

Surrogatbaserad optimering av avgiftsnivåer i trängselavgiftssystem

Rapport

2015-09-30

Joakim Ekström

Kommunikations- och transportsystem, Institutionen för teknik och naturvetenskap
Linköpings universitet, SE-601 74 Norrköping
E-post: joakim.ekstrom@liu.se

Ida Kristoffersson

Trafikanalys, Sweco TransportSystem AB
Gjörwellsgatan 22, Box 340 44, SE-100 26 Stockholm
E-post: ida.kristoffersson@sweco.se

Nils-Hassan Quttineh

Optimeringslära, Matematiska institutionen
Linköpings universitet, SE-581 83 Linköping
E-post: nils-hassan.quttineh@liu.se

Slutrapportering av projektet ”Optimala vägavgifter med dynamiska modeller” som genomförts på uppdrag av Trafikverket (TRV 2013/19714)

Sammanfattning

Trängselskatt finns idag i både Stockholm och Göteborg, och utformningen av dessa trängselskattesystem kommer att justeras framöver med avseende på avgiftsnivå, placering och tidpunkt. För Stockholm finns beslut om ändring från januari 2016 och i Göteborg ändrades avgiftsnivåerna i januari 2015. I detta projekt utvecklas metoder som ska kunna ge stöd vid justering av avgiftsnivåer, så att en så stor samhällsekonomisk nytta som är möjligt uppnås med trängselskattesystemet.

För storstadsområden, där det under rusningstrafik är trängsel i delar av nätverket, är trängselskatt främst intressant att analysera med dynamiska transportmodeller. Tidigare utveckling av metoder för optimal avgiftssättning har dock främst fokuserat på statiska modeller, exempelvis Emme, som har kända problem med att korrekt uppskatta förändring i restider när det är trängsel i delar av trafiknätverket. I detta projekt har vi därför tillämpat surrogat-baserad optimering, som är en metodansats som ställer få krav på vilken transportmodell som används. Den dynamiska transportmodellen Regent/VisumDUE finns sedan tidigare implementerad för Stockholmsregionen, och har därför även använts i detta projekt. VisumDUE är en makroskopisk nätutläggningsmodell med dynamiskt ruttval, och Regent är en efterfrågemodell som innehåller resgenerering, färdmedelsval och destinationsval för arbetsresor¹.

Surrogat-baserad optimering erbjuder ett ramverk för optimering av problem med beräkningsmässigt kostsamma målfunktioner. Genom att approximera en funktionsyta till samplade punkter från den kostsamma målfunktionen, kan optimeringen istället göras över den approximerade funktionsytan. För Regent/VisumDUE tar utvärderingen av ett givet trängselskattescenario ca tio timmar, och det är denna beräkningstid som gör målfunktionen kostsam. Givet ett antal samplade punkter, görs ytterligare sampling utifrån en given strategi för att förbättra approximationen, så kallad iterativ sampling. Inom ramverket finns dock en mängd möjligheter för hur de olika komponenterna designas. Därför är det svårt att utvärdera surrogat-baserad optimering med endast Regent/VisumDUE. En statisk transportmodell har därför använts för att utvärdera ett antal kombinationer av samplingsstrategi och funktionsyta. Den mest lovande kombinationen har sedan även utvärderats med Regent/VisumDUE. För att vara praktiskt tillämpbart i framtiden har fokus i projektet varit att utvärdera hur metodansatsen fungerar när antalet möjliga tulluppsättningar är kraftigt begränsat (20-40 stycken).

De tullsnitt som har använts som grund i projektet är trängselskatt i Stockholm på nuvarande tullring, på Essingeleden samt på innerstadsbroarna. Skatten är differentierad med avseende på riktning, vilket ger sex olika skattenivåer att optimera. Optimeringen har gjorts för trängselskattnivå under maxtimmen. I det dynamiska fallet har trängselskattens nivå utanför maxtimme funnits med som indata, men samma tidsprofil som på nuvarande tullring har antagits i alla scenarier (avgiftstrappa 50%, 75%, 100%, 75%, 50%). Utvärderingen med den statiska transportmodellen visar att lösningar nära globalt optimum kan uppnås med endast 40 utvärderade trängselskattnivåer, och en tydlig förbättring av den samhällsekonomiska nyttan uppnås redan vid 20 utvärderade trängselskattnivåer.

¹ Detta gäller den version av Regent som integrerades med VisumDUE under 2013. Sedan dess har Regent utvecklats och innehåller nu efterfrågemodeller för både arbetsresor och övriga resor.

Även med ett kraftigt begränsat antal utvärderingar av den kostsamma målfunktionen i Regent/VisumDUE, har vi visat att det är möjligt att använda metodansatsen. En tydlig förbättring av den samhällsekonomiska nyttan uppnås med endast 22 utvärderade trängselskattenivåer. Ytterligare experiment skulle dock behövas för att undersöka hur stor denna förbättring är i förhållande till vad som skulle kunna uppnås.

Innehållsförteckning

1. Introduktion.....	1
1.1 Bakgrund.....	1
1.2 Syfte.....	2
2. Val av optimeringsmetod och transportmodell	2
2.1 Begränsningar med statisk modell	2
2.2 Metoder för optimering av avgiftssystem med dynamiska transportmodeller.....	3
2.3 Val av optimeringsmetod	4
2.4 Tillgängliga dynamiska transportmodeller	4
2.5 Val av dynamisk transportmodell.....	5
3. Surrogatbaserad optimering av trängselavgifter	5
3.1 Surrogatbaserad optimering	5
3.2 Den kostsamma målfunktionen	7
3.3 Surrogatmodellen.....	7
3.4 Experimentdesign.....	9
3.5 Samplingsstrategi	9
3.6 Implementering.....	11
4. Utvärdering av metodansats	11
4.1 Tillvägagångssätt för utvärdering av metodansats	11
4.2 Utvärdering med den statiska transportmodellen.....	12
4.3 Utvärdering med VisumDUE och Regent	14
4.4 Utvärderingsmått	17
5. Resultat.....	17
5.1 Statisk modell	17
5.2 VisumDUE med Regent	20
6. Diskussion.....	22
7. Slutsatser	23
Referenser	24

1. Introduktion

1.1 Bakgrund

Trängselskatt har de senaste åren införts både i Stockholm och Göteborg med minskning av biltrafikflöden och kortare köer som följd (Eliasson m.fl. 2009; Börjesson m.fl. 2012; Börjesson och Kristoffersson 2015). I Göteborg höjdes avgiften från januari 2015 och från 2016 finns beslut om att även höja avgiften i Stockholm samt införa trängselskatt på Essingeleden (Sveriges Riksdag 2014). Det är troligt att trängselskatten i Stockholm och Göteborg kommer justeras ytterligare framöver med avseende på avgiftsnivå, placering, tidpunkt m.m. För att veta mer om vilka justeringar som ger störst samhällsekonomisk nytta behöver nya metoder utvecklas. I Sverige är den avgift som tas ut i Stockholm och Göteborg en skatt, men internationellt är det vanligare att det betraktas som en avgift, och därför kommer vi fortsättningsvis att använda benämningen trängselavgift.

En tumregel är att trängselavgift gör mest nytta om den tas ut vid flaskhalsar, så att köer uppströms minskar och det kringliggande vägnätet avlastas. Att designa ett trängselavgiftssystem är dock svårt, främst på grund av risken att flytta köer – i tid och rum – när bilisterna ändrar sitt beteende för att undvika trängselavgiften. Stockholm, med dess flaskhalsar på broarna runt innerstaden, var ett relativt enkelt designobjekt, medan ruttvalsproblematiken var större i Göteborg. Inför införandet av trängselavgift i Stockholm och Göteborg har experter kallats in, vilka har tagit fram några olika designförslag och utvärderat dessa med hjälp av transportmodeller. Det har dock inte funnits en övergripande metod, som ser till att designen av trängselavgiftssystemet går mot det system som ger störst samhällsekonomisk nytta.

Designen av de svenska trängselavgiftssystemen baserades på prognoser med den nationella transportmodellen Sampers, vilken i dagsläget använder den statistiska nätutläggningsmodellen Emme för ruttval och beräkning av utbudsdata så som restider i biltrafiken. Det har visat sig att prognoser med Sampers gett en tillräckligt bra bild av hur stora *flödesminskningar* man kan räkna med vid införandet av trängselavgift, men att *restidsminskningarna* underskattats stort (Eliasson m.fl. 2013). För att göra mer realistiska prognoser av restidsminskningar och därmed den samhällsekonomiska nyttan behöver dynamiska transportmodeller användas (Engelson och van Amelsfort 2011; Börjesson och Kristoffersson 2014).

Med statistiska transportmodeller (exempelvis Sampers/Emme) kan optimala avgiftsnivåer i ett trängselavgiftssystem som maximerar den samhällsekonomiska nyttan (s.k. marginalkostnadsprissättning) enkelt beräknas, givet att samtliga vägavsnitt kan avgiftsättas. Även om antagandet om avgiftssättning, med differentierade avgifter på samtliga vägsträckor, är orealistiskt ur ett praktiskt perspektiv är marginalkostnadsprissättande trängselavgifter intressanta att beräkna. Marginalkostnadsprissättande trängselavgifter resulterar nämligen i ett systemoptimum, som ger en viktig övre gräns för hur stor samhällsekonomisk nytta som det är möjligt att uppnå med ett trängselavgiftssystem. De ger även viktiga insikter i hur ett bra trängselavgiftssystem kan utformas i verkligheten och det finns utvecklade metoder för att givet en marginalkostnadsprissättning lokalisera trängselavgifter och bestämma avgiftsnivåer på ett fåtal vägavsnitt. Även under antagande att det finns restriktioner på var avgifter tas ut, exempelvis endast vid förutbestämda tullportaler, finns utvecklade metoder att optimera både nivå på tullar och lokalisering av tullportaler, se exempelvis Ekström et al. (2014). Dessa metoder bygger vanligen på heuristiska algoritmer som kräver att ett stort antal olika designer för trängselavgiftssystemet utvärderas, eller på algoritmer

som utnyttjar analytiska samband i den statistiska transportmodellen för att förutsäga hur trafiken påverkas vid en liten förändring av en avgiftsnivå (derivata-information).

Tyvärr har statistiska transportmodeller en begränsad möjlighet att korrekt uppskatta köbildning och restider på vägavsnitten, vilket ger en stor osäkerhet i de avgiftsnivåer som beräknas. Dynamiska nätverksmodeller – exempelvis VisumDue, Transmodeller, Contram, Metropolis, Dynameq och Mezzo² – ger istället en avsevärt bättre möjlighet att beskriva hur köer breder ut sig i ett trafiknätverk både i tid och i rum, samt uppskattar restider på enskilda vägavsnitt med större noggrannhet. Dock saknar de flesta dynamiska nätverksmodeller de analytiska samband som gör det möjligt att enkelt bestämma optimala trängselavgifter. Det är också ovanligt att en dynamisk nätverksmodell är kopplad till en efterfrågemodell, vilket är problematiskt då trängselavgifter generellt har stor påverkan på efterfrågan, t ex. överflyttning till kollektivtrafik.

1.2 Syfte

Det övergripande syftet med projektet är att undersöka om tidigare utvecklade metoder för optimering av trängselavgiftssystem kan appliceras på en, för svenska förhållanden, intressant dynamisk transportmodell. Vi begränsar oss här till transportmodeller som idag finns utvecklade och implementerade för Stockholm, samt till att studera trängselavgiftssystem som bygger på passager över tullringar/tullsnitt. Samt om en sådan metodansats kan användas för att förbättra den samhällsekonomiska nyttan med ett trängselavgiftssystem.

Detta syfte kan brytas ner till följande aktiviteter i projektet:

1. Presentera en översikt över forskningsläget kring metoder för optimering av trängselavgiftssystem som kan användas tillsammans med dynamiska transportmodeller.
2. Utvärdera lämpligheten hos idag använda dynamiska transportmodeller att användas vid optimering av trängselavgiftssystem.
3. Tillämpa vald optimeringsmetod med vald dynamisk transportmodell och utvärdera resultatet med avseende på förbättrat samhällsekonomiskt överskott.

2. Val av optimeringsmetod och transportmodell

2.1 Begränsningar med statisk modell

Trängselavgifter som maximerar den samhällsekonomiska nyttan har främst studerats med statiska modellansatser, där varken köutbredning i tid eller rum studeras. Standardanalysen av optimala trängselavgifter med denna typ av modeller ger att bilförare som grupp upplever en negativ nytta när trängselavgifter införs om inte intäkten återförs till denna grupp. Börjesson och Kristoffersson (2014) visar dock att om man tar hänsyn till dynamiska effekter (både med avseende på effekter från köer som sprids i ett trafiknätverk och effekter på efterfrågan till följd av förändrat val av avresetid) och variation i tidsvärden, kan gruppen bilresenärer uppleva en positiv nytta från införandet av trängselavgifter, även om de inte får del av intäkterna. Liknande brister i den statistiska analysen har påvisats av Engelson and van Amelsfort (2011) och Kristoffersson (2013), som tydligt visar att det nationella statistiska modellsystemet som används i Sverige (Sampers) kraftigt undervärderar nyttan från införandet av trängselavgift i Stockholm, jämfört både med vad som har uppmätts men också jämfört beräkningar från dynamiska transportmodeller.

² De uppräknade dynamiska modellerna finns alla implementerade för Stockholm med varierande kvalitet i kalibreringen.

Huvuddelen av den forskning som skett kring optimal utformning av trängselavgiftssystem har utgått från användandet av statistiska transportmodeller för att modellera resenärernas förändrade resmönster. Det finns främst tre anledningar till detta:

1. För storstadsområden har det tills nyligen varit främst statistiska modeller som har funnits tillgängliga, och mer dynamiska modeller har främst använts för att studera mindre delområden av trafiknätverket.
2. I statistiska modeller finns matematiska samband som gör att effektiva optimeringsmetoder kan utvecklas, dessa samband saknas vanligen i dynamiska modeller.
3. Beräkningstiden är avsevärt högre för dynamiska modeller jämfört med statistiska modeller, vilket gör att optimeringsmetoder som kräver utvärdering av ett stort antal olika trängselavgiftssystem inte är lämpliga att använda tillsammans med dynamiska modeller.

2.2 Metoder för optimering av avgiftssystem med dynamiska transportmodeller

Publicerad forskning kring optimal utformning av trängselavgiftssystem, där trafikantbeteendet modelleras med dynamisk transportmodell, är begränsat till relativt få publikationer. I detta projekt har vi begränsat oss till metodansatser som ska kunna tillämpas med idag använda kommersiella dynamiska transportmodeller. Det vill säga att vi inte förväntar oss känna till detaljer om hur nätutläggning och ruttval sker. Dynamiska transportmodeller delas vanligen upp i kategorierna analytiska modeller och simuleringsbaserade modeller (Peeta och Ziliaskopoulos 2001). De analytiska modellerna utgår från analytiska samband mellan trafikflöde och restider, medan i simuleringsbaserade modeller ges restiden av simulerade interaktioner mellan fordon. I detta projekt har vi valt att fokusera på metodansatser som kan användas tillsammans med båda typer av transportmodeller. Givet dessa begränsningar finns i huvudsak tre olika metodansatser; flakshalsminimerande metoder, stokastiska sökmetoder och surrogatbaserade optimeringsmetoder.

Flaskhalsminimerande metoder (exempelvis de Palma m.fl. 2005 och Yang m.fl. 2010) utgår från tumregeln att trängselavgift gör mest nytta om den tas ut vid flaskhalsar, så att köer uppströms minskar och det kringliggande vägnätet avlastas. Problemet med denna ansats är dock att köer då istället flyttas i tid och rum. Dessutom är denna typ av metod inte är direkt tillämpningsbar för att bestämma avgifter på tullringar och tullsnitt om dessa tullar inte sammanfaller med flaskhalsarna i trafiknätverket.

Stokastiska sökmetoder, såsom meta-heuristiker (exempelvis genetiska algoritmer) har vanligen använts för att optimera både lokalisering av trängselavgiftsportaler, samt avgifternas nivåer, med statistiska transportmodeller. Genetiska algoritmer för både optimal lokalisering och avgiftssättning av trängselavgifter presenteras i Shepherd och Sumalee (2004) och en kombination av genetiska algoritmer (för lokalisering av tullportaler) och "simulated annealing" (för avgiftssättning) beskrivs av Zhang och Yang (2004). Grunden för flertalet av de stokastiska sökmetoder som används för optimering av trängselavgiftssystem är slumpvis manipulering av en existerande lösning. För en överblick av stokastiska sökmetoder som utvecklats för optimal utformning av trängselskattesystem, se exempelvis Ekström (2012). Gemensamt för dessa metoder är att ett stort antal olika lösningar behöver utvärderas. Detta begränsar dessa metoders möjlighet att tillämpas tillsammans med dynamiska transportmodeller, som ofta är beräkningsmässigt krävande.

Surrogatbaserad optimering används i de Palma m.fl. (2005) för ett mindre exempelnätverk i transportmodellen Metropolis, som inkluderar modell för att val av avresetid och byte till

kollektivtrafik, och i Chen et. al. (2014) för ett motorvägsnätverk över Maryland i USA med transportmodellen DynusT som modellerar efterfrågan att resa som fix. Dessutom använder Chow och Regan (2014) surrogatbaserad optimering för att bestämma optimala avgiftsnivåer i en statisk transportmodell, med särskild hänsyn till robusthet i transportsystemet.

I en surrogatbaserad optimeringsmetod samplas ett antal initiala trängselavgiftsnivåer och en surrogatmodell (responsyta) anpassas sedan till dessa punkter. Optimeringen kan sedan göras på surrogatmodellen. Vanligen kombineras surrogatbaserad optimering med metoder för att iterativt sampla ytterligare punkter.

2.3 Val av optimeringsmetod

Detta projekt fokuserar på metoder för optimering av trängselavgiftssystem som kan implementeras och användas i närtid, med idag (eller i en nära framtid) använda transportmodeller. Med detta fokus bedöms surrogatbaserad optimering som den mest intressanta optimeringsansatsen att gå vidare med. Motiveringen till detta val är att:

1. Surrogatbaserad optimering är en metodansats som kan tillämpas med många olika typer av transportmodeller, med det enda kravet att transportmodellen kan leverera indata till den samhällsekonomiska nyttoberäkningen.
2. Surrogatbaserad optimering kräver begränsad interaktion med transportmodellen, vilket gör den enkel att implementera.
3. Exempel på tillämpning kring optimering av trängselavgifter finns (Chen m.fl. 2014) med goda resultat.

Punkt 1 utgör det viktigaste motivet, då det i dagsläget är oklart vilken dynamisk transportmodell som i framtiden kommer att användas i Sverige vid utvärdering av effekter från trängselavgiftssystem. Surrogatbaserad optimering har dock en begränsning i dimensionen på optimeringsproblemet, och metoden kan förväntas fungera sämre för problem där många differentierade avgiftsnivåer ska optimeras, alternativt där både avgiftsnivå och plats för att ta ut avgifterna är variabelt. I detta projekt har vi begränsat antalet beslutsvariabler (differentierade tullnivåer i avgiftssystemet) till sex stycken.

2.4 Tillgängliga dynamiska transportmodeller

Den dynamiska transportmodell som i dagsläget är bäst kalibrerad för Stockholm är Contram. Denna modell har länge använts för att beskriva effekter av åtgärder som påverkar trängsel. I Contram ingår dock ingen efterfrågemodellering, vilket är en stor brist vid analys av trängselavgiftssystem. Både i Stockholm och i Göteborg var anpassningsstrategin för arbetsresor nästan uteslutande att byta till kollektivtrafik (Börjesson och Kristoffersson 2015). För att ta hänsyn till effekt på efterfrågan har därför Contram ofta tillämpats med efterfråge-matriser från Sampers som indata, men ingen feedback har gjorts. Denna Sampers-Contram-tillämpning blir inkonsistent då restider från Contram inte påverkar efterfrågan. För att utvärdera tidpunktsval har Contram även integrerats med efterfrågemodellen Silvester, men i denna ingår ett mycket förenklat färdmedelsval och både resgenerering och destinationsval saknas (Kristoffersson och Engelson 2009).

Ett parallellt forskningsprojekt kallat IHOP har som syfte att utveckla ett mer konsistent verktyg för storstadsanalyser än ovan beskrivna tillvägagångssätt med Sampers och Contram. Under våren 2015 kommer IHOP del2 påbörjas för att gå vidare med utveckling av storstadsmodell för vägtrafik. IHOP

del 1 utvärderade möjligheten att knyta samman kommersiellt använda dynamiska transportmodeller med en efterfrågemodell för Stockholm. Resultatet som vid start av detta projekt var tillgängligt från IHOP-projektet var två olika Stockholmsmodeller, en i VisumDUE och en i Transmodeller (Almroth m.fl. 2014). Båda dessa transportmodeller är kopplade med efterfrågemodellen Regent, och kan således användas för att ge indata vid beräkning av samhällsekonomisknytta vid införandet av ett specifikt trängselavgiftssystem. Dessa två transportmodeller utgjorde vid projektstarten de enda tillgängliga dynamiska modellerna för en svensk stad, där både förändringar i ruttval och efterfrågan beskrivs på ett tillfredställande sätt.

2.5 Val av dynamisk transportmodell

Som beskrevs i föregående avsnitt var VisumDUE och Transmodeller de enda dynamiska transportmodeller som fanns integrerade med en efterfrågemodell för svenska förhållanden vid projektets start. Det är därför naturligt att valet främst har stått mellan dessa två transportmodeller. Även om båda dessa modeller finns implementerade för Stockholm, har ingen av modellerna kalibrerats för att användas i verklig policyanalys. Båda dessa modeller är beräkningsmässigt krävande (10-40h för att utvärdera en uppsättning tullar).

Oavsett val av dynamisk transportmodell kan man konstatera att de för nätverk av realistisk storlek är beräkningskrävande. Surrogatbaserad optimering är en metodklass, och inom denna klass kan en mängd val göras för att utforma optimeringsmetoden för en specifik tillämpning. Att använda en dynamisk transportmodell för denna utvärdering begränsar antalet experiment, och därför kommer en statisk transportmodell att användas för en första utvärdering av den surrogatbaserade optimeringsmetoden.

Den dynamiska transportmodell som vi har valt att använda i projektet är VisumDUE. De främsta motiven bakom detta beslut är:

1. I projektgruppen finns erfarenhet från tidigare användning av VisumDUE. Tidigare erfarenheter av den transportmodell som används i projektet har bedöms som viktig för att kunna nå resultat inom ramen för projektets budget.
2. Erfarenheter från IHOP-projektet är att VisumDUE kräver mindre beräkningstid än Transmodeller.
3. VisumDUE är en transportmodell som bygger på makroskopiska samband mellan flöden och restider medan Transmodeller är en simuleringsbaserad transportmodell. Detta betyder att VisumDUE har större likheter med den statiska transportmodell som kommer att användas i projektet.

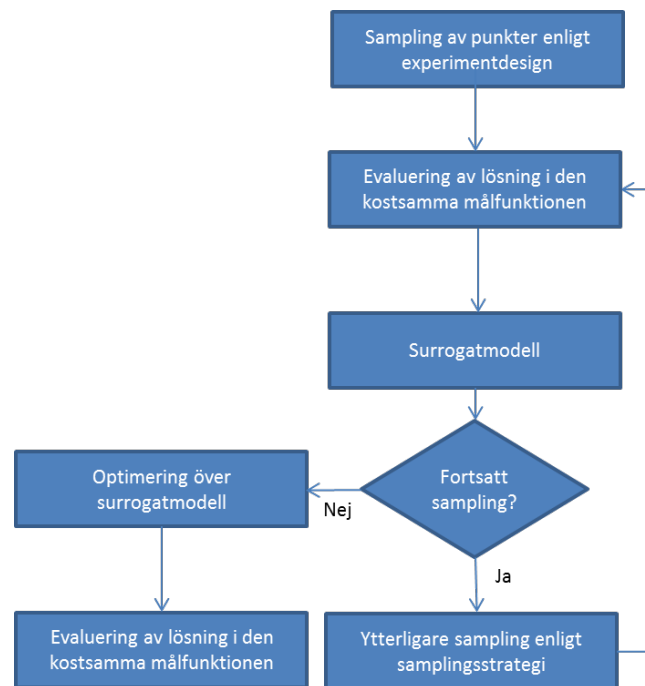
3. Surrogatbaserad optimering av trängselavgifter

3.1 Surrogatbaserad optimering

Surrogatbaserad optimering används för optimeringsproblem där utvärdering av målfunktionen är beräkningsmässigt kostsam. Att bestämma den samhällsekonomiska nyttan för en given uppsättning tullnivåer i ett trängselavgiftssystem kräver att en trafikmodell körs för att bestämma hur trafikflöden och restider förändras. Om trafikmodellen är beräkningskrävande kommer utvärderingen av en uppsättning trängselavgifter att vara kostsamt. Surrogatbaserad optimering använder dels den riktiga målfunktionen för att bestämma den samhällsekonomiska nyttan för ett mindre antal uppsättningar tullnivåer (så kallade samplade tullnivåer), dels anpassas en funktionsyta till de samplade tullnivåerna

som används för att approximera den samhällsekonomiska nyttan för de tullnivåer som inte utvärderas med den riktiga målfunktionen. Optimering av tullnivåer kan sedan istället göras över den approximerade funktionsytan.

Huvudkomponenterna för surrogatbaserad optimeringsmetod illustreras i Figur 1. Huvudkomponenterna är den kostsamma målfunktionen, den initiala samplingen av tulluppsättningar (hädanefter kallad för experimentdesign), den iterativa samplingen av ytterligare tulluppsättningar, samt den anpassade funktionsytan. Till skillnad från Chen et al. (2014), fokuserar vi i detta projekt på tillämpningar av surrogatbaserad optimering där det totala antalet samplade tullnivåer är kraftigt begränsat i antal.



Figur 1: Huvudkomponenter i en surrogatbaserad optimeringsmetod

Den anpassade funktionsytan används för att approximera den samhällsekonomiska nyttan för de tulluppsättningar som inte har utvärderats i den kostsamma målfunktionen. På så sätt erhålls en analytisk funktion som kan ersätta den kostsamma målfunktionen i en optimeringsmetod. Den anpassade funktionsytan kan dock mycket väl vara icke-konvex, och att bestämma optimum för denna yta är i sig ett svårt optimeringsproblem. Eftersom den anpassade funktionsytan är billig att utvärdera kan dock optimeringsmetoder som utvecklats för att lösa icke-konvexa problem användas. I detta projekt används genetiska algoritmer och "simulated annealing" i kombination, för att bestämma (nära) optimala tulluppsättningar för den anpassade funktionsytan.

Experimentdesignen bestämmer valet av initiala tulluppsättningar att utvärdera i den kostsamma målfunktionen. Både valet av antal initiala tulluppsättningar och de specifika tullnivåerna är en del av experimentdesignen. Efter att de initiala tulluppsättningarna har utvärderats i den kostsamma målfunktionen kan funktionsytan anpassas till dessa punkter. Vanligen använda funktionsuttryck är radiala basfunktioner (RBFer) (Gutmann 2001) och Krigingmodeller (Sacks m.fl. 1989; Jones, Schonlau, och Welch 1998).

Utöver de initiala tulluppsättningarna kan ytterligare punkter samplas för att iterativt förbättra approximationen av den kostsamma målfunktionen. Den iterativa samplingsstrategin beror vanligen på både i vilket område goda lösningar hittills hittats (lokal sökstrategi) och i vilket områden som funktionsytan förväntas ge sämst beskrivning av den kostsamma målfunktionen (global sökstrategi). Det totala antalet utvärderade tulluppsättningar är lika med det initiala antalet samplade tullnivåer adderat med de iterativt samplade tullnivåerna.

3.2 Den kostsamma målfunktionen

Vid optimering av avgiftsnivåer i trängselavgiftssystem utgörs målfunktionen av det samhälls-ekonomiska överskottet som uppnås med en given tulluppsättning. Detta är ett väl etablerat mått på effektiviteten hos ett trängselavgiftssystem (se exempelvis de Rus 2010). Det samhälls-ekonomiska överskottet bestäms genom att addera förändringen av nyttorna för användarna samt nyttorna för samhället (vanligen uttryckt i SEK). För att bestämma dessa nyttoförändringar för en given tulluppsättning måste trafikflöden och restider i trafiknätverket bestämmas. Detta görs med hjälp av en trafikmodell, som kan innehålla både delmodeller för hur reseefterfrågan och ruttval påverkas. För att utvärdera en uppsättning tullnivåer måste således en transportmodell köras, vilket för stora nätverk är mycket tidskrävande. Det är alltså körningen av transportmodellen som gör att målfunktionen är kostsam att beräkna.

I detta projekt används dels en statisk transportmodell och dels VisumDUE med Regent för att beräkna trafikflöden och restider givet en tulluppsättning. Båda dessa modeller är sammankopplade med en modul för att beräkna efterfrågan. För den statistiska modellen används en aggregerad Stockholmsmodell som bygger på indata från efterfrågemodellen T/RIM (Engelson och Svalgård 1995) och för VisumDUE används efterfrågemodellen Regent.

För en given tulluppsättning, här betecknad τ , kan länkflöden och efterfrågan beskrivas som funktion av τ , dvs. $v(\tau)$ samt $q(\tau)$. Länkflöden och efterfrågan erhålls från transportmodellen, för ett givet τ . Från efterfrågemodellen erhålls även restiden i respektive reserelation som $\pi(q(\tau))$. Den samhälls-ekonomiska nytta kan nu beskrivas som $S(v(\tau), \pi(q(\tau)), \tau)$. Den kostsamma målfunktionen $F(\tau)$ har här formulerats som $F(\tau) = -S(v(\tau), \pi(q(\tau)), \tau)$ för att erhålla ett minimeringsproblem.

Förändringen av den samhälls-ekonomiska nyttan kan beskrivas i termer av hur resenärernas upplevda nyttor och kostnader förändras samt hur samhällets nyttor och kostnader förändras vid förändringar av tullnivåerna. Förändringen av resenärernas nyttor och kostnader kan beräknas med logsumman som enkelt erhålls från de efterfrågemodeller som används i detta projekt. I detta projekt har vi begränsat oss till tullintäkterna som samhällets nyttor, men kostnader i termer av exempelvis utsläpp, buller, vägslitage och olyckor kan enkelt inkluderas om det går att beskriva dessa som funktion av länkflöden och restider (eller hastigheter).

3.3 Surrogatmodellen

Surrogatmodellen ger en approximation av målfunktionsvärdet för tulluppsättningar som ännu inte är utvärderade med den kostsamma målfunktionen. Två olika typer av funktionsytor används i detta projekt som surrogatmodeller, RBFer (Gutmann 2001) och Krigingmodeller (Sacks m.fl. 1989; Jones, Schonlau, och Welch 1998).

Radiala basfunktioner (RBF)

Anta en mängd $X = \{x^{(i)} \in \mathfrak{R}^d, i = 1 \dots n\}$ av n samplade punkter, och låt vektorn y beteckna deras målfunktionsvärden. Interpolationsfunktionen $s(x)$ definieras av en viktad summa av radiala basfunktioner $\phi(r)$ enligt:

$$s(x) = \sum_{i=1}^n \lambda_i \phi\left(\|x - x^{(i)}\|_2\right) + b^T x + a. \quad (1)$$

Så länge som alla punkter i X är unika, så kan parametrarna λ , b och a bestämmas unikt genom att lösa ett linjärt ekvationssystem som innehåller både X och y (se exempelvis Holmström, Quttineh, och Edvall 2008 för en detaljerad beskrivning av detta ekvationssystem). Således är interpolationsfunktionen $s(x)$ unikt bestämd av de samplade punkterna. Numeriska problem uppstår när två samplade punkter ligger väldigt nära varandra, men detta kan enkelt hanteras i den metod som används för att bestämma nya punkter att sampla.

Olika funktionsuttryck kan användas i den radiala basfunktionen $\phi(r)$. För exempel på funktioner som vanligen används se exempelvis Holmström, Quttineh, och Edvall (2008). I detta projekt används den vanligt använda funktionen $\phi(r) = r^3$.

Krigingmodeller

Krigingmodellen beskriver varje punkt i interpolationsfunktionen som en realisation av en stokastisk variabel med medelvärde μ och standardavvikelse σ^2 . Estimering av μ och σ^2 görs genom s.k. "Maximum Likelihood Estimation" (MLE), givet X och y . Interpolationsfunktionen, $f(x)$, kan uttryckas som

$$f(x) = \mu + \mathbf{r}' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\mu), \quad (2)$$

där fet stil indikerar vektorer och matriser. \mathbf{R} är en korrelationsmatris som beskrivs av korrelationsfunktionen för alla par av samplade punkter

$$\text{Corr}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = e^{-D(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})}$$

där

$$D(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sum_{k=1}^d \theta_k |x_k^{(i)} - x_k^{(j)}|^p \quad (3).$$

I (3) är d antalet variabler i det ursprungliga optimeringsproblemet och p är en parameter som antingen kan fixeras till ett värde mellan 1 och 2, eller skattas i estimeringen. I Chen m.fl. (2014), som studerar ett liknande optimeringsproblem, har en gausisk korrelationsfunktion använts ($p=2$) men i detta projekt har även en exponentiell korrelationsfunktion använts ($p=1$). θ_k är en skalningsparameter för varje dimension, och skattas i estimeringen.

Jämförelse mellan RBF och Kriging

RBF och Krigingmodeller är båda interpolationsmetoder som används som surrogatmodeller, och som ofta ger snarlika resultat när de används i en surrogatmodell (Quttineh 2012). Locatelli (2013)

visar att Krigingmodeller innefattas i teorin för RBFer. Möjligheten att tolka Krigingmodeller i termer av medelvärde och standardavvikelse ger dock avsevärt bättre möjligheter att utveckla bra samplingsstrategier.

Den viktigaste skillnaden mellan att använda någon av de typiska funktionsytorna i en RBF och att använda Krigingmodeller, är att i en typisk RBF saknas parametrar som är beroende av dimensionen, men i en Krigingmodell ger θ_k varje dimension en individuell vikt.

3.4 Experimentdesign

Experimentdesignen bestämmer vilka punkter som initialt samplas från den kostsamma målfunktionen.

För att kunna anpassa en RBF krävs att antalet punkter som samplas är minst $d + 1$, där d är dimensionen på problemet (antal beslutsvariabler). Om vi exempelvis vill anpassa en RBF för ett problem med sex variabler, måste vi initialt minst sampla sju punkter. För de Krigingmodeller som används i detta projekt finns inget liknande krav, men om den linjära komponenten, μ , i (2) byts ut mot ett polynom kommer också Krigingmodellen att kräva ett minimum av samplade punkter. I Müller (2012) rekommenderas antalet punkter som initialt väljs till $2 \cdot (d + 1)$, men det noteras också att det i de fall där antalet möjliga funktionsevalueringar är kraftigt begränsat kan vara värt att välja färre punkter initialt till förmån för fler punkter i den iterativa samplingen.

En experimentdesign som har de samplade punkterna uniformt spridda är ofta önskvärd eftersom det resulterar i att responsytans approximationsfel sprids ut uniformt. Detta kallas för experimentdesignens "space-filling" egenskap, och har en kraftig influens för hur väl surrogatmodellen kommer approximera den kostsamma funktionen. "Latin Hypercube sampling" (Latin Hypercube Design, LHD, Forrester m.fl. 2008) har goda "space-filling" egenskaper. Genom att använda en maximin LHD från Forrester m.fl., så kan vi dessutom säkerställa att de initial samplade punkterna separerar så mycket som möjligt i L^1 -normen.

Förutom att använda sig av en maximin LHD visar Quttineh och Holmström (2009) fördelarna med att använda hörnpunkter i Lösningrummet samt inrepunkter med lågt målfunktionsvärde (för minimeringsproblem). Att använda hörnpunkter leder dock till ett stort antal initial samplade punkter (det finns 2^d hörnpunkter, där d är problemets dimension). Vi har därför begränsat oss till ett mindre antal uniforma punkter.

3.5 Samplingsstrategi

Samplingsstrategier ("infill strategies") delas vanligen upp i ett-steps- (one-stage) och två-stepsmetoder (two-stage). I en ett-stepsmetod utnyttjas inte den anpassade responsytan när man bestämmer nästa punkt att sampla. Två-stepsmetoder utnyttjar istället information från den anpassade responsytan för att bestämma nästa punkt att sampla. I detta projekt har vi använt två olika två-steps samplingsstrategier. "Candidate point sampling" har använts tillsammans med både RBF surrogatmodellerna och Krigingmodellerna och "Expected improvement sampling" har använts tillsammans med Krigingmodellerna.

Kandidatpunktssampling

Kandidatpunktssampling ("Candidate point sampling", Regis och Shoemaker 2007) bygger på att ett stort antal kandidatpunkter samplas och evalueras i surrogatmodellen. Dessutom bestäms för varje

kandidatpunkt avståndet till den närmast tidigare samplade punkten. Målfunktionsvärdet i surrogatmodellen och avståndet till närmast tidigare samplade punkt vägs sedan samman till ett gemensamt kandidatvärde. Den punkt med högst kandidatvärde väljs sedan som nästa punkt att sampla. På så vis görs en sammanvägning av målet att sampla punkter i områden där det ännu inte samplats någon punkt, samt att sampla punkter i områden där bra lösningar förväntas erhållas. Denna samplingsstrategi ställer inte heller krav på vilken typ av responsyta som används.

För att bestämma ett gemensamt kandidatvärde i punkten x , $K(x)$, måste målfunktionsvärde och avståndsmått skalas. Låt $F(x)$ vara interpolationsfunktionens värde i punkten x , och låt \underline{F} vara målfunktionsvärdet för den kandidatpunkt med lägst målfunktionsvärde och \bar{F} vara målfunktionsvärdet för den kandidatpunkt med högst målfunktionsvärde. För en samplad punkt, x , bestäms det skalade målfunktionsvärdet $K_1(x)$ som $K_1(x) = (\bar{F} - F(x))/(\bar{F} - \underline{F})$. Som avståndsmått används det euklidiska avståndet mellan en samplad kandidatpunkt och den närmast tidigare samplade punkten. Låt $G(x)$ beskriva det euklidiska avståndet mellan punkten x och den närmast tidigare samplade punkten. Det minsta och högsta avståndet för någon av kandidatpunkterna beskrivs av \underline{G} respektive \bar{G} . Det skalade avståndsmåttet $K_2(x)$ bestäms som $K_2(x) = (\bar{G} - G(x))/(\bar{G} - \underline{G})$. Givet vikter för målfunktionsvärde och avstånd (w_1 och w_2) ges kandidatvärdet för en punkt x av $K(x) = w_1 K_1(x) + w_2 K_2(x)$.

Viktningen av de två måtten ger en stor effekt på vilka punkter som väljs. En stor vikt på målfunktionsvärde ger punkter som ligger nära tidigare samplade punkter med lågt målfunktionsvärde (eftersom vi har formulerat målfunktionen som ett minimeringsproblem) och stor vikt på avståndsmåttet ger punkter som ligger långt från tidigare samplade punkter. Det är önskvärt att sampla punkter både i närheten av tidigare bra lösningar och i områden där punkter inte tidigare samplats. Vi har därför valt att iterativt uppdatera w_1 i steg om 0.1 från 0 till 1, och w_2 ges då som $w_2 = 1 - w_1$. Efter 10 iterationer sätts w_1 till 0 för att sedan åter uppdateras.

Kandidatpunkter skapas i två steg. I första steget skapas $1000 \cdot d$ kandidatpunkter genom att störa den hittills bäst funna lösningen. Ytterligare $1000 \cdot d$ kandidatpunkter skapas genom att dra slumpvisa punkter uniformt över hela Lösningsrummet.

“Expected improvement”

Den stora fördelen med Krigingmodeller är att de kan tolkas i termer av medelvärde och standardavvikelse. Denna information kan sedan användas för att utforma effektiva samplingsstrategier. En sådan strategi är förväntad förbättring av målfunktionsvärdet, s.k. ”Expected Improvement Sampling” (Jones, Schonlau, och Welch 1998). Krigingmodeller ger dels en skattning av den kostsamma funktionens värde, $f(x)$, i punkten x , samt en skattning av osäkerheten, $\sigma(x)$. Låt f_{min} vara det hittills bäst funna värdet på den kostsamma funktionen. Förbättringen av f_{min} definieras då som $I = \max\{0, f_{min} - f(x)\}$. Den förväntade förbättringen, $ExpI(\bar{x})$, i punkten \bar{x} , kan sedan formuleras som

$$ExpI(\bar{x}) = \begin{cases} (f_{min} - f(\bar{x}))\Phi\left(\frac{f_{min}-f(\bar{x})}{\sigma(\bar{x})}\right) + \sigma(\bar{x})\phi\left(\frac{f_{min}-f(\bar{x})}{\sigma(\bar{x})}\right) & \text{om } \sigma(\bar{x}) > 0 \\ 0 & \text{om } \sigma(\bar{x}) = 0, \end{cases} \quad (4)$$

där $\Phi\left(\frac{f_{min}-f(\bar{x})}{\sigma(\bar{x})}\right)$ är fördelningsfunktionen och $\phi\left(\frac{f_{min}-f(\bar{x})}{\sigma(\bar{x})}\right)$ täthetsfunktionen för en normalfördelad stokastisk variabel med medelvärde $f_{min} - f(\bar{x})$ och standardavvikelse $\sigma(\bar{x})$. För härledning av $\sigma(x)$ från (2) se Jones, Schonlau, och Welch (1998).

I "Expected Improvement Sampling" (EI-sampling) ges nästa punkt att sampla, \hat{x} , som $\hat{x} = \arg\{\max_x \text{ExpI}(x)\}$. Vilket kräver att $\text{ExpI}(x)$ maximeras. Detta är vanligen ett icke-konvext optimeringsproblem, som i detta projekt har lösts med en kombination av heuristiker som bygger på genetiska algoritmer och "simulated annealing". Den första termen i ExpI favoriserar punkter där surrogatmodeller ger bra lösningar, medan den andra termen favoriserar punkter där osäkerheten är hög. På så sätt uppnås med detta mått en bra avvägning mellan att välja samlingspunkter där tidigare bra lösningar har erhållits och i områden som där få punkter tidigare samplats.

Ett potentiellt problem är att $\sigma(x)$ är en skattad osäkerhet, och med få samplade punkter ger $\sigma(x)$ en underskattning av den verkliga osäkerheten (Jones 2011). I detta projekt har vi därför använt en korrektionsfaktor som utnyttjar resultat från korsvalidering, och som Jones (2011) visar kan förbättra skattningen av $\sigma(x)$.

Ytterligare en samplingsstrategi, som är relaterad till EI-sampling är att välja den punkt där den skattade osäkerheten är som högst (Locatelli 2013). Detta kommer att resultera i en samplingsstrategi som ger god spridning av samplade punkter i lösningsrummet, men helt utan koppling till de områden där bra lösningar hittills har erhållits.

3.6 Implementering

De olika delkomponenterna i en metod som bygger på surrogatbaserad optimering kan implementeras relativt enkelt. Däremot kan det vara lättare eller svårare att skapa en koppling mellan optimeringsmetoden och den kostsamma målfunktionen. I detta projekt har den surrogatbaserade optimeringsmetoden implementerats i MATLAB. Som grund har MATLAB-paketet "Surrogate Model Optimization Toolbox" (Müller 2012) använts.

En viktig del i surrogatbaserad optimering är interaktionen mellan samplingsmetoderna (både före experimentdesign och samplingsstrategi) och den kostsamma målfunktionen. För den statistiska modellen har dessa två delar integrerats i MATLAB. För VisumDUE har det varit relativt få tulluppsättningar att utvärdera och därför har inget gränssnitt mellan MATLAB och VisumDUE använts. Istället har beräkningarna i MATLAB och VisumDUE gjorts separat och sedan överförts manuellt. Enkelheten i surrogatbaserad optimering gör att vi utan någon större förändring av metoden kan använda andra transportmodeller som i framtiden bedöms som intressanta.

4. Utvärdering av metodansats

4.1 Tillvägagångssätt för utvärdering av metodansats

Surrogatbaserad optimering är en metodklass som erbjuder en mängd valmöjligheter för hur de olika komponenterna i metoden utformas. I detta projekt har vi fokuserat på en tillämpning tillsammans med en dynamisk transportmodell, vilket gör antalet evalueringar av den kostsamma målfunktionen kraftigt begränsat, eftersom beräkning av ruttval och efterfrågan för en given tulluppsättning är mycket tidskrävande.

Ett antal olika kombinationer av experimentdesign, samplingsstrategier och val av surrogatmodell har utvärderats i projektet. Utvärderingen har skett i två steg, på liknande sätt som görs i Chen et al. (2014a). I första steget har en statisk transportmodell använts och i det andra steget den dynamiska transportmodellen. Anledningen till detta är att den dynamiska transportmodellen (här VisumDUE med Regent för beräkning av efterfrågan) kräver cirka tio timmar för att utvärdera en uppsättning trängselavgifter. Detta gör det omöjligt att utvärdera ett stort antal kombinationer av experimentdesign, samplingsstrategi och funktionsform för surrogatmodellen. Den statiska transportmodellen kräver istället endast sex minuter för att utvärdera en uppsättning trängselavgifter, vilket gör att ett stort antal olika versioner av metoden kan utvärderas. Att utvärdera metodansatsen med en statisk modell, trots att syftet är att använda den med en dynamisk modell, är inte okomplicerat. Samtidigt ger den statiska modellen möjlighet att jämföra lösningar från den surrogatbaserade optimeringen med lösningar som tagits fram med andra tillgängliga metoder. Nackdelen är att det kan finnas effekter från dynamiskt ruttval i Visum DUE som påverkar hur metoden bör utformas. I det andra steget väljer vi ut en kombination av surrogatmodell, experimentdesign och samplingsstrateg från första steget, och som sedan används tillsammans med den dynamiska transportmodellen. I det andra steget har vi dock ingen möjlighet att jämföra med andra metoder, och det är svårt att utvärdera hur bra lösningar som erhålls. Däremot kan mått på hur väl funktionsytan är anpassad mot de samplade uppsättningarna trängselavgifter tas fram, och jämföras mellan den statiska och dynamiska modellen. Effekten av samplingsstrategi kan också utvärderas och jämföras med resultat från den statiska modellen, genom att jämföra den samhällsekonomiska nyttan för bäst funna tulluppsättning före och efter den iterativa samplingen genomförs.

4.2 Utvärdering med den statiska transportmodellen

Beskrivning av den statiska modellen

Den statiska modellen använder ett aggregerat nätverk för att beskriva väginfrastrukturen i Stockholmsregionen. Ruttval sker enligt Wardrops användarjämvikt och efterfrågan beräknas med en logitmodell som innehåller valen att resa med bil, resa med kollektivtrafik eller att helt avstå från att resa. Nätverket har 392 länkar och 40 zoner, vilket resulterar i 1560 OD-par. Modellen beskriver resande under morgonens rusningstimma, och i Ekström et al. (2014) finns en mer detaljerad beskrivning av modellen. Eftersom nätverket är kraftigt aggregerat, men efterfrågan dåligt kalibrerad mot detta nätverk vet vi att trängselnivåerna är något höga, och den samhällsekonomiska nyttan av trängselskatt riskerar därför att överskattas. Modellen bedöms heller inte vara tillräckligt väl kalibrerad för att resultat från modellen ska kunna användas som beslutsunderlag, däremot bedöms modellen ge kostnader och nyttor i relevant storlek vid införande av trängselavgifter i en stad av Stockholmsstorlek.

Utvärdering sker genom optimering av tullnivåer för sex olika avgiftssnitt: dagens tullring in/ut, Essingeleden N/S och innerstadsbroarna N/S. Dels motsvarar dessa snitt tullsystem som tidigare har diskuterats som möjliga utvidgningar av dagens system, men framför allt ger det ett optimeringsproblem med samma antal beslutsvariabler som används i Chen et al. (2014), vilket underlättar jämförelser med tidigare publicerade resultat. Maximala avgiftsnivåer är satta till 60 SEK per passage, och flerpasageregeln samt Lidingöundantaget är ej inkluderat i modellen.

Beräkning av samhällsekonomisk nytta

För den statistiska modellen formuleras det samhällsekonomiska nyttomåttet, som utgör målfunktionen, som förändringen i samhällsekonomisk nytta jämfört med 0-tull scenariot. För en given tullupsättning, τ , erhålls länkflöden och efterfrågan $v(\tau)$ samt $q(\tau)$, samt restid i respektive reserelation $\pi(q(\tau))$ från den statistiska modellen. Låt I utgöra mängden OD-par och A mängden länkar. Förändringen i samhällsekonomisk nytta kan uttryckas som

$$\Delta S(v(\tau), \pi(q(\tau)), \tau) = US(v(\tau), \pi(q(\tau)), \tau) + OS(v(\tau), \pi(q(\tau)), \tau) - S_0$$

där US är resenärernas nyttoöverskott, OS samhällets nyttoöverskott och S_0 den samhällsekonomiska nyttan i 0-tullscenariot. Resenärernas nyttoöverskott består av summan av kostnader och nyttor för genomförda resor, och samhällets nyttoöverskott i detta projekt endast består av tullintäkter. Notera att resenärernas kostnader och nyttor består av både tid och pengar, men kan med hjälp av ett genomsnittligt tidsvärde formuleras som antingen kostnad och nyttor i tid eller pengar.

För den aktuella modellen ges förändringen av samhällsekonomisk nytta, i SEK, av logsumman

$$\Delta S(v(\tau), \pi(q(\tau)), \tau) = \frac{1}{\lambda} \sum_{i \in I} T_i \ln \frac{A_i e^{\lambda(\pi_i^0 - \pi_i(q_i(\tau)))} + K_i}{T_i} + \sum_{a \in A} \tau_a (v_a(\tau))$$

där λ, A, K, T och π^0 är parametrar i efterfrågemodellen. Den första summan ger förändringen i resenärernas nyttoöverskott och den andra summan samhälletsintäkter. För en beskrivning av ingående parametrar och en mer ingående diskussion av formuleringen av förändring i samhällsekonomisk nytta hänvisas till Ekström (2012).

Genomförande

För den statistiska modellen har två olika Krigingmodeller utvärderats, tillsammans med både kandidatpunktsampling och EI-sampling med korrektionsfaktor. De två Krigingmodellerna skiljer sig åt i valet av parameter p i (3). En gausisk korrelationsfunktion erhålls genom att sätta $p = 2$, och en exponentiell korrelationsfunktion erhålls genom att sätta $p = 1$. RBF som surrogatmodell har utvärderats för kubisk RBF tillsammans med kandidatpunktsampling. För samtliga kombinationer av surrogatmodell och samplingsstrategi har fyra olika experimentdesigner använts. Initial sampling görs med maximin LHD (MLHD) och sju samplade punkter (minimum för RBF) eller 14 samplade punkter (minsta rekommenderade). Dessa experimentdesigner betecknas med MLHD7 respektive MLHD14. För de sju respektive 14 samplade punkterna skapas ytterligare två experimentdesigner genom att lägga till hörnpunkter motsvarande 0-tull scenario samt alla tullar med max avgift (60 SEK) samt de inre punkterna motsvarande 25%, 50% och 75% av den maximala tullnivån. Experimentdesign med uniforma tullnivåer betecknas ”-U”. Totalt resulterar detta i 20 kombinationer, som listas i Tabell 1.

Samtliga varianter av MLHD sampling samt kandidatpunktsampling innehåller slumpstal, och därför har tio upprepade experiment genomförts för varje kombination. För varje kombination har samma slumpstalsupsättningar använts för att bättre kunna göra jämförelser.

För att jämför med resultaten i Chen m.fl. (2014) har dessutom samtliga tre surrogatmodeller (RBF med kandidatpunktsampling och Krigingmodeller med EI-sampling) utvärderats med 67 initialt

samlade tullnivåer, 62 av dessa MLHD sampling och resterande fem är samma uniforma tullnivåer som tidigare använts.

Tabell 1: Utvärderade kombinationer av surrogatmodell, experimentdesign och samplingsstrategi

Kombination nr	Surrogatmodell			Experimentdesign				Samplingsstrategi	
	RBF	KRIG $p=1$	KRIG $p=2$	MLHD7	MLHD7-U	MLHD14	MLHD14-U	KAND	EI
1	X			X				X	
2	X				X			X	
3	X					X		X	
4	X						X	X	
5		X		X				X	
6		X			X			X	
7		X				X		X	
8		X					X	X	
9		X		X					X
10		X			X				X
11		X				X			X
12		X					X		X
13			X	X				X	
14			X		X			X	
15			X			X		X	
16			X				X	X	
17			X	X					X
18			X		X				X
19			X			X			X
20			X				X		X

4.3 Utvärdering med VisumDUE och Regent

Beskrivning av den dynamiska modellen

VisumDue-modellen med Regent är framtagen inom projektet IHOP. I slutrapporten för det projektet finns en mer ingående beskrivning av modellerna (Almroth m.fl. 2014). Modellen avser Stockholms Län, har 1240 zoner (vilka överensstämmer med Sampers prognosområden för Stockholms Län) och ca 17000 länkar. Modellen är bara översiktligt kalibrerad och inte avsedd att användas för policyanalys i dagsläget.

I efterfrågemodellen Regent ingår resgenerering, färdmedelsval (bil som förare, bil som passagerare, kollektivtrafik, cykel och gång) och destinationsval. I VisumDue ingår ruttval och utbudsberäkning. VisumDue är en makroskopisk modell som kan hantera stora nätverk, samtidigt som den beräknar en dynamisk användarjämvikt där hänsyn tas till att svängflödeskapaciteter är begränsade och att länkar uppströms kan blockeras av köer som växer bakåt. Eftersom restider och trängselavgifter påverkar efterfrågan på bilresor behöver iteration göras mellan Regent och VisumDUE. I den använda versionen av regent är endast arbetsresor implementerade. För arbetsresor görs fem iterationer mellan efterfrågan och utbud och i varje iteration används method of successive averages (MSA) för att gå i riktning mot konvergens.

Varje VisumDUE-körning tar mellan 1.5-3 timmar för att nå intern konvergens i nätutläggningen, beroende på hur hårt belastat vägnätet är. Ju högre trängselavgifter desto fler väljer att inte resa med bil, vilket leder till färre bilar att lägga ut på vägnätet och därmed snabbare konvergens i nätutläggningen. Körtiden för Regent är kort i sammanhanget, några minuter per iteration. Detta leder till att hela Regent-VisumDUE tar ca 8-15 timmar att köra, d.v.s. över natt.

Simuleringsperiod är morgonperioden 6:30-9:00³, med en tidsuppdelning på kvartar under maxtimme (7:30-8:30) och halvtimmar utanför maxtimme. I VisumDUE finns möjlighet att lägga in tidsberoende trängselavgifter. Dagens trängselavgiftssystem i Stockholm har tidsprofilen (50%, 75%, 100%, 75%) av maxavgift under perioden 6:30-9:00. Vi har valt att behålla denna tidsprofil för alla trängselavgiftsscenarioer som testas. Samhällsekonomisk nytta beräknas dock bara för maxtimme.

Figur 2 visar de centrala delarna av Stockholmsnätverket i VisumDUE med avgiftsbelagda länkar i lila. Lidingö-undantaget är inte implementerat i modellen. Avgiftsbelagda länkar tillhör ett av sex snitt: dagens tullring in/ut, Essingeleden N/S eller innerstadsbroarna N/S. Totalt ingår 41 länkar (riktningsberoende) i VisumDUE i dessa snitt. För att ändra trängselavgift på länkarna behöver både avgift i maxtimme och avgift i tre tidsperioder utanför maxtimme ändras. I vanliga fall ändras trängselavgift i VisumDUE (precis som i VISUM) via användargränssnittet genom att dubbelklicka på länken och ändra dessa fyra värden per länk. För varje trängselavgiftscenario innebär detta därmed att ändra 164 värden. Då en mängd scenarier körs i projektet märkte vi snart att denna process behövde automatiseras. Ett program har därför skrivits i språket *R* där endast trängselavgiftens nivå på de olika snitten behöver anges (sex värden). Programmet genererar sedan de data som behövs till en .NET-fil som kan läsas in tillsammans med VisumDUE-nätverket och automatiskt skriva över avgiftsnivån (både maxtimme och övriga tidsperioder) på de aktuella länkarna.



Figur 2: Utsnitt ur VisumDUE-vägnätet för Stockholm med trängselavgiftssnitten markerade i lila.

³ Tidsuppdelningen har justerats jämfört med IHOP-projektet så att maxtimma för trängselavgift stämmer överens med maxtimma för biltrafikutbud.

I VisumDUE efterfrågan uppdelad på tre efterfrågesegment: arbetsresor, övriga resor och yrkestrafik. De tre segmenten har olika tidsvärden, se Tabell 2. Som nämnts ovan är det bara arbetsresor som itereras med Regent. Övriga resor skulle också kunna itereras med Regent, men denna funktionalitet var inte implementerad i Regent vid projektets start. Därmed fick övriga resor hanteras som en fast tilläggs-matrix. Yrkestrafik hanteras också som en fast matrix.

Tabell 2: Tidsvärden för de tre efterfrågesegmenten i VisumDUE

Efterfrågesegment	Tidsvärde
Arbetsresor	87
Övriga resor	59
Yrkestrafik	100

Beräkning av samhällsekonomisk nytta

Samhällsekonomisk nytta är målfunktionen vid utvärdering av olika uppsättningar trängselavgifter även i det dynamiska fallet. Den samhällsekonomiska nyttan beräknas relativt nyttan i situationen utan trängselavgift. De termer som ingår i den samhällsekonomiska nyttan är: restidsvinster/kostnader, trängselavgiftskostnad, anpassningskostnader, samt totala summan intäkter. Anpassningskostnaden beräknas endast för arbetsresor eftersom bara dessa itereras med efterfrågemodellen. Restidsvinster/kostnader, trängselavgiftskostnad och anpassningskostnad ingår alla i logsumman som Regent beräknar för arbetsresor. Samhällsekonomiska nyttan S_s trängselavgiftsscenario s beräknas från resultatet av körning med Regent/VisumDUE för maxtimme kl. 7:30 - 8:30 enligt

$$S_s = \frac{1}{c} \left[\sum_{z \in Z} \sum_{a \in A_z} \ln(1 + e^{u_{a,s}}) - \sum_{z \in Z} \sum_{a \in A_z} \ln(1 + e^{u_{a,0}}) \right] - \sum_{i \in I} [(\beta_x \pi_{i,s} + \tau_{i,s} - \beta_x \pi_{i,0}) q_{x,i} + (\beta_y \pi_{i,s} + \tau_{i,s} - \beta_y \pi_{i,0}) q_{y,i}] + \sum_{l \in L} \tau_l v_l$$

Logsumman beräknas för varje agent a i Regent och summeras sedan över alla agenter, A_z , som startar i zon z , för att därefter summeras över alla zoner, Z , i Stockholms län. Logsumman beräknas på resgenereringsnivån. Eftersom Regent är en nästlad logit-modell med nivåerna (nerifrån och upp) destinationsval, färdmedelsval och resgenerering, kommer nyttorna från destinations- och färdmedelsvalsnivåerna att komma med i logsumman på resgenereringsnivån. Resgenereringen har endast två alternativ: inte resa (nytta $u = 0$ ger $e^0 = 1$) och resa (nytta $u_{a,s}$ för agent a i trängselavgiftsscenario s och $u_{a,0}$ för agent a i scenario 0 utan trängselavgift). Nyttan u innehåller en kalibreringskonstant samt logsumman från färdmedelsvalnivån multiplicerad med en resgenereringsparameter. På motsvarande sätt innehåller nyttorna på färdmedelsvalnivån logsumman från destinationsvalet, samt en bilkonkurrensvariabel och färdmedelskonstanter. Nyttorna på destinationsvalsnivån innehåller restid och reskostnad (biljettkostnad/avståndskostnad och eventuell trängselavgift), samt några få socioekonomiska parametrar (kvinna/man). Kostnadsparametern c skalar om logsumman till SEK. Övriga resor och yrkestrafik hanteras som fasta matriser i VisumDUE. Förändringen av den samhällsekonomisk nytta för dessa ärenden beräknas som skillnaden mellan generaliserad reskostnad i situationen med och utan trängselavgift. Index för övriga resor och yrkestrafik anges med x respektive y . Tidsvärde för respektive efterfrågesegment betecknas β , och restiden respektive trängselavgiften för OD-par i och scenario s betecknas $\pi_{i,s}$ och $\tau_{i,s}$. Länkflödet på

länk l betecknas v_l och q utgör efterfrågan i respektive OD-par och efterfrågesegment. Mängden OD-par och länkar betecknas I respektive L .

Genomförande

För VisumDUE med Regent har kombination 18 från Tabell 1 valts. Denna kombination utgörs av en Krigingmodell med gaussisk korrelationsfunktion och EI-sampling tillsammans med tolv initialt samplade tulluppsättningar, varav sju är framtagna med MLHD och fem är de tidigare beskrivna uniforma tulluppsättningarna. Vidare motivering till detta val ges i kapitel 5.

4.4 Utvärderingsmått

För den statistiska modellen sker utvärdering mot optimum som beräknas med hjälp av MATLABs paket för globaloptimering. Från surrogatmodellen presenteras utvärdering av surrogatmodellens optimallösning i den kostsamma målfunktionen som medelvärde från de upprepade experimenten med olika slumpvalsuppsättningar samt 95% konfidensintervall.

För både den statistiska och dynamiska modellen presenteras korsvalideringsresultat med både normaliserat rotmedelkvadratfel (NRMSE) och normaliserat maximalt absolut fel (NMAE). Det första av dessa ger en uppskattning av medelfelet över hela domänen, medan NMAE måttet ger en uppskattning felet där det är som störst. Korsvalideringen görs genom s.k. utelämna-en validering, där en av de samplade punkterna utelämnas för att sedan skatta om surrogatmodellen och därefter approximera målfunktionsvärdet i den utelämnade punkten. Det approximerade målfunktionsvärdet kan sedan jämföras med det kända målfunktionsvärdet. Fördelen med utelämna-en validering är att det inte kräver att nya punkter utvärderas i den kostsamma målfunktionen. För en genomgående beskrivning av utelämna-en validering, NRMSE och NMAE i ett liknande sammanhang se Chen m.fl. (2014).

Dessutom kommer surrogatmodellens maximala samhällsekonomiska nytta också presenteras ("estimated global optimum", EGO). Detta mått ger dock ingen information om hur väl den aktuella responsytan approximerar den kostsamma funktionen. Däremot ger EGO en approximation av den kostsamma funktionens optimala värde.

Vad som också är intressant är bidraget från den iterativa samplingen, därför utvärderas surrogatmodellens optimum i den kostsamma målfunktionen efter att den initiala samplingen har genomförts för att sedan kunna jämföras med de lösningar som erhålls under den iterativa samplingen.

5. Resultat

5.1 Statisk modell

I resultaten som presenteras i Tabell 3 till 5 avser $S(\hat{x})$ samhällsekonomisk nytta vid utvärdering av surrogatmodellens optimallösning (\hat{x}) i den kostsamma målfunktionen och EGO avser surrogatmodellens målfunktionsvärde för samma lösning.

Målfunktionsvärde för nära globalt optimum (GO), framtaget med meta-heuristiker enligt tidigare beskrivning, är 607 130.

Tabell 3: Samhällsekonomsnytta, $S(\hat{x})$ (i SEK/h), efter 20, 30 respektive 40 samplade tulluppsättningar

Kombination nr	Efter 20 samplade tulluppsättningar		Efter 30 samplade tulluppsättningar		Efter 40 samplade tulluppsättningar	
	Medel	95% konf. +/-	Medel	95% konf. +/-	Medel	95% konf. +/-
1	546 235	21 297	578 350	9 348	590 723	7 044
2	560 343	14 156	584 300	9 763	592 797	6 352
3	505 194	22 652	570 627	16 435	589 693	13 189
4	538 991	20 261	574 762	12 937	586 898	6 654
5	498 239	36 464	516 199	30 265	536 629	35 573
6	540 621	20 962	558 171	18 288	562 985	18 753
7	518 449	25 268	549 513	23 120	575 677	9 777
8	509 769	24 665	555 662	26 824	567 838	25 143
9	556 173	23 140	568 082	21 839	575 087	15 169
10	582 052	9 049	589 877	6 585	592 605	6 302
11	567 148	17 385	581 766	10 871	585 358	11 050
12	565 737	13 573	584 609	8 499	587 637	7056
13	541 512	37 030	575 102	18 209	585 412	13 473
14	567 983	10 774	587 442	7 534	596 470	4 575
15	540 852	26 011	570 081	22 185	588 664	7 058
16	536 008	21 985	585 438	9 047	596 470	6 209
17	593 269	5 800	600 799	2 306	602 595	1 237
18	583 990	6 209	598 306	3 228	602 972	1 735
19	558 742	19 496	591 155	8 573	598 320	3 901
20	553 054	17 653	589 917	8 036	598 948	4 406

För 20 samplade tulluppsättningar är det tre olika kombinationer med EI-sampling som särskilt utmärker sig (10, 18 och 19). Det är dels Krigingmodellen med gausisk korrelationsfunktion och MLHD7/MLHD7-U samt Krigingmodellen med exponentiell korrelationsfunktion och MLHD7-U. Ser man till osäkerheten i resultaten och hur de olika kombinationerna presterar vid ett större antal samplade tulluppsättningar är det Krigingmodellen med gausisk korrelationsfunktion och MLHD7/MLHD7-U som ger högst samhällsekonomsnytta.

Notera också att för alla kombinationer av surrogatmodell och samplingsstrategi, förutom Kriging med gausisk korrelationsfunktion och EI-sampling, ger experimentdesignen MLHD7-U bäst resultat. Detta är en experimentdesign med totalt tolv samplade tulluppsättningar, varav fem är de uniforma tulluppsättningarna. Att MLHD14/ MLHD14-U (14 respektive 19 initial samplade tulluppsättningar) alltid presterar sämre än MLHD7-U visar att de iterativt samplade tulluppsättningarna är viktigare än ett större antal initialt samplade tulluppsättningar.

Slutligen kan vi konstatera att med endast 20 evalueringar av den kostsamma målfunktionen nås i genomsnitt 92.3% av GO, och med 20 ytterligare iterationer når vi 97.7% av GO.

Tabell 4: Medelvärde av NRMSE, NMAE och $|S(\hat{x}) - EGO|$ efter 20 samplade tulluppsättningar med 10 upprepade experiment

Kombination nr	NRMSE (%)	NMAE	$ S(\hat{x}) - EGO $
1	32.3	1.64	11 133
2	36.4	1.97	13 256
3	41.0	2.55	38 579
4	33.2	1.92	46 001
5	32.4	1.96	0.00
6	40.7	1.70	0.63
7	37.7	2.10	0.00
8	42.8	2.16	0.00
9	19.3	1.77	1 313
10	29.6	1.70	2 285
11	21.9	1.79	1 343
12	38.3	2.09	337
13	26.5	1.50	5 782
14	37.8	1.64	11 752
15	37.5	2.17	10 637
16	42.9	2.04	11 674
17	14.3	1.31	3 984
18	25.7	1.60	7 487
19	20.6	1.69	2 910
20	36.6	2.05	10 230

I Tabell 4, som visar genomsnittliga värden på NRMSE, NMAE och $|S(\hat{x}) - EGO|$ efter 20 samplade tulluppsättningar, är det tydligt att Krigingmodellen med exponentiell korrelationsfunktion ger en väldigt lite skillnad mellan uppskattat globalt optimum (EGO) och det värde som erhålls när denna lösning utvärderas i den kostsamma målfunktionen ($S(\hat{x})$). Detta beror uteslutande på att vi med denna Krigingmodell erhåller optimum i eller nära redan samplade punkter.

De tre kombinationer som presterar bäst utifrån NRMSE är 17, 9 och 19, och utifrån NMAE är 17, 13 och 18. Kombinationerna 9,13 och 19 presterar samtliga dåligt utifrån uppnådd samhällsekonomisk nytta efter 20 samplade tulluppsättningar. En viss försiktighet bör man därför ha när man uttalar sig om vilken kombination som är bäst baserat på endast NRMSE och NMAE.

Tabell 5: NRMSE (%), NMAE, EGO och $S(\hat{x})$ efter 67 och 97 samplade tullnivåer (N)

Surrogatmodell	N	NRMSE	NMAE	EGO	$S(\hat{x})$
Kriging $p=1$	67	26.6	2.46	601 051	590 942
Kriging $p=2$		19.8	2.12	608 207	594 461
kubisk RBF		10.23	1.00	610 891	596 657
Kriging $p=1$	97	18.41	2.11	618 294	604 770
Kriging $p=2$		12.84	1.71	606 626	605 754
kubisk RBF		6.95	0.78	599 181	599 702

Resultaten som presenteras i Tabell 5 syftar främst till att kunna göra relevanta jämförelser med resultaten från Chen m.fl. (2014). Chen m.fl. applicerar surrogatbaserad optimering på ett

motorvägsnät i Maryland, USA, med sex tullvariabler att bestämma värde på. Fem av dessa avser vanliga tullnivåer och den sjätte variabeln avser storleken på tullarna "off-peak" som procentsats av tullarna under "peak". Jämförelsen med de sex tullvariablerna vi arbetar med i våra analyser haltar därför delvis, men dimensionen på optimeringsproblemet är densamma. Trafikmodellen som används är en simuleringsbaserad dynamisk trafikmodell med fix efterfrågan. Chen m.fl. använder två olika varianter av Krigingmodeller, där den ena versionen hanterar brus i målfunktionen. Detta är viktigt för simuleringsbaserade trafikmodeller, men inte relevant för VisumDUE som är en deterministisk modell. Efter de 67 initialt samplade tulluppsättningarna erhåller Chen m.fl. 2.91% NRMSE och 2.26 NMAE med motsvarande Kriging $p = 2$, och efter ytterligare 30 iterativt samplade tulluppsättningar med EI-sampling nås 2.92% NRMSE och 2.74 NMAE med motsvarande brushanterade version av Kriging $p = 2$. Vad gäller medelfelet ser vi tydligt att vi har avsevärt högre osäkerheter i resultaten med den statistiska modellen, medan det absoluta felet är mindre. Vi erhåller alltså för liknande surrogatmodeller och samplingsstrategier, med samma antal samplade punkter, en avsevärt sämre approximation av den kostsamma målfunktionen. En möjlig förklaring till de stora skillnaderna är att elastisk efterfrågan ger en målfunktion som är svårare att approximera.

5.2 VisumDUE med Regent

Val av surrogatmodell, samplingsstrategi och experimentdesign

För experiment med VisumDUE har en Krigingmodell med gausisk korrelationsfunktion använts, tillsammans med EI-sampling och en experimentdesign som består av sju samplade punkter med MLHS, de fem tidigare beskrivna uniforma tulluppsättningarna samt två ytterligare tulluppsättningar. De två ytterligare uppsättningarna är dels den nuvarande trängselavgiften på tullringen i Stockholm, och ingen trängselavgift i övrigt, och dels en tulluppsättning som är utgör ett expertförslag.

Resultaten med den statistiska modellen visar tydligt att en Krigingmodell med gaussisk korrelationsfunktion och EI-sampling är den bästa kombinationen av surrogatmodell och samplingsstrategi. Om vi endast beaktar resultaten med statistisk modell, är en experimentdesign med endast sju initiala tulluppsättningar att föredra när det totala antalet funktionsevalueringar är starkt begränsat. Som tidigare diskuterats visar tidigare forskning dock att ytterligare inre punkter med bra målfunktionsvärden är önskvärt, särskilt om punkter på randen av det tillåtna området ger sämre målfunktionsvärden. För den statistiska modellen är detta inte ett problem då även randpunkter ger hög samhällsekonomisk nytta. En första utvärdering av experimentdesign med endast sju punkter enligt MLHD-sampling, som alla ligger på randen, ger ingen tulluppsättning med positiv samhällsekonomisk nytta. Därför har de fem uniforma punkterna, varav tre är inre punkter, samt den nuvarande tullringen i Stockholm och expertförslaget, använts för att skapa en experimentdesign med bättre målfunktionsvärden. De tulluppsättningar som initialt samplas finns listade i Tabell 6 tillsammans med den resulterande samhällsekonomiska nyttan. I Tabell 6 avser "MLHD" de punkter som är samplade med MLHD-sampling, "U" de uniforma punkter som samplats, samt "Tullring" och "Expert" de ytterligare initialt samplade punkter som valt baserat på nuvarande tullring och förslag från expert.

Experiment med den statistiska modellen har utvärderats efter tjugo tulluppsättningar, varav åtta är genererade under den iterativa samplingen. För VisumDUE med Regent utgör fjorton tulluppsättningar den initiala samplingen, och därefter skapas ytterligare åtta under den iterativa

samplingen för att kunna göra en jämförelse av hur den iterativa samplingen fungerar med de två olika modellerna.

Tabell 6: Initial samplade tullnivåer

	Essingeleden		Innterstadsbroarna		Tullring		Samhällsekonomisk nytta (SEK/h)
	Norrg.	Söderg.	Norrg.	Söderg.	In	Ut	
MLHD1	30	10	20	0	50	0	-536 424
MLHD2	10	60	30	10	40	40	-7 402
MLHD3	0	0	0	20	0	30	-16 648
MLHD4	0	0	60	30	30	20	-416 996
MLHD5	50	20	40	40	60	50	-438 640
MLHD6	20	30	10	60	20	60	-271 827
MLHD7	40	50	50	50	10	10	-174 695
U1	0	0	0	0	0	0	0
U2	15	15	15	15	15	15	349 560
U3	30	30	30	30	30	30	45 047
U4	45	45	45	45	45	45	-335 677
U5	60	60	60	60	60	60	-817 663
Tullring	0	0	0	0	20	20	232 562
Expert	50	20	20	10	30	10	-54 157

Resultat från experiment

I Tabell 7 presenteras de tulluppsättningar som utgör optimallösningen till surrogatmodellens funktionsyta före (A) och efter (B) att den iterativa samplingen genomförs samt den tulluppsättning som motsvarar dagens tullring i Stockholm (C), tillsammans med resulterande samhällsekonomisk nytta. Den samhällsekonomiska nyttan uppdelad på intäkt och anpassningskostnad ges i Tabell 8, tillsammans med uppskattad intäkt per år för respektive tulluppsättning. Tabell 9 presenterar NRMSE och NMAE efter varje iterativt samplad tulluppsättning lagts till.

Tabell 7: Samhällsekonomisk nytta för tulluppsättning före (A) och efter (B) iterativ sampling, samt för nuvarande tullring i Stockholm (C).

	Essingeleden		Innterstadsbroarna		Tullring		Samhällsekonomisk nytta (SEK/h)
	Norrgående	Södergående	Norrgående	Södergående	In	Ut	
A	11.86	11.82	11.55	8.73	16.49	15.02	353 590
B	10.15	6.07	14.93	10.23	16.72	14.16	406 300
C	0	0	0	0	20.00	20.00	232 562

Tabell 8: Samhällsekonomisk nytta uppdelad i anpassningskostnad och intäkt, samt uppskattad intäkt per år.

	Konsumentöverskott (SEK/h)	Intäkt (SEK/h)	Uppskattad intäkt per år (MSEK/år)
A	436 572	790 162	1104
B	388 727	795 027	1111
C	338 740	571 332	798

Tabell 9: NRMSE, NMAE efter varje iterativt samplad tulluppsättning

Iteration	NRMSE (%)	NMAE
0	80.0	2.01
1	76.4	1.92
2	68.7	1.77
3	65.7	1.59
4	71.8	1.77
5	72.3	1.75
6	67.1	1.75
7	63.9	1.83
8	65.0	1.76

Resultaten i Tabell 7 visar tydligt att den surrogatbaserade optimeringen lyckas hitta tulluppsättningar som ökar den samhällsekonomiska nyttan. Det är särskilt intressant att undersöka hur stor förbättring som erhålls med den iterativa samplingen. I detta fall har den samhällsekonomiska nyttan ökat med 14.9% efter den iterativa samplingen jämfört med optimum före den iterativa samplingen. Som jämförelse ökar den iterativa samplingen den samhällsekonomiska nyttan med 4%-18% när vi använder samma kombination av surrogatmodell och samplingsstrategi med den statistiska modellen. Den genomsnittliga förbättringen är 11% med statistisk modell. Optimeringen över surrogatmodellen före den iterativa samplingen ökar, för VisumDUE med Regent, målfunktionsvärdet med endast 1.2% jämfört med den samplade punkt som har högst samhällsekonomisk nytta. Som jämförelse ökar den samhällsekonomiska nyttan med 0%-15% med den statistiska modellen. Den genomsnittliga ökningen är 3.9% och medianökningen är 3.1%.

I Tabell 9 är det tydligt att VisumDUE med Regent ger en målfunktion som är avsevärt svårare att approximera med en surrogatmodell. Det maximala felet (NMAE) är inte särskilt mycket större, men medelfelet (NRMSE) tyder på en sämre förmåga att approximera den kostsamma målfunktionen. Samtidigt är det tydligt att tillsammans med EI-sampling fungerar modellen relativt väl, och tar fram en lösning som förbättrar den samhällsekonomiska nyttan.

Ur ett policyperspektiv är det intressant att se på uppdelningen mellan konsumentöverskott och intäkt. Den lösning som erhålls från den surrogatbaserade optimeringen (B) ger en samhällsekonomisk nytta som är 98% högre jämfört med nuvarande tullring, och som ökar intäkterna med 39%. Samtidigt ökar konsumentöverskottet endast med 15%. Det är alltså frågan om en relativt kraftig ökning av intäkterna, utan att den upplevda kostnaden för resenärerna ökar i samma utsträckning.

6. Diskussion

För både den statistiska och dynamiska transportmodellen är det tydligt att den iterativa samplingen förbättrar den samhällsekonomiska nyttan, i jämförelse med den nytta som uppnås efter den initiala samplingen. En rimlig fråga är dock huruvida utvärderingen med statistisk modell bidrar till att välja den kombination av surrogatmodell, experimentdesign och samplingsstrategi som fungerar bäst med den dynamiska transportmodellen.

Två alternativa ansatser till den utvärdering som gjorts i detta projekt är:

- (i) Att använda samma nätverk, samma uppdelning av resor i efterfrågesegment och samma efterfrågemodell, för både den statiska och dynamiska transportmodellen. Detta skulle göra det möjligt att urskilja effekter som beror på den dynamiska modellen.
- (ii) Att använda en mindre beräkningskrävande dynamisk transportmodell (exempelvis genom att använda ett mindre nätverk). Detta skulle göra det möjligt att även jämföra lösningen från den surrogatbaserade optimeringen med global optimum.

För (i) är svårigheten att även med den statiska modellen skulle beräkningstiden, för nätverket som har använts med Regent/VisumDUE, göra det svårt att analysera alla de olika kombinationerna av surrogatmodell, experimentdesign och samplingsstrategi som utvärderats med den statiska modellen i denna rapport. För (ii) är utmaningen att konstruera ett mindre nätverk som fortfarande innehåller graden av komplexitet som finns i ett större nätverk. Således är det inte säkert att de slutsatser som dras från experiment med det mindre nätverket är relevanta för fallet med ett fullstort trafiknätverk. Den ansats vi har använt i detta arbete liknar till stor del angreppssättet från Chen m.fl. (2014), men med den viktiga skillnaden att vi använder en statisk modell som representerar ett faktiskt trafiksystem.

Det är värt att notera att den surrogatbaserade optimeringsmetoden vi har använt endast är beroende av möjligheten att beräkna en samhällsekonomisk nytta för en given tulluppsättning. Detta gör att den kan användas med alla modellsystem som givet information om aktuella avgiftsnivåer returnerar en samhällsekonomisk nytta.

7. Slutsatser

I detta projekt har surrogatbaserad optimering använts för att utvärdera dess tillämpbarhet vid optimering av avgiftsnivåer för tullsnitt i trängselavgiftssystem. Utvärderingen med statisk transportmodell visar att metodansatsen, med under 40 körningar av transportmodellen, når mycket nära de resultat som har erhållits med stokastiska sökmetoder. Resultaten visar också att EI-sampling är en mycket bra metod för att iterativt generera ytterligare tulluppsättningar att sampla. Det är användandet av EI-sampling som gör att Krigingmodellen är mest lämpad för denna typ av optimeringsproblem, då Krigingmodellen med kandidatpunktssampling ger resultat i samma nivå som RBF-modellen.

Även med ett kraftigt begränsat antal utvärderingar av den kostsamma målfunktionen i den dynamiska transportmodellen har vi visat att det är möjligt att använda metodansatsen. En tydlig förbättring av den samhällsekonomiska nyttan uppnås med endast 22 utvärderade tulluppsättningar. Ytterligare experiment skulle dock behövas för att undersöka hur stor denna förbättring är i förhållande till vad som skulle kunna uppnås.

En så god approximation av den kostsamma målfunktionen som uppnås med en dynamisk transportmodell i Chen m.fl. (2014) har inte kunnat återupprepas i detta projekt. En avsevärt sämre approximation erhålls även med VisumDUE och Regent i jämförelse med den statiska modellen. Frågan är om en simuleringsbaserad transportmodell med dynamisk efterfrågan skulle ge en ytterligare svårare målfunktion att approximera. Att använda den befintliga kombinationen av Transmodeller och Regent som finns implementerad för Stockholms Län skulle vara en intressant jämförelse.

En intressant framtida forskningsfråga är hur en enklare modell (exempelvis en statisk modell) och en mer detaljerad modell (exempelvis en dynamisk modell) kan kombineras i en surrogatbaserad optimeringsmetod. Särskilt intressant blir det för efterfrågemodellen Regent som finns koppad till både statistiska och dynamiska transportmodeller.

Referenser

- Almroth, A., S. Berglund, O. Canella, L. Engelson, G. Flötteröd, D. Jonsson, I. Kristoffersson, och J. West. 2014. *Further development of Sampers and modelling of urban congestion*. IHOP final report. http://swopec.hhs.se/ctswps/abs/ctswps2014_010.htm.
- Börjesson, M., J. Eliasson, M. Beser-Hugosson, och K. Brundell-Freij. 2012. "The Stockholm congestion charges—5 years on. Effects, acceptability and lessons learnt". *Transport Policy* 20: 1–12.
- Börjesson, M., och I. Kristoffersson. 2014. "Assessing the welfare effects of congestion charges in a real world setting". *Transportation Research Part E: Logistics and Transportation Review* 70: 339–55.
- . 2015. "The Gothenburg congestion charge - Traffic effects, design and politics". *Submitted*.
- Chen, X.M., L. Zhang, X. He, C. Xiong, och Z. Li. 2014. "Surrogate-Based Optimization of Expensive-to-Evaluate Objective for Optimal Highway Toll Charges in Transportation Network". *Computer-Aided Civil and Infrastructure Engineering* 29 (5): 359–81.
- De Palma, A., M. Kilani, och R. Lindsey. 2005. "Congestion pricing on a road network: A study using the dynamic equilibrium simulator METROPOLIS". *Transportation Research Part A* 39 (7-9): 588–611.
- De Rus, Ginés. 2010. *Introduction to Cost-Benefit Analysis: Looking for Reasonable Shortcuts*. Cheltenham, Glos, UK; Northampton, Mass., USA: Edward Elgar, \$c c2010.
- Ekström, J., L. Engelson, och C. Rydergren. 2014. "Optimal toll locations and toll levels in congestion pricing schemes: a case study of Stockholm". *Transportation Planning and Technology* 37 (4): 333–53.
- Ekström, Joakim. 2012. "Optimization Approaches for Design of Congestion Pricing Schemes". Doktorsavhandling, Linköping: Linköpings universitet.
- Eliasson, J., M. Börjesson, D. van Amelsfort, K. Brundell-Freij, och L. Engelson. 2013. "Accuracy of congestion pricing forecasts". *Transportation Research Part A: Policy and Practice* 52: 34–46.
- Eliasson, J., L. Hultkrantz, L. Nerhagen, och L. Rosqvist. 2009. "The Stockholm congestion-charging trial 2006: Overview of effects". *Transportation Research Part A* 43 (3): 240–50.
- Engelson, Leonid, och S. Svalgård. 1995. "The travel demand and land-use model for Stockholm region". Inregia AB.
- Engelson, L., och D. van Amelsfort. 2015. "The role of volume-delay functions in forecasting and evaluating congestion charging schemes: the Stockholm case". *Transportation Planning and Technology* 38 (6): 684–707.
- Forrester, Alexander, Andrés Sóbester, Andy Keane, och Wiley InterScience (Online service). 2008. *Engineering Design via Surrogate Modelling a Practical Guide*. Hoboken, N.J.; Chichester: Wiley ; John Wiley [distributör]. <http://public.eblib.com/choice/publicfullrecord.aspx?p=366798>.
- Gutmann, H. -M. 2001. "A Radial Basis Function Method for Global Optimization". *Journal of Global Optimization* 19 (3): 201–27. doi:10.1023/A:1011255519438.
- Holmström, Kenneth, Nils-Hassan Quttineh, och Marcus M. Edvall. 2008. "An Adaptive Radial Basis Algorithm (ARBF) for Expensive Black-Box Mixed-Integer Constrained Global Optimization". *Optimization and Engineering* 9 (4): 311–39. doi:10.1007/s11081-008-9037-3.
- Jones, Donald R. 2011. "Improved Quantification of Prediction Error for Kriging Response Surfaces". presenterad vid IMA Hot Topics Workshop Uncertainty Quantification in Industrial and Energy Applications: Experiences and Challenges, juni 2.

- Jones, Donald R., Matthias Schonlau, och William J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization* 13 (4): 455–92.
- Kristoffersson, I. 2013. "Impacts of time-varying cordon pricing: Validation and application of mesoscopic model for Stockholm". *Transport Policy* Vol. 28 (juli): p. 51–60.
- Kristoffersson, I., och L. Engelson. 2009. "A dynamic transportation model for the Stockholm area: Implementation issues regarding departure time choice and OD-pair reduction". *Networks and Spatial Economics* 9 (4): 551–73.
- Locatelli, Marco. 2013. *Global optimization: theory, algorithms, and applications*. MOS-SIAM series on optimization. Philadelphia: Mathematical Programming Society : Society for Industrial and Applied Mathematics.
- Müller, Juliane. 2012. *User guide for Modularized Surrogate Model Toolbox*. Teknisk rapport. Tampere University of Technology.
- Peeta, Srinivas, och Athanasios K. Ziliaskopoulos. 2001. "Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future". *Networks and Spatial Economics* 2001 (3-4): 233–65.
- Quttineh, Nils-Hassan. 2012. "Models and Methods for Costly Optimization and Military Decision Support Systems". Doktorsavhandling, Linköping: Linköpings universitet.
- Quttineh, Nils-Hassan, och Kenneth Holmström. 2009. "The Influence of Experimental Designs on the Performance of Surrogate Model Based Costly Global Optimization Solvers". *Studies in Informatics and Control* 18 (1): 87–95.
- Regis, Rommel G., och Christine A. Shoemaker. 2007. "A Stochastic Radial Basis Function Method for the Global Optimization of Expensive Functions". *INFORMS Journal on Computing* 19 (4): 497–509. doi:10.1287/ijoc.1060.0182.
- Sacks, Jerome, William J. Welch, Toby J. Mitchell, och Henry P. Wynn. 1989. "Design and Analysis of Computer Experiments". *Statistical Science* 4 (4): 409–23.
- Shepherd, Simon, och Agachai Sumalee. 2004. "A Genetic Algorithm Based Approach to Optimal Toll Level and Location Problems". *Networks and Spatial Economics* 4 (2): 161–79. doi:10.1023/B:NETS.0000027771.13826.3a.
- Sveriges Riksdag. 2014. "Förändrad trängselskatt och infrastruktursatsningar i Stockholm - Debatt och beslut". mars 26. <http://www.riksdagen.se/sv/Debatter--beslut/Debatter-och-beslut-om-forslag/Arendedebatter/?did=H101SkU24>.
- Yang, Hai, Wei Xu, Bing-sheng He, och Qiang Meng. 2010. "Road Pricing for Congestion Control with Unknown Demand and Cost Functions". *Transportation Research Part C: Emerging Technologies* 18 (2): 157–75. doi:10.1016/j.trc.2009.05.009.
- Zhang, Xiaoning, och Hai Yang. 2004. "The Optimal Cordon-Based Network Congestion Pricing Problem". *Transportation Research Part B: Methodological* 38 (6): 517–37. doi:10.1016/j.trb.2003.08.001.