week 36, 2006

Based on the results we conclude that:

- We have not found a good match of data from the three databases
- Most matches can be found for delay-causes registered as “power supply”.
- There are more matches between data registered in BMS and TIOS compared to Banedata and TIOS

The results indicate that the registrations in BMS, and especially Banedata, are insufficient. We therefore asked the owners of the two data systems about this. They told us that not all incidents were reported in and registered in the systems, and the number of not registered incidents varied over time and between lines. This means that if we had studied other lines and other time periods we could have got better results.

Automatic linking of information between the three data bases

Linking information automatically between the three data sources should be possible because the data systems are meant to register information about the same incidents. In addition all the three databases holds information about time and location and if the incidents affected the train traffic or not. It should therefore also be possible to link the information, technically. Yet, our analyses have demonstrated that this is not possible. There are especially two main-reasons for this:

The first reason is that delay causes registered in TIOS are not always registered at the location where they occurred but where the effect of the delay causes occurred. For some of the registrations it is possible through the comment field to check where the delay cause occurred. But this information is text based which means it is difficult to use it for automatic linking of information. Furthermore, many of the registered delay causes lack a descriptive comment and not all comments include information about location. The second reason is that the registrations in BMS, and especially Banedata, seems to be insufficient. Many of the delay causes registered in TIOS can not be found in the two other databases, especially when it comes to delays caused by signal and track faults. Also important is that several of the registrations in BMS and Banedata which is marked as a fault that did not affect the operations, obviously did affect the operations.

The time registration in BMS and Banedata is also a challenge. In these two data bases a fault is registered when it occurred. In addition it is possible to find information about how long time it took before the fault was repaired. But it is only possible to use the first information when you search in the databases. If you search out information for the same period of time in the three

databases, you risk that some of the registered delays causes in TIOS are results of faults which has occurred, and is registered, outside the chosen period of time in BMS and Banedata.

To achieve automatic linking we think it is necessary to develop a general code-system for all three databases that directly tie the information together. It could be designed in such a way that each infrastructure fault reported in BMS was given a unique number, where the same number is used when the fault is registered in Banedata. If the personnel that do the registrations of delay causes in TIOS are given access to this information, they could use the same codes when they register the delay causes. In addition, the completeness of registrations in BMS and Banedata has to improve if linking of information should be possible.

Conclusion

This paper gives input to how it could be possible to identify root causes for delays caused by infrastructure faults and to estimate the effects of these faults by linking information from several data sources. In our study we have analysed data systems used in the Norwegian railways, but the findings may also be useful for railway industries in other countries. This because we assume that similar databases to those we have studied does exist in most countries. Linking of information between the punctuality database and the infrastructure databases should be possible, technically, but our analyses have demonstrated that this is not possible. This is due to that delay-causes registered in the punctuality database are not always registered at the location where they occurred but often where the effect of the delay-cause occurred. Another reason is that the registrations in the infrastructure databases seem to be insufficient. To achieve automatic linking of information between the data sources we recommend the railway industry to develop a general code-system for all three databases that directly tie the information together.

References

Aune A. (1996) *Kvalitetsstyrte bedrifter (Total quality management)*, Ad Notam Gyldendal, Oslo, Norway

Capiello C, Francalanci C and Pernici B (2004) 'Data Quality Assessment from the User's Perspective'. *Proceedings of the 2004 international workshop on Information Quality in Information Systems*, pp. 68-73

Jernbaneverket (2006) *Punctuality report 2005*, Oslo, Norway

Jernbaneverket (2005) Norwegian railway administrations' punctuality statistics from 2005, based on TIOS data.

Lundin M., Elmquist, A.L. and Hammarlund S. (2005) '*Incitament för ökad punktlighet på järnväg – fortsättningsprojekt (Incentives for improved punctuality)*', TFK report 2005:3, Borlänge, Sweden.

NEA Transport research and training (2003). 'BOB railway case – benchmarking passenger transport in railways – final report', Rijswijk, The Netherlands

Olsson, N.O.E. and Haugland, H., (2004): 'Influencing factors on train punctuality – results from some Norwegian studies', *Transport Policy*, **11** (4), pp. 387-397

Redman T.C. (2005) *Data Quality - the field guide*, Digital Press, USA, p.73

Rudnicki, A. (1997) 'Measures of regularity and punctuality in public transport operation', *Transportation system, Proceedings volume from the IFAC symposium, Chania, Greece 16-18 June 1997*, Vol. 2, pp 661-666

Shiba, S., Graham, A. and Walden D. (1993) *The New American TQM, Four Practical revolutions in Management*, Productivity Press, Portland, Or

Veiseth M. and Bititici U.S. (2005) *Performance measurement in railway operations – improvement of punctuality and reliability*, Proceedings of the PMA conference, London, UK, July, 2005

Veiseth, M. (2002) 'Punctuality in railway operations', *Master theses, Norwegian University of Science and Technology*, Trondheim, Norway