

Surrogate-based optimization of cordon toll levels in congested traffic networks

Joakim Ekström, Ida Kristoffersson and Nils-Hassan Quttineh

Linköping University Pre-Print



N.B.: When citing this work, cite the original article.

This is the pre-peer reviewed version of the following article:

Joakim Ekström, Ida Kristoffersson and Nils-Hassan Quttineh, Surrogate-based optimization of cordon toll levels in congested traffic networks, 2016, Accepted for publication in Journal of Advanced Transportation.

Copyright: Wiley: 12 months

<http://eu.wiley.com/WileyCDA/>

Preprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-126894>

Surrogate-based optimization of cordon toll levels in congested traffic networks

Joakim Ekström^{*1}, Ida Kristoffersson² and Nils-Hassan Quttineh³

¹Communications and Transport Systems, Department of Science and Technology, Linköping University, Norrköping, Sweden

²Sweco, Stockholm, Sweden

³Optimization, Department of Mathematics, Linköping University, Linköping, Sweden

2015

Abstract

The benefits, in terms of social surplus, from introducing congestion pricing schemes in urban networks are depending on the design of the pricing scheme. The major part of the literature on optimal design of congestion pricing schemes is based on static traffic assignment, which is known for its deficiency in correctly predict travels in networks with severe congestion. Dynamic traffic assignment can better predict travel times in a road network, but are more computational expensive. Thus, previously developed methods for the static case cannot be applied straightforward. Surrogate-based optimization is commonly used for optimization problems with expensive-to-evaluate objective functions. In this paper we evaluate the performance of a surrogate-based optimization method, when the number of pricing schemes which we can afford to evaluate (due to the computational time) is limited to between 20 and 40. A static traffic assignment model of Stockholm is used for evaluating a large number of different configurations of the surrogate-based optimization method. Final evaluation is done with the dynamic traffic assignment tool VisumDUE, coupled with the demand model Regent, for a Stockholm network including 1 240 demand zones and 17 000 links. Our results show that the surrogate-based optimization method can indeed be used for designing a congestion pricing scheme which return a high social surplus.

1 Introduction

Road pricing continues being suggested as a tool for traffic management in urban areas. In Sweden, Stockholm and Gothenburg have adopted congestion charges, and road

^{*}Corresponding author, joakim.ekstrom@liu.se

pricing in terms of congestion charging has previously been implemented in Singapore and London. The benefits from introducing congestion charges are well established in the literature of transportation economics [1–3], but also from practical experience in Stockholm [4, 5] and in London [6].

There has recently been a growing interest in analyzing road pricing schemes in urban areas using dynamic traffic assignment (DTA) models. The motivation behind this development is the problem for traditional (static) traffic assignment models to accurately predict travel time savings in networks with severe congestion, when evaluating road pricing schemes [7–9]. In theory, first-best pricing can be achieved by introducing marginal social cost pricing (MSCP) tolls [10]. In practice there are, however, limitations on how the charging system may be designed, and second-best pricing schemes are usually suggested. Further discussions on implementation and public acceptability of congestion charges can for instance be found in [11–15].

In a second-best pricing scheme the objective is to maximize the social surplus, under constraints on toll locations and/or toll levels. Finding second-best optimal toll levels and their locations in urban road traffic networks has so far mainly been studied using either derivative-free heuristics (for example genetic algorithms and simulated annealing) or ascent methods, together with static traffic assignment models. For an overview of developed methods we refer to [16]. Both derivative-free heuristics and ascent methods rely on fast computations of the road users response (traffic flows, travel times and demands), given the road pricing scheme. For the case of ascent methods, they also rely on fast computations (or rather approximation) of derivatives. DTA models are either based on analytical relationships or on simulation of traffic conditions. Using DTA models for evaluating the road users’ response to a pricing scheme is, however, more computationally expensive (as in time consuming) in comparison with static traffic assignment models. Previously developed optimization methods for the static case are therefore not suitable to use together with DTA. For the DTA case, a grid search method for optimizing a single time dependent toll in a hypothetical three link network is presented in [17], and a surrogate-based optimization method with a simulation-based DTA is applied to a highway network model of Maryland in [18].

Another approach is adopted in [19] for combined optimization of toll locations and their individual toll levels. Instead of performing the optimization with the DTA model directly, toll location and levels are in a first step optimized with a static traffic model (STA) [20]. In a second step they are evaluated in a DTA model (the Silvester/Contram model [21]), and the social surplus is computed more accurately for the selected cordon design. This approach allows for combined optimization of the cordon structure and toll levels, and shows the applicability of combining STA and DTA models in the process of optimizing and evaluating toll locations and levels. Assuming that the DTA model is able to more accurately model the spatial and temporal distribution of congestion, the toll locations and levels based on the static model can most likely be further improved. Thus there is a need for tuning the toll levels in order to further improve the social surplus when it can be more accurately estimated using a DTA model.

While second-best pricing is a difficult optimization problem, both in the static and the dynamic case, MSCP tolls can in the static case be computed by solving a convex optimization problem, if all users are assumed to have the same value of time (only one user class). In the dynamic case, time dependent MSCP tolls are computed for analytical DTA models in [17, 22, 23]. For the simulation-based DTA models, approximately time dependent MSCP tolls are computed in [24]. When introducing congestion pricing there are, however, practical considerations which makes the application of MSCP rather limiting.

In [18] encouraging results are reported for the application of surrogate-based optimization with a simulation-based DTA model, for optimizing toll levels in a cordon pricing scheme. Similar approaches have earlier been developed for road pricing with a DTA model [25] and for a robust pricing problem with a STA model [26]. Surrogate-based optimization, for example in terms of response surfaces, is commonly used for optimization problems with expensive-to-evaluate objective functions. The surrogate model is used for approximating the costly function, and the optimization is then performed on the surrogate model instead. The performances of optimization methods based on surrogate models are, however, dependent on three components; experimental design, infill strategy and choice of surrogate model itself. The experimental design will give the initial set of toll levels for which the costly function needs to be evaluated, the infill strategy iteratively determine additional sets of toll levels to be evaluated by the costly function, and the choice of surrogate model will give the functional form to be fitted to the sampled sets of toll levels. For a thorough review on the application of surrogate-based optimization in traffic modeling and planning related problems we refer to [18].

While some very interesting insights on the applicability of surrogate models for optimizing toll levels within a DTA model are presented in [18], their traffic model does not include travel demand modeling (TDM). With reduced or increased travel times/travel costs it is well known that the travel demand will be affected, and a major part of the socio-economic benefits for introducing congestion pricing is related to reduction of the travel demand. Thus, bringing in TDM improves the realism of the traffic model and allows for a more accurate computation of the social surplus. In both Stockholm and Gothenburg the main adaptation strategy was work trips changing from car to public transport [27]. The introduction of travel demand modeling within a DTA model is therefore important, but will increase the computational time. For the TDM-DTA model used in this paper the computational time is approximately ten hours for evaluating one set of toll levels. Thus to sample 67 initial points, and 30 additional infill points, as is done in [18], is not feasible.

This paper focuses on evaluating the performance of a surrogate-based optimization framework for optimizing toll levels in cordon pricing schemes, when the number of possible evaluations of the costly objective function is limited to a small number. To be able to evaluate a large number of different configurations of the surrogate-based optimization framework, we will use a STA model based on user equilibrium route

choice, including a TDM for the Stockholm region. The use of the STA model also allows us to use stochastic-search methods (meta-heuristics) to compute close to global optimal toll levels. This would not be possible using a DTA model for a network of similar size. Based on these evaluations, the surrogate-based optimization framework will be used on a recently developed DTA model for the Stockholm region which is based on VisumDUE analytical traffic assignment software and Regent demand model [28]. The traffic models used within this work have been roughly calibrated for actual traffic conditions in the Stockholm region, but not to the extent that they can be used as a basis for actual policy decisions on congestion pricing.

The main contributions from this paper are the design and evaluation of the surrogate-based optimization framework when the number of different sets of toll levels to evaluate is limited to a small number and when TDM and DTA are combined. This is, to our knowledge, the first attempt to evaluate a methodology for road toll optimization, with a combined TDM-DTA model, for a large scale traffic network.

The remainder of the paper is outlined as follows. In Section 2 the toll level setting problem is presented and followed by the methodological approach in Section 3. Evaluation of the surrogate-based optimization framework with the STA model is presented in Section 4. Based on the evaluation with the STA model, a configuration to be used with the DTA model is chosen and applied in Section 5. Finally the results are discussed in Section 6, and in Section 7 we conclude our findings and suggest further research directions.

2 The toll level setting problem

This paper focuses on the toll level setting problem (TLP), which is to find optimal toll levels for a set of cordon tolls in a congestion pricing scheme with the objective to maximize the social surplus (S). The cordon design is assumed to be predetermined, and it is only the toll levels that are being optimized. The methodology presented in this paper should, however, be possible to extend to level setting problems in other forms of congestion pricing schemes, as long as there are suitable traffic models for evaluating them, for example distance-based or access-based pricing schemes. Thus the costly function is in our case the social surplus measurement, and the decision variables to adjust are the different toll levels. These toll levels are usually bounded from below by zero, and from above by a maximum toll level which the government/operator has decided.

The social surplus measure is the sum of the consumer surplus (CS) plus the toll revenues (R) minus additional external effects [29]. For simplicity we will assume that both the consumer surplus and the revenues are given in the unit of SEK. If the consumer surplus is given in travel time, the value of time (VOT), possibly differentiated for each user class, can be used to transform time into a monetary unit. In this paper we will neglect additional external effects, and the *change* in social surplus ΔS from the no-toll

scenario is expressed as

$$\Delta S = \Delta CS + \Delta R,$$

where ΔCS and ΔR imply the change in consumer surplus and revenues. The consumer surplus is the consumer benefits minus the consumer costs, and if the travel demand is given by a discrete choice travel demand model, the consumer surplus can be computed by the logsum [30]. For fixed travel demand, the change in consumer surplus is simply equal to the change in total travel costs.

Let \mathcal{C} be the set of defined cordons, and \mathcal{M} the set of time periods, for which the toll levels are differentiated. The matrix τ , then consists of cordon toll levels for each combination of cordon and time period, τ_{cm} with $c \in \mathcal{C}$ and $m \in \mathcal{M}$. Note that for a time static traffic model τ is reduced to a vector of cordon toll levels for the modeled time period. The set of feasible cordon toll levels are given by

$$\mathcal{T} : \{ \tau | \tau_{cm}^L \leq \tau_{cm} \leq \tau_{cm}^U, c \in \mathcal{C}, m \in \mathcal{M} \},$$

where τ_{cm}^L and τ_{cm}^U are the lower and upper bound for each cordon toll and time period.

For simplicity we let the traffic state, in terms of network travel costs, flows and demand be represented by \mathbf{u} . This representation can easily be extended to multiple user classes and time periods. While the computation of the social surplus measure itself is trivial, it requires knowledge about the traffic state corresponding to a given set of toll levels, which is computed using a traffic model. The cordon tolls, τ , are added to the generalized link costs in the traffic model. Thus, the computational time for evaluating one set of toll levels is depending on the computational effort required for computing the corresponding traffic state as function of toll levels, $\mathbf{u}(\tau)$. The change in social surplus can then be formulated as a function of the traffic state

$$\Delta S(\mathbf{u}(\tau)) = \Delta CS(\mathbf{u}(\tau)) + \Delta R(\mathbf{u}(\tau)), \quad (1)$$

where the consumer surplus and the revenue are now implicit functions of the toll levels. The optimization problem can be expressed as

$$\max_{\tau \in \mathcal{T}} \Delta S(\mathbf{u}(\tau)). \quad (2)$$

3 Methodology

Surrogate-based optimization is commonly applied to low dimensional engineering design problems, which either have a black-box objective function or an objective function which is costly (as in computational time). In this section we present the surrogate-based optimization framework, and its components, which will later be applied to the TLP.

3.1 The surrogate-based optimization framework

The main components of the framework are the costly objective function (in our case (1), the social surplus), the response surface (the surrogate model itself), the initial sampling (determined by the experimental design) and the infill sampling. An overview of the surrogate-based optimization framework is given in Figure 1. Note that the infill sampling is an iterative process which will make use of the fitted surrogate model to various degrees, but the actual optimization of toll levels over the surrogate model is not necessarily done in each iteration. In contrast to [18], this paper focuses on the application of a surrogate modeling framework to a road pricing problem in which the possible number of objective function evaluations is very limited. Based on the computational time from running the DTA model it is reasonable to evaluate somewhere between 20 and 40 sets of toll levels in the costly objective function, that is, the sum of the initially sampled points and infill sampled points is in the region of 20 to 40 samples.

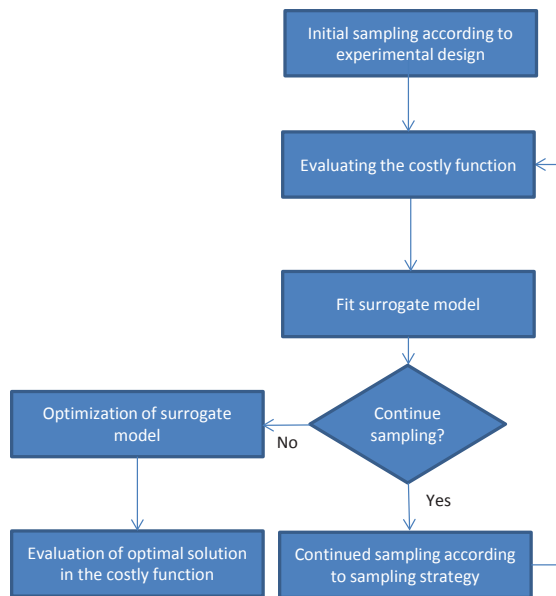


Figure 1: Conceptual idea of the surrogate-based optimization framework.

The response surface is used for predicting the costly function values for points (in our case a certain combination of cordon toll levels) which have not yet been sampled, and will thus provide an inexpensive approximation of the costly objective function; in fact, one gets both an analytical expression of the response surface itself as well as for its derivatives. Hence, instead of performing the optimization on the costly function itself, the optimization can be performed on the response surface. The response surface will in general be a non-convex function, and finding its optimal solution is thus not trivial either, although the evaluation of a non-sampled point can be done with very little computational effort. Here a multi-started simulated annealing method has been

used for finding the global optimum of the response surface, denoted s^* . Both linear and non-linear constraints can be included in the framework. For the problem studied in this paper, only box constraints on the toll level variables are required. Such constraints can easily be included in the standard methods used for finding the optimum of the response surface. Figure 2 illustrates a costly function and corresponding surrogate models based on 25 and 100 samples points, respectively.

The experimental design is used for choosing an initial set of points to sample, for which the costly objective function will be evaluated. In addition to these points, a set of pre-determined points can optionally be specified. Given the initial sampling, the response surface can be constructed. Commonly used response surfaces are radial basis functions (RBF) [32], and Kriging models (sometimes referred to as DACE models) [33, 34]. While the initial sampling is used for constructing the response surface, additional points can be sampled in order to iteratively improve the approximation of the costly function. The generation of infill samples is determined by the infill strategy, and is commonly both depending on where in the variable domain the response surface predict good solutions to be found (local search strategy) as well as of what part of the domain has been less explored (global search strategy). Initial samples and infill samples constitute the total number of sampled points.

The performance of the framework is sensitive to the problem dimension. In the field of surrogate-based optimization, problems with more than 15 variables are considered as high-dimensional problems [35]. The type of surrogate models evaluated in this paper are suitable for low and moderate dimension problems, and for the basis of our evaluation we consider a problem with six variables (same dimension as in [18]). We, however, extend our evaluation to the case of ten variables as well.

In standard textbook literature, surrogate-based optimization is commonly presented for a minimization problem. Even though the TLP is a maximization problem we will in this section present the surrogate-based optimization framework for a minimization problem. As a basis for the implementation of the surrogate modeling framework pre-

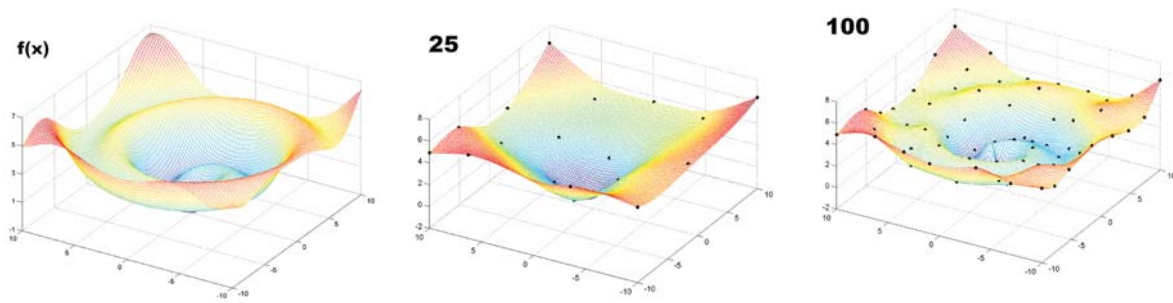


Figure 2: The leftmost picture is the true costly function $f(x)$ to be optimized. The following pictures are the surrogate models based on 25 and 100 sampling points n , respectively. Source [31].

sented in this paper, the Surrogate Model Optimization Toolbox for MATLAB [36] has been used.

3.2 Response surfaces

In this section we introduce two type of functions commonly used as response surfaces, the Radial Basis Function (RBF) interpolation [32] and the Kriging model interpolation [33, 34], applied to a costly function, $f(x)$ with $x \in \mathbb{R}^d$, and box constraints given by $x \in \Omega$.

3.2.1 Radial basis functions

Given n distinct points $\mathbf{x} \in \Omega$, with the evaluated function values $\mathbf{y} = f(\mathbf{x})$, the radial basis function interpolant $s(x)$ has the form

$$s(\bar{x}) = \sum_{i=1}^n \lambda_i \cdot \phi(\|\mathbf{x}^{(i)} - \bar{x}\|_2) + b^T x + a, \quad (3)$$

with $\lambda \in \mathbb{R}^n$, $b \in \mathbb{R}^d$, $a \in \mathbb{R}$, where ϕ is commonly chosen as the cubic spline $\phi(r) = r^3$ or the thin plate spline with $\phi(r) = r^2 \log r$. The unknown parameters λ , b and a are obtained as the solution of the system of linear equations

$$\begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ c \end{pmatrix} = \begin{pmatrix} \mathbf{y} \\ 0 \end{pmatrix}, \quad (4)$$

where Φ is the $n \times n$ matrix with $\Phi_{ij} = \phi(\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|_2)$ and

$$P = \begin{pmatrix} x_1^T & 1 \\ x_2^T & 1 \\ \vdots & \vdots \\ x_n^T & 1 \end{pmatrix}, \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}, c = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_d \\ a \end{pmatrix}, \mathbf{y} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}. \quad (5)$$

If the rank of P is $d + 1$, then the matrix $\begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix}$ is nonsingular and the linear system (4) has a unique solution [37]. Thus we have a unique RBF interpolation function $s(x)$ to the costly $f(x)$, defined by the points \mathbf{x} , if $n \geq d + 1$.

3.2.2 Kriging models

Kriging models¹ describes every point in the interpolation function as a realization of random variables, normally distributed with mean μ and variance σ^2 . Kriging models

¹The application of Kriging models as surrogate models was first discussed in a paper with title ‘‘Design and Analysis of Computer Experiments’’ (DACE) [33], thus Kriging models designed for use as surrogate models are sometime referred to as DACE models.

can be included in the theory of RBF interpolation functions [38], but interpreting Kriging models in terms of mean value and standard deviation allows for the development of more advanced infill strategies.

Estimates of μ and σ^2 are obtained from Maximum Likelihood Estimation (MLE) with respect to the n sampled points \mathbf{x} and their corresponding function values $\mathbf{y} = f(\mathbf{x})$. Using a matrix \mathbf{R} of correlation values between sampled points, the Kriging interpolant for a non-sampled point \bar{x} is defined by

$$s(\bar{x}) = \mu + \mathbf{r}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\mu) \quad (6)$$

where \mathbf{r} is the vector of correlations between \bar{x} and \mathbf{x} , μ is the estimated mean, and $\mathbf{r}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\mu)$ represents the adjustment to this prediction based on the correlation of sampled points \mathbf{x} . The correlation function is defined as

$$\text{Corr} [\mathbf{x}^{(i)}, \mathbf{x}^{(j)}] = e^{-D(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})} \quad (7)$$

where $D(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ is a distance formula. Compared with Euclidean distance, where every variable is weighted equally, we will utilize the distance formula from [33]

$$D(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sum_{k=1}^d \theta_k \cdot \left| x_k^{(i)} - x_k^{(j)} \right|^{p_k} \quad \theta_k > 0, p_k \in [1, 2] \quad (8)$$

which is designed to capture functions more precise. The exponent p_k is related to the smoothness of the function in the k th dimension. Values of p_k near 2 corresponds to smooth functions and values near 1 to less smoothness. Parameter θ_k controls the impact of changes in variable x_k .

Note that when the distance between $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ is small, the correlation is close to one. When the distances increase, the correlation approaches zero. A large value for θ will affect the distance to grow faster, which leads to a decrease in the correlation. Thus, different dimensions can be given different weights, allowing for a more accurate correlation function.

3.3 Merit functions

To determine additional points to sample, where the costly objective function should be evaluated, merit functions are used. The only qualifications needed in a merit function are the ability to identify unexplored regions and/or investigate promising areas. A purely global merit function would be to find the point most distant from all sampled points, that is, maximizing the minimal distance to sampled points. The other extreme would be to have a purely local merit function. Such a function would find s_{min} , the global minimum of the surrogate model, and choose this as the new point to sample.

A key factor here is to balance these ambiguous goals, to somehow find both local and global points. The global search is necessary in order to avoid getting stuck in a local minima, but will be inefficient for finding the global optimal solution to the costly

function. Infill strategies use a merit function to determine the next point to sample, based on information from an already fitted response surface. They are commonly divided into one-stage and two-stage methods. In a one-stage infill strategy the search for infill points does not make use of the fitted response surface, while a two-stage strategy in the first stage fit the response surface to already sampled points, and in the second stage make use of the fitted model to decide where to sample the next point.

For this paper we will evaluate two different two-stage methods. The first one is candidate point sampling [39], in which candidate points are sampled and evaluated both based on their proximity to already sampled points, and on their predicted objective function value when evaluated in the surrogate model. This method is applicable to both RBF and Kriging interpolation functions. The second one is Expected Improvement (EI) sampling [34], which makes use of the interpretation of Kriging models in terms of mean value and standard deviation. This sampling strategy can therefore only be applied with Kriging models.

It is possible to develop algorithms that suggests multiple infill points in each iteration, see for example [40, 41], which is very efficient whenever the costly function evaluations can be performed in parallel. For reasons explained later in Section 4, this is not possible for the scope of this paper.

3.3.1 Candidate point sampling

Candidate point (CAND) sampling evaluates a large number of points based on their value when being evaluated in response surface, and on their distance to previously sampled points. The surrogate model objective function value and the distance to the closest sampled point are weighted together, and the candidate point with the minimum candidate values is then chosen. Thus a trade-off between local and global search can be obtained. Also note that this sampling strategy can be used with any surrogate model.

To use both objective function value and distance in the same candidate value, requires scaling. Let $K(\tilde{x})$ denote the candidate value at \tilde{x} , and let \hat{x}_1 and \check{x}_1 be the candidate points with maximum and minimum surrogate objective function values. The scaled objective function value is $K_1(\tilde{x}) = (s(\tilde{x}) - s(\check{x}_1)) / (s(\hat{x}_1) - s(\check{x}_1))$.

Let $G(\tilde{x})$ denote the Euclidean norm between \tilde{x} and the closest already sampled point. The candidate points with longest and shortest distance to the closest already sampled point are denoted \hat{x}_2 and \check{x}_2 respectively. The scaled distance $K_2(\tilde{x})$ is then given by $K_2(\tilde{x}) = (G(\hat{x}_2) - G(\tilde{x})) / (G(\hat{x}_2) - G(\check{x}_2))$. Using weights w_1 and w_2 the weighted candidate value is given as $K(\tilde{x}) = w_1 K_1(\tilde{x}) + w_2 K_2(\tilde{x})$.

A large value on w_1 and a small value on w_2 will favor candidate points in areas where the response surface returns low function values, and with an opposite weight setting, candidate points in less explored areas will be favored. To have an infill strategy sampling points in both areas, we will use an iterative scheme for selecting values on w_1 and w_2 . Starting with $w_1 = 0$ and $w_2 = 1$ we will increase w_1 by 0.1 each time a new point is sampled, and update $w_2 = 1 - w_1$. When w_1 reaches 1, it will be reset to 0, and w_2 reset to 1.

Finally we need to decide how to sample the candidate points. Two sets of candidate points will be sampled. The first set is created by sampling $1000d$ points by a small random perturbation of the so far best found point and the second set is created by sampling another $1000d$ points uniformly in the solution space. Thus we will ensure to both sample points where we know that good solutions have already been obtained, and in less explored areas.

3.3.2 Expected improvement sampling

With a Kriging model we both get an estimate of the costly objective function value, $s(\bar{x})$ in the point \bar{x} , and an estimate of the uncertainty $\sigma(\bar{x})$. Let f_{\min} be the lowest found objective function value for the costly function. The improvement at a point \bar{x} from the best found solution f_{\min} is then defined as

$$I(\bar{x}) = \max [0, f_{\min} - f(\bar{x})]. \quad (9)$$

Equation (9) states that if we evaluate a new point, with objective function value $f(\bar{x}) < f_{\min}$, then an improvement equal to $f_{\min} - f(\bar{x})$ has been reached, otherwise the improvement is 0.

Not knowing the actual costly function value at \bar{x} we can still make use of information from the Kriging model to estimate the expected improvement, $ExpI(\bar{x})$, at the point \bar{x} . The expected improvement can be formulated as in [34]

$$ExpI(\bar{x}) = \begin{cases} (f_{\min} - f(\bar{x})) \Phi \left(\frac{f_{\min} - f(\bar{x})}{\sigma(\bar{x})} \right) \sigma(\bar{x}) \Psi \left(\frac{f_{\min} - f(\bar{x})}{\sigma(\bar{x})} \right) & \text{if } \sigma(\bar{x}) > 0 \\ 0 & \text{if } \sigma(\bar{x}) = 0, \end{cases} \quad (10)$$

where $\Phi(\cdot)$ and $\Psi(\cdot)$ are the probability density function and cumulative density function respectively.

With EI sampling, the next point to sample, \hat{x} , is given by $\hat{x} = \arg \{ \max_x ExpI(x) \}$. The maximization of $ExpI(x)$ is however, in general, a non-convex continuous optimization problem. To solve this problem we will apply a multi-started simulated annealing algorithm.

Note that whenever we evaluate $ExpI$ at an already sampled point, $x^{(i)}, i \in 1 \dots n$, $\sigma(\bar{x})$ will be zero, and the expected improvement becomes zero. Thus by maximizing $ExpI(x)$ a point will never be sampled more than once.

The first term in $ExpI$ is favoring points at which the response surface predicts good solutions, while the other term is favoring points in which the uncertainty is high. Thus this measurement provides a balance between a local and global behavior, when being used as merit function in an infill strategy.

One potential problem using EI sampling is that $\sigma(x)$ is an estimated standard deviation, and with few sampled points, $\sigma(x)$ will give an underestimation of the actual standard deviation, thus underestimating the uncertainty at point x [42]. The effect of an underestimation of the standard deviation will be a more local behavior of the EI

sampling, since the second term, favoring global exploration, will be smaller. Therefore a correction factor has been applied, presented in [42], to improve the estimation of the actual standard deviation. The correction factor uses leave-one-out cross validation (which will later be discussed in Section 3.5.1) to approximate the underestimation, and compute a factor (> 1) to be multiplied with the standard deviation. This factor does not take into account that the quality of the approximation will differ over the domain.

3.4 Experimental design

All surrogate-based algorithms need an initial set of sample points in order to get going. To build the first response surface, a minimum of $n \geq d + 1$ sample points is required where d is the dimension of the problem to be solved. For example, if we are to fit a model for a TLP with six variables, we need at least seven initially sampled toll levels. So how should one choose these initial points? The procedure of choosing this initial set is often referred to as Experimental Design, or sometimes Design of Experiments. An experimental design in which the points are uniformly spread will result in a response surface which has the property that the model accuracy is also uniformly spread. This is called the space-filling property of the experimental design and is important for the performance of the surrogate model optimization framework. Latin hypercube design (LHD) possess good space-filling properties [43], and by maximizing the minimum distance between the points we can ensure that the experimental design results in sampled points which are also separated as much as possible in any suitable norm. Such a design is referred to as maximin LHD (MLHD).

In addition to MLHD, [44] shows the benefit of using additional corner points of the search space, as well as using interior points with low objective function values (for a minimization problem). This would, however, greatly increase the number of initial samples, and in this paper we have instead used a combination of a few corner and interior points.

3.5 Measurements of model accuracy

To measure the model accuracy, the leave-one-out cross-validation technique has been applied [45]. We will in this section also discuss the use of the estimated global optimum and global optimum when measuring model accuracy.

3.5.1 Leave-one-out cross-validation

The basic principle behind this approach is to leave one of the sampled points out, re-fit the response surface, and then compare the estimate from the re-fitted model, with the actual measurement from the sampled points. In [46] it is noted that leave-one-out cross-validation gives a measurement of the model sensitivity to lost information. Solely using this measurement may result in acceptance of a model as accurate while it is only insensitive to lost information, or we reject an accurate model because it is sensitive to

lost information. The benefit of using leave-one-out cross validation is that it does not require any additional sampling, which is crucial when evaluating the performance of the surrogate-based optimization framework with a truly costly function.

Three different measurements have been computed from the leave-one-out cross validation; normalized root mean square error (NRMSE), normalized maximum absolute error (NMAE) and Pearson correlation coefficient (PCC). NRMSE gives an estimate of the mean error over the complete domain, thus measuring the overall accuracy of the surrogate model, and NMAE gives an estimate of the maximum error, thus measuring the local accuracy. PCC gives the correlation between true and estimated objective function values at the sampled points, from leave-one-out cross validation. Thus it would be desirable to have a model with low NRMSE and NMAE, and high PCC values [47].

For a set of n sampled points the NRMSE is given by

$$\sqrt{\frac{\sum_{i=1}^n (s(x_i) - f(x_i))^2}{\sum_{i=1}^n (s(x_i))^2}},$$

and NMAE is given by

$$\frac{\max_{1 \leq i \leq n} |s(x^{(i)}) - f(x^{(i)})|}{\frac{1}{n} \sum_{i=1}^n \left(s(x^{(i)}) - \frac{1}{n} \sum_{j=1}^n f(x^{(j)}) \right)^2}.$$

The PCC value is given by

$$r = \frac{n \sum_{i=1}^n s(x^{(i)}) f(x^{(i)}) - \sum_{i=1}^n s(x^{(i)}) \sum_{i=1}^n f(x^{(i)})}{\sqrt{n \sum_{i=1}^n (s(x^{(i)}))^2 - (\sum_{i=1}^n s(x^{(i)}))^2} \sqrt{n \sum_{i=1}^n (f(x^{(i)}))^2 - (\sum_{i=1}^n f(x^{(i)}))^2}}.$$

3.5.2 Global optimum and estimated global optimum

For the STA model, global optimum is computed using MATLABs simulated annealing method for multiple starting points. The resulting objective function value (f^*) cannot be guaranteed to actually be global optimal, but we believe it is reasonably close to the global optimal solution. By evaluating the optimal toll level solution from the optimization on the response surface in the costly objective function, we can compare this function value with f^* . This gives a measurement of how good solution that we can produce with the surrogate-based optimization framework for a given combination of response surface, infill strategy and experimental design. For the Regent/VisumDUE model no global optimum can be estimated, and thus f^* cannot be computed and used in the evaluation process.

Optimizing the toll levels over the response surface results in an estimate of the global optimal objective function value (denoted s^*), which can then be compared with the same toll levels evaluated in the costly function. It is, however, difficult to use this comparison in the evaluation process. Let us for instance assume that the optimum of the response surface correspond with one of the sampled points. The optimal response

surface function value and the costly objective function value will then be equal, and it is easy to determine that we have actually found the true global optimum. This is a false conclusion since there could be a non-sampled point with a better costly objective function value.

4 Evaluation of the surrogate-based optimization framework

The surrogate-based optimization framework offers a variety of possible combinations of response surfaces and sampling strategies. Our aim is to apply the framework to the Regent/VisumDUE model [28] of the Stockholm region, which requires approximately ten hours of computational time to evaluate one set of toll levels. The computational time limits the number of costly function evaluations which are reasonable in a practical application, and we will therefore focus our attention to surface-sampling combinations which can produce high quality solutions within 20 to 40 costly function evaluations. It is in the case of the TDM-DTA model that function evaluations are costly, and our evaluation of the framework is therefore carried out in two steps, similar to [18].

In the first step a large number of different combinations of response surfaces, experimental designs and infill sampling strategies are evaluated using a static traffic model with user equilibrium route choices and a binomial logit demand model for the Stockholm region. While both the static and dynamic models represent the Stockholm region, the STA model requires approximately six minutes for evaluating one set of toll levels, and allows for a large number of combinations to be evaluated. On the other hand, the STA model lacks a detailed road network and the realism that comes from the spatial and temporal distribution of congestion in the VisumDUE model. With the STA model over 16 000 costly function evaluations have been carried out during 180 hours. To use the dynamic model for this number of function evaluations would have required over 22 years of computational time. The implementation of the STA model also allows for efficient usage of parallel processing, which is not possible with our current setup of the Regent/VisumDUE model.

In the second step we choose the most promising combination, and evaluate it with the Regent/VisumDUE model. Another benefit from using the STA model in the first step is that other global optimization methods, such as genetic algorithms and simulated annealing, can be applied for comparison. Thus we can compare the solutions produced by the surrogate-based optimization framework with close to global optimal solutions.

4.1 The static traffic model

The STA model of the Stockholm region uses an aggregated traffic network of Stockholm (Figure 3), with 392 links and 40 zones (1 560 OD-pairs). It models the morning rush hour, for a single user class with VOT equal to 1.2 SEK per minute. Route choice is modeled by a Wardrop equilibrium and travel demand is given by a multinomial logit

model, including the choices of travel by car, by public transport or not traveling at all. The demand model is roughly calibrated for the aggregated network. Therefore the model should not be used for actual policy evaluation purposes for the Stockholm region, but for evaluating the surrogate-based optimization framework we believe it is a good representation of a congested urban area, including the complexity of travel choices related to both route and mode. The model has been used in several evaluations of road toll optimization methods [20, 48], and a detailed description of the model is given in [20].

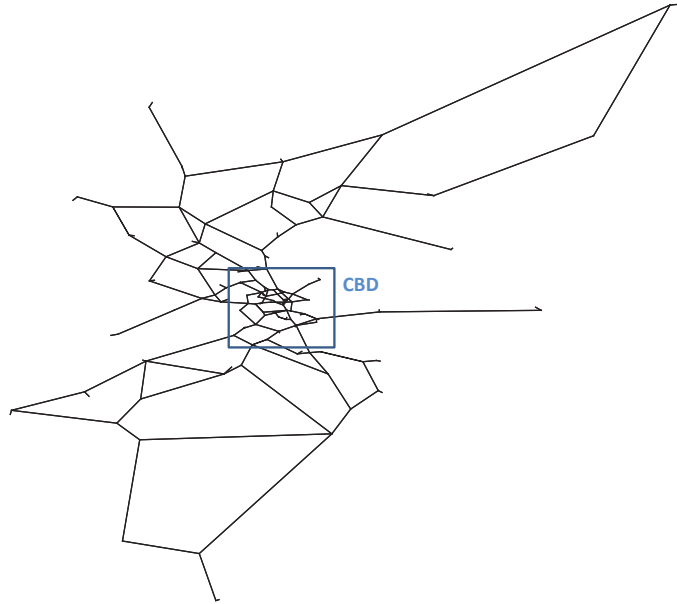


Figure 3: Stockholm network model used in the static traffic model.

The change in social surplus is given by (1). For the STA model, the change in consumer surplus, ΔCS , is given by the logsum from the multinomial logit demand model, and the change in revenues, ΔR , is given by the sum of the collected tolls. For a more detailed description of the computation of the social surplus measurement for this model see [20].

Comparing STA and DTA for a case when a congestion pricing scheme is introduced in a road traffic network, one can expect the improvement of the social surplus to be larger when the road users response is evaluated with the DTA (given that it is a well-designed pricing scheme) [9]. Here, we, however, use an STA model with an aggregated network, in comparison with a much more detailed network with the DTA model. Thus we will have the problem of overestimating the congestion on some roads in the network [20], and therefore the maximum improvement of the social surplus is actually higher with the STA model in our case.

4.2 Experimental setup

Two cordon pricing scenarios for Stockholm will be considered in the evaluation process. The first one, Scenario 1, consist of three cordons; the ring toll currently implemented in Stockholm, tolls on the bypass highway and a toll cordon covering the inner city bridges. This set of cordon tolls corresponds to one of the possible extensions of the current pricing scheme which has been discussed [49]. For each cordon, the toll levels are differentiated based on direction, resulting in a total of six different toll level variables. The second scenario, Scenario 2, is based on the current toll ring in Stockholm and the bypass highway. Here the current toll ring is further differentiated into four sub-cordons, each with their individual toll level, differentiated depending on direction. This results in a total of eight differentiated toll levels for the toll ring, and additionally two tolls for the bypass highway. For both scenarios the maximum toll level of each cordon is set to 60 SEK.

Scenario 1 will be the scenario primarily used for the evaluation of the surrogate-based optimization framework, and this is also the scenario which will be used when evaluating the framework with the DTA model. Scenario 2 gives an optimization problem with higher dimension, and will primarily be used for sensitivity analysis of the performance of the framework with respect to the problem dimension. The two scenarios are illustrated in Appendix A.

Setup for Scenario 1

In [18] response surfaces which can handle noisy costly functions are evaluated. For the traffic models used within this work, this is not necessary since they are deterministic.

Four different response surfaces have been evaluated; Kriging models with exponential ($p = 1$), Gaussian ($p = 2$) and generalized exponential ($1 \leq p \leq 2$) correlation functions, and RBF with cubic interpolation function. Since the costly objective function is known to be non-convex we have chosen to only evaluate global optimization infill strategies, thus rejecting the suboptimal approaches used in [18].

The infill strategies which has been evaluated are the candidate point (CAND) and the expected improvement (EI) sampling strategies. CAND sampling has been applied with all response surfaces, and EI sampling with the correction factor² has been applied with all Kriging models.

Four different experimental designs have been evaluated for each combination of surrogate model and infill sampling strategy. The experimental designs used are all of maximin LHD (MLHD) type, based on seven and fourteen sampled points (sets of toll levels), and optionally augmented with five points corresponding to uniform toll levels; two additional corner points (zero and maximum toll levels) and three interior points (25%, 50% and 75% of maximum toll levels). We denote these experimental designs ED- x , $x \in \{7,14,7U,14U\}$, where the number is the sample size and U indicates the

²Preliminary evaluations without the correction factor showed a more local optimizing behavior and this configuration was therefore not included in the full evaluation.

augmented designs. The minimum number of samples required to fit an RBF is 7, and according to [36] the recommended minimum number of sampled points is 14. The complete set of combinations evaluated in Scenario 1 are listed in Table 1.

Table 1: Evaluated combinations of experimental design, surrogate model and infill strategy.

Number	Surrogate model	Infill strategy	Exp. Design
	RBF		ED-
1-4	Cubic	CAND	7, 7U, 14, 14U
	Kriging		ED-
5-8	$p = 1$	CAND	7, 7U, 14, 14U
9-12		EI	7, 7U, 14, 14U
13-16	$p = 2$	CAND	7, 7U, 14, 14U
17-20		EI	7, 7U, 14, 14U
21-24	$1 \leq p \leq 2$	CAND	7, 7U, 14, 14U
25-28		EI	7, 7U, 14, 14U

Note that MLHD and CAND sampling include random sampling of toll levels. Therefore ten repetitions of the experiments have been carried out for each combination listed in Table 1. The same ten sets of random seeds have been used for each combination.

Additionally, for comparison with the results in [18], Scenario 1 has been evaluated with four different surrogate models, for 67 initially sampled toll levels and 30 additionally infill samples (CAND sampling for RBF and EI sampling for the Kriging models). Scenario 1 has the same problem dimension as the experiment with DynusT in [18], although the DynusT experiments include five toll level variables and a sixth variable to model off-peak tolls as proportional to peak toll levels.

Setup for Scenario 2

For Scenario 2, the main objective is to evaluate how sensitive the performance is to the problem dimension. Therefore the experiments have been limited to the Kriging model with Gaussian correlation function and EI sampling and to the RBF. Three different experimental design have been used; MLHD with 11, 22 and 33 samples, equal to $d + 1$, $2(d + 1)$ and $3(d + 1)$ respectively, extended with the five uniform samples from Scenario 1 (denoted ED-11U, ED-22U and ED-33U respectively). ED-11U and ED-22U correspond to ED-7U and ED-14U in Scenario 1. The third one (ED-33U) has been added to evaluate the performance with an even higher number of initial samples. As before, ten repetitions have been carried out for each combination.

4.3 Numerical results

It is natural that the main evaluation of the surrogate-based optimization framework is done by evaluating the toll levels which maximize the response surface in the costly objective function. The toll levels obtained from maximization of the response surface are denoted $\hat{\tau}$, and the corresponding costly objective function value is given by $\Delta S(\hat{\tau})$. Additionally, the PCC, NRMSE and NMAE values are reported.

Scenario 1

Results from the experiment with the Scenario 1 cordon layout are presented in Figure 4 and Figure 5. The global optimal objective function value, f^* , has been estimated to 607 130 SEK.

Comparisons with results from [18] are presented in Table 2. Here $\Delta S(\hat{\tau})$ is also included to illustrate the performance of the surrogate-based optimization framework when using a higher number of samples in the experimental design. For these comparisons only one set of experiments has been carried out, with the same random seed, due to the increased computational burden. Thus the initial set of sampled toll levels is equal for all four cases.

Table 2: PCC, NRMSE, NMAE, s^* and $\Delta S(\hat{\tau})$ after 67 and 97 sampled sets of toll levels (N), including comparison with results from [18].

Surrogate model	N	PCC	NRMSE	NMAE	s^*	$\Delta S(\hat{\tau})$
Kriging $p = 1$	67	0.89	0.27	2.46	601 051	590 942
Kriging $p = 2$		0.85	0.20	2.12	608 207	594 461
Kriging $1 \leq p \leq 2$		0.73	0.27	2.46	621 231	573 375
RBF cubic		0.96	0.10	1.00	610 891	596 657
Kriging $p = 2$ from [18]		-	0.03	2.26	-	-
Kriging $p = 2$ from [18] with noise		-	0.03	2.74	-	-
Kriging $p = 1$	97	0.95	0.18	2.11	618 294	604 770
Kriging $p = 2$		0.94	0.13	1.71	606 626	605 754
Kriging $1 \leq p \leq 2$		0.87	0.18	2.03	603 472	598 760
RBF cubic		0.98	0.07	0.78	599 181	599 702
Kriging $p = 2$ from [18] with noise		-	0.03	2.90	-	-

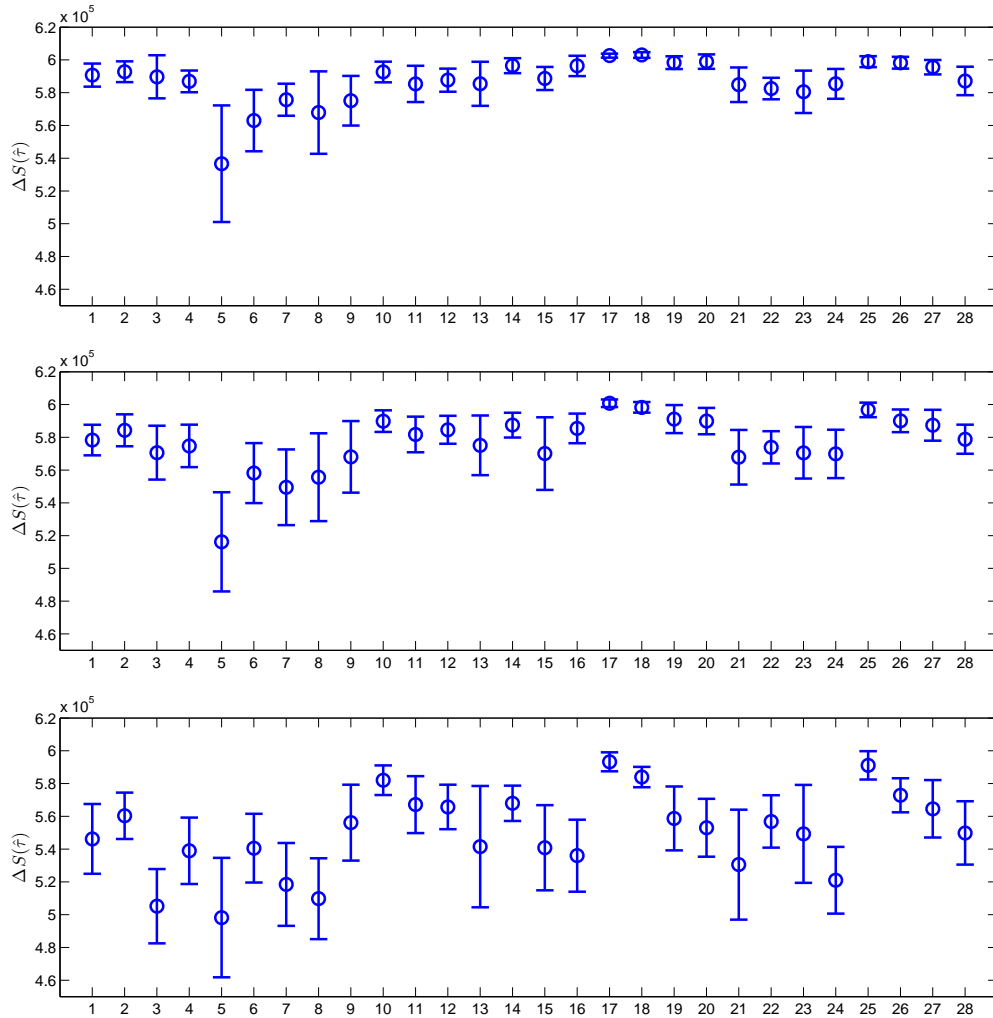


Figure 4: Mean social surplus for Scenario 1, $\Delta S(\hat{\tau})$ in SEK, after 20 (bottom), 30 (middle) and 40 (top) sampled sets of toll levels, together with 95% confidence intervals.

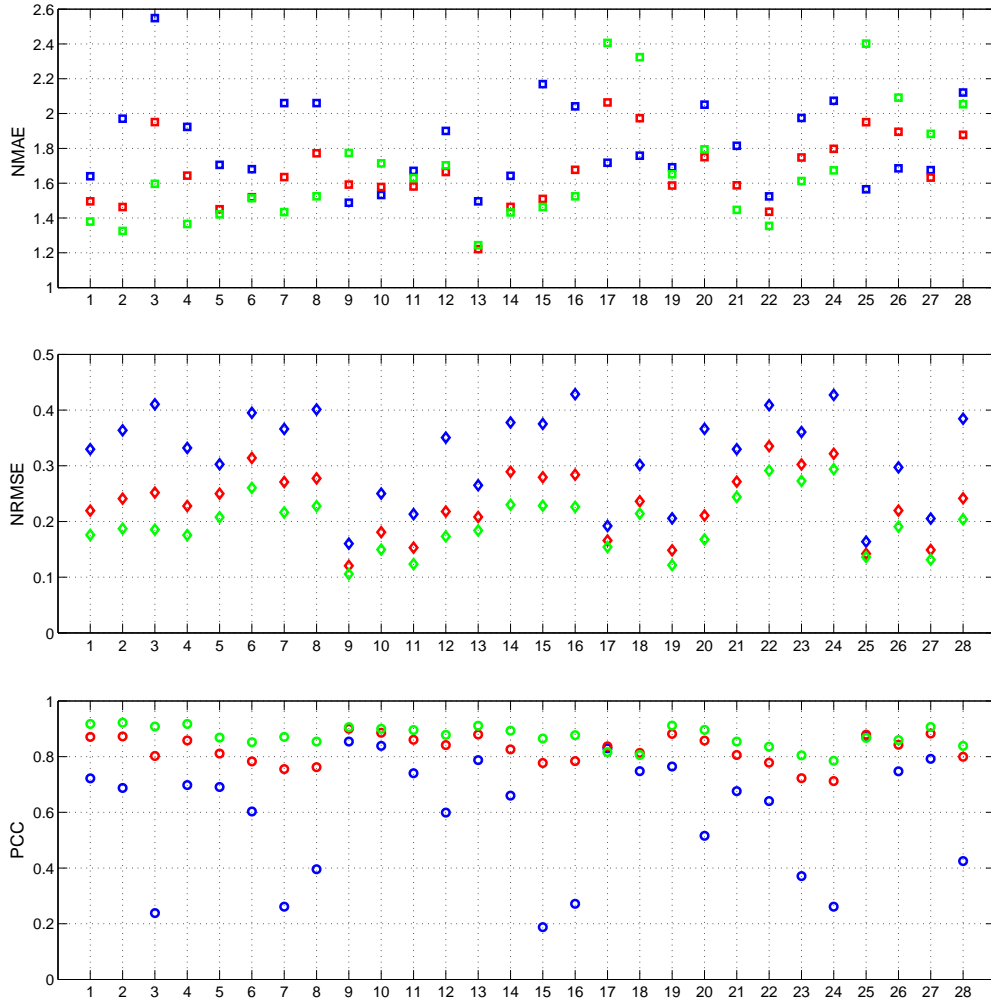


Figure 5: PCC, NRMSE and NMAE values for Scenario 1 after 20 (blue), 30 (red) and 40 (green) infill samples

Scenario 2

The global optimal objective function value (f^*) has been estimated to 580 195 for Scenario 2. Mean improvements of the social surplus are presented in Figure 6 after 40, 60 and 80 sampled sets of toll levels, and NRMSE, NMAE and PCC mean values are presented in Figure 6.

After a total of 80 sampled sets of toll levels, all surrogate models result in mean $\Delta S(\hat{\tau})$ around 97 to 98% of f^* . The differences between the best and worst performing configuration are larger after 40 sampled set of toll levels, where the Kriging model with ED-11U clearly the best performing one, followed by Kriging with ED-22U and RBF with ED11-U.

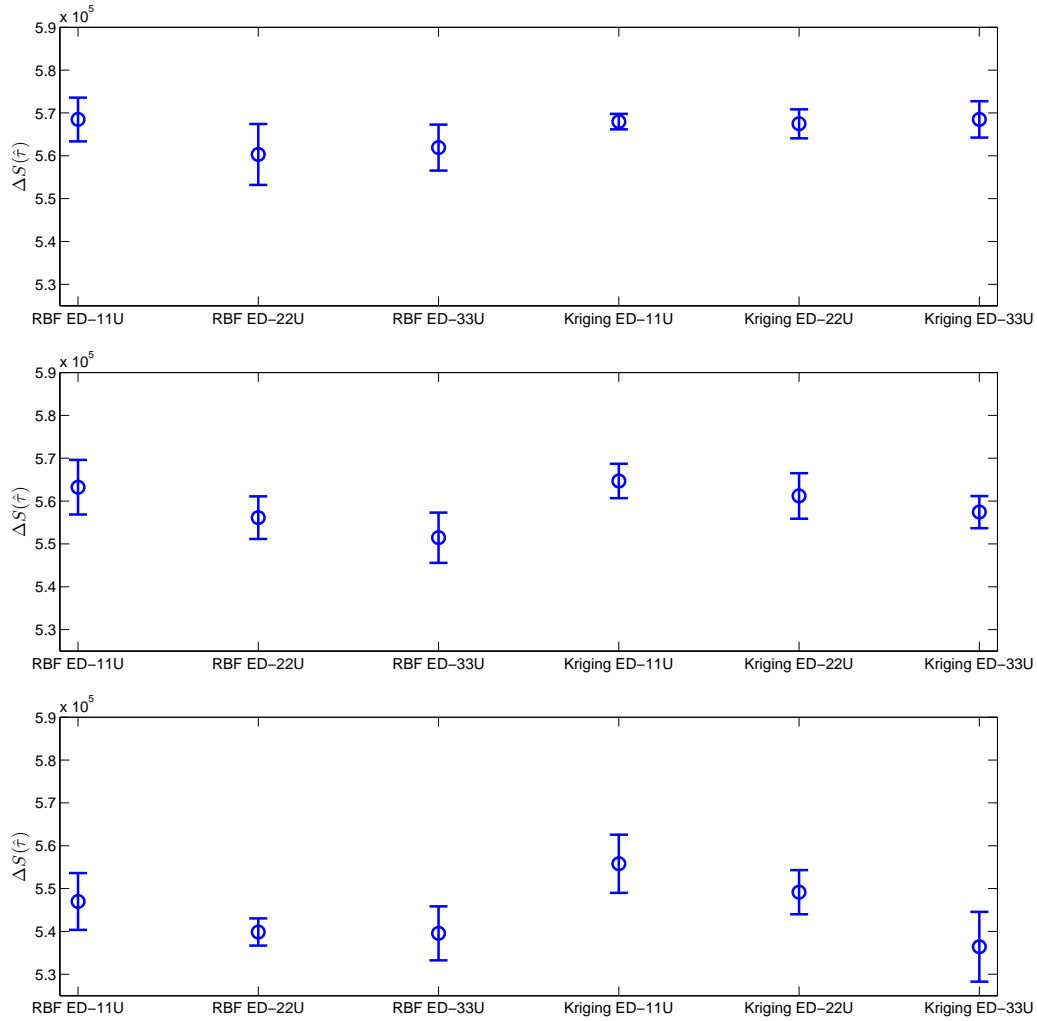


Figure 6: Mean social surplus for Scenario 2, $\Delta S(\hat{\tau})$ in SEK, after 20 (bottom), 30 (middle) and 40 (top) sampled sets of toll levels, together with 95% confidence intervals.

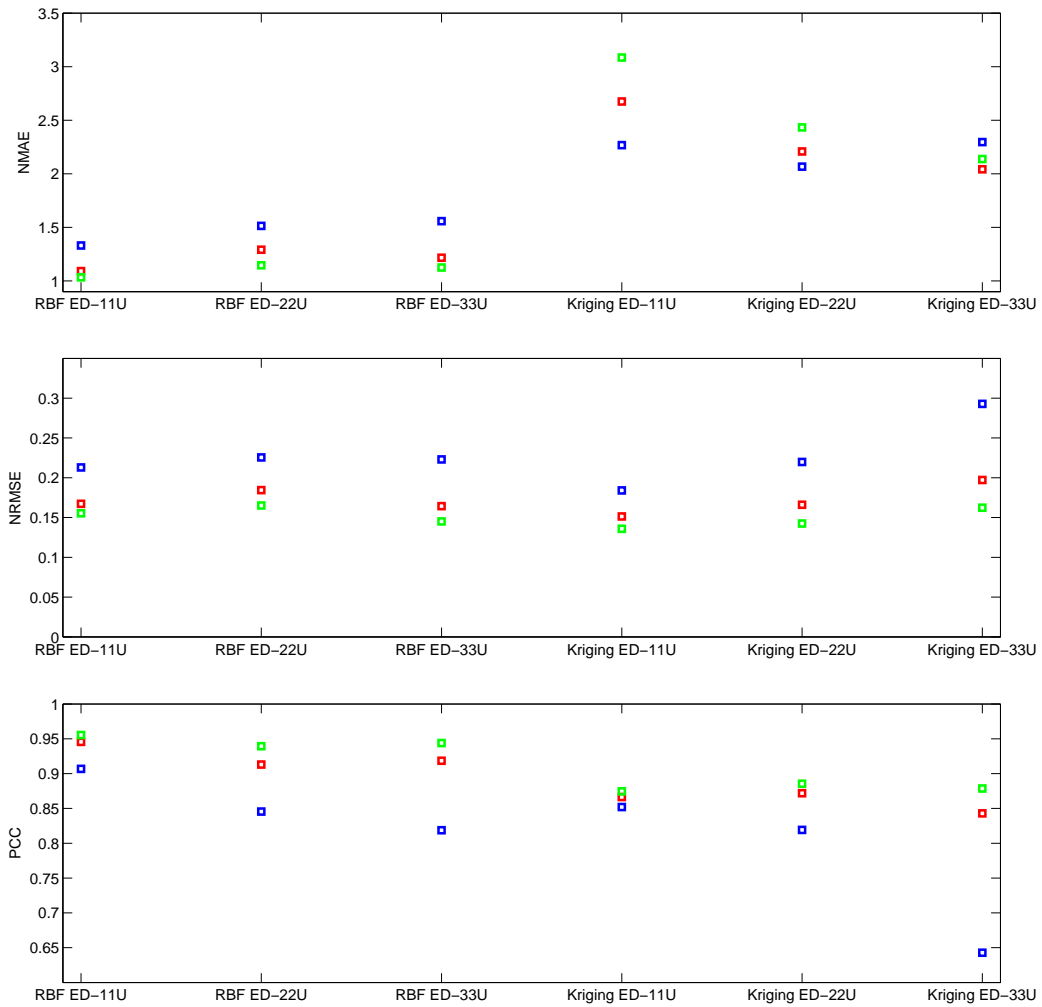


Figure 7: PCC, NRMSE and NMAE values for Scenario 2 after 20 (blue), 30 (red) and 40 (green) infill samples

4.4 Analysis

In terms of mean social surplus improvement (Figure 4), the top five combinations from Table 1, after 20 iterations, are number 17, 25, 18, 10 and 26. They all provide mean objective function values in the range of 94.3% to 97.7% of f^* . Based on the 95% confidence intervals we can with 95% confidence reject the hypothesis that all of the combinations are as good as number 17. In fact, only when comparing number 17 with 25, 18 and 10 we cannot reject this. These four combinations are all Kriging models with EI sampling and ED-7, with or without the additional five uniform toll levels, and they are also among the best performing ones after 40 samples.

Comparing the different combinations based on their PCC, NRMSE and NMAE values after 20 sampled sets of toll levels, it is difficult to draw any conclusions. This difficulty is illustrated in Figure 8, in which the PCC, NRMSE and NMAE values are plotted against $\Delta S(\hat{\tau})$ for each individual repetition after 20 sampled sets of toll levels in Scenario 1. The solutions with $\Delta S(\hat{\tau})$ within 2% of f^* are marked in red. Clearly there are combinations which produce good solutions with a rather low value on PCC and high values on NRMSE and NMAE. The opposite holds for solutions with lower objective function value. Thus one should use these measurements carefully when discussing which combination is the best one.

CAND sampling in general result in large confidence intervals for $\Delta S(\hat{\tau})$. Thus there is a less chance of actually returning high objective function values if the method is run only once. In contrast, Kriging with Gaussian function and EI sampling has the smallest confidence interval after 20, 30 and 40 samples (with the exception of ED-14U after 20 samples). This combination is also the best performing one when the number of samples is increased.

Comparing RBF and the Kriging models with CAND sampling, they produce similar results in term of $\Delta S(\hat{\tau})$. Thus it seems that the main contribution from using Kriging models comes from the possibility to use EI sampling. While CAND sampling includes generation of random numbers, EI sampling is purely deterministic (not considering solving problem (10) with stochastic search methods). This is reflected in the results, as the confidence intervals are smaller for all but a few cases of EI sampling.

One interesting aspect of using EI sampling is that it relies on approximation of uncertainties from the Kriging model. It is known [42] that this uncertainty is in general underestimated with few samples, and an underestimation of the uncertainty will lead to a more local optimizing behavior. In Figure 5 we can see that the NMAE values are increasing when adding infill samples with the EI sampling, and decreasing with additional infill samples when using CAND sampling. This is an effect from the local optimizing behavior, generating infill samples in a small region for which the response surface will have a good fit but which will give a worse fit in areas with less samples (increased NMAE value). This behavior is not entirely bad, and will at some point appear when using EI sampling, as we can always expect a local behavior eventually with this sampling technique. If zooming in on a small area early, this might, however, lead to local optimal solutions.

Given the differences between our experiments and the DynusT experiments in [18] it is still interesting to compare the results presented in Table 2. In terms of NMAE the results are similar, but considering the mean error across the domain, NRMSE, it seems that it is more difficult to approximate our problem (with static assignment and elastic demand) compared with the problem in [18] (using DynusT traffic assignment and fixed demand). This limited comparison is not enough to make any definitive conclusions, but it is still an interesting observation.

It is also interesting to see that with 67 initial samples and 30 infill samples, the

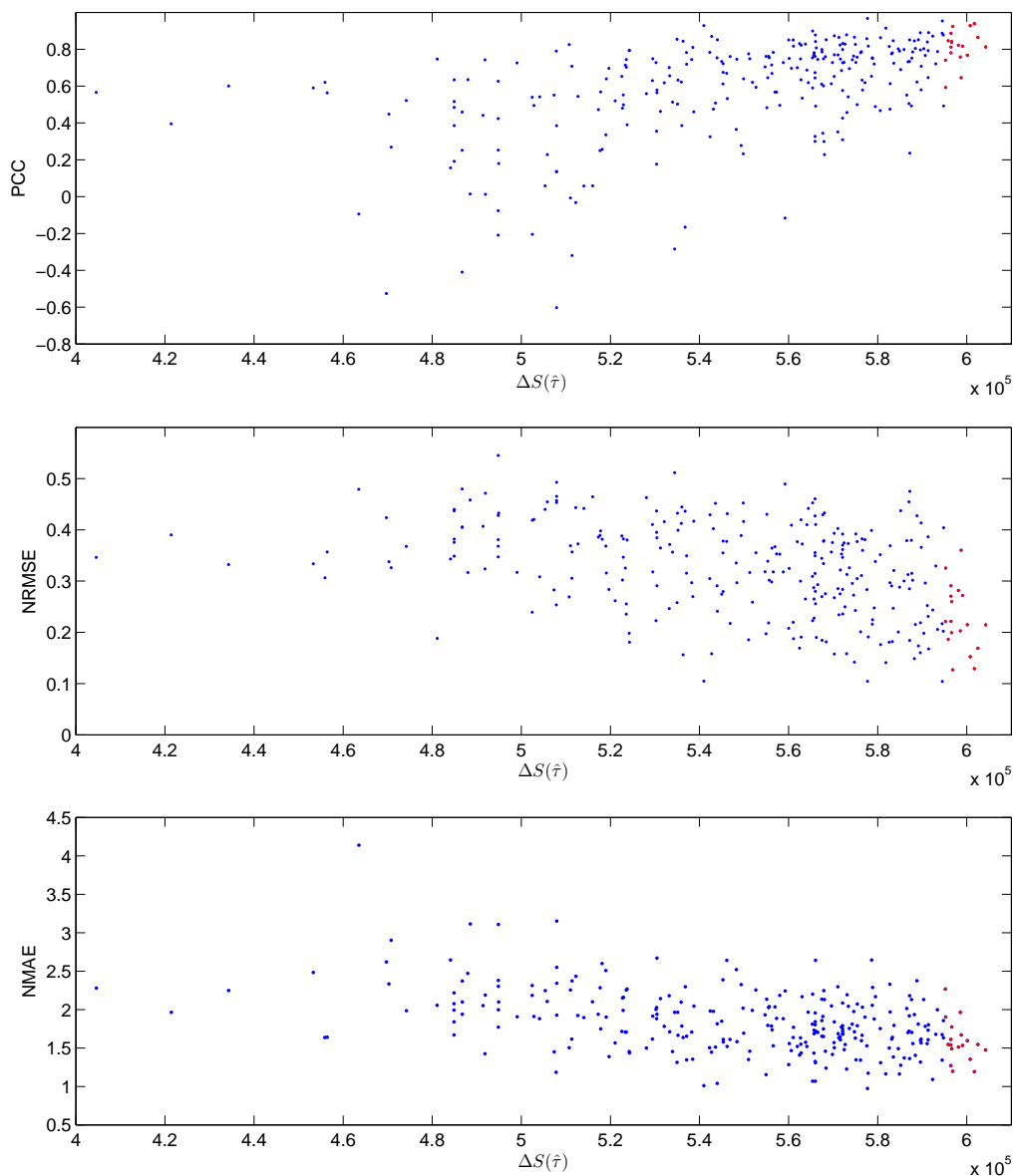


Figure 8: PCC, NRMSE and NMAE plotted against $\Delta S(\hat{\tau})$ after 20 sampled sets of toll levels.

Kriging models with $p = 1$ and $p = 2$ achieve an objective function value within 1% of f^* . It is also clear that after the initially 67 sampled points, the RBF surrogate model is the best performing one, but the CAND infill sampling strategy is inferior in comparison with EI sampling, in terms of improving the costly objective function value.

In Scenario 2, the problem of few samples becomes apparent. With few samples at hand, the actual variability of the costly objective function value is not captured, which leads to over-fitted response surfaces, and especially to an underestimation of the uncertainty in the Kriging model. This result in a more local optimizing behavior of the EI sampling when the initial set of samples is given by ED-11U compared to ED-33U. The local optimizing behavior can be seen from the rapid increase in the NMAE value and decrease in the NRMSE value as the infill samples are added. Still, with the more local behavior, good solutions are obtained more rapidly in comparison to the two alternative sets of initial samples, for which the EI-sampling will spend more time generating samples in less explored areas. The CAND sampling used with the RBF has a built-in mechanism of always sampling in less explored areas at certain intervals, and the effect of increasing NMAE with additional infill samples is not appearing for the RBF case. Over all, the RBF response surface results in much smaller NMAE values and higher PCC values in comparison with the Kriging model. This difference, however, does not seem to affect the performance of the response surfaces in terms of mean social surplus improvements. Comparing the NMAE and PCC values between Scenario 1 and 2, it is clear that the RBF response surface produce lower NMAE values and higher PCC values for Scenario 2, without performing better than the Kriging response surface. This suggests that the RBF response surface is more sensitive to over-fitting, in comparison to the Kriging response surface.

5 Application to a dynamic traffic model for the Stockholm region

5.1 The Regent/VisumDUE model

The dynamic model of Stockholm, used within this project, has a demand module (Regent) and supply module (VisumDUE), and includes 1 240 demand zones and 17 000 links. An overview of the modeling system is given in [28].

The Regent demand module includes trip generation, mode choice (car driver, car passenger, public transport, bicycle and walk) and destination choice. Regent exists in different version, and the version used within this work only includes commuting trips. VisumDUE is the supply module and includes dynamic route choice and network loading for fixed demand. The network loading is based on a macroscopic traffic flow model, capable of handling large networks. VisumDUE can take turning capacities in intersections and blocking back from queues at congested links into account. Thus it is capable of describing both temporal and spatial evolution of congestion in the network, in contrast to the STA model applied in the previous section. On the other hand,

deterministic route choice, and macroscopic attributes and variables are similar to what is used in the STA model. The VisumDUE model includes fixed demand matrices for other trips and commercial traffic, which together with commuting trips from Regent results in three different user classes. VOT for each user class is given in Table 3.

Table 3: Value of time for each demand segment.

Demand segment	Value of time (SEK/h)
Commuting trips	87
Other private trips	59
Commercial car trips	100

For each evaluated congestion pricing scenario, five iterations have been carried out between Regent and VisumDUE, the demand for commuting trips has been updated with the method of successive averages (MSA) between each iteration. Each traffic assignment carried out with VisumDUE takes between 1.5 and 3 hours computing time, depending on the level of congestion in the road network, which results in a total computational time for evaluating one scenario with Regent/VisumDUE of 8 to 15 hours (the demand calculation in Regent is done within a few minutes of computational time). Typical computational time of 10 hours has been observed.

The evaluation of each scenario is done for the morning, 6:30-9:00. Within this timer period, a time discretization of 15 minute periods is applied in the peak-hour (7:30-8:30) and half-hour periods outside this period. The currently implemented congestion charging scheme in Stockholm is applied with a time profile of 50%, 75% and 100% of the maximum fee (See Table 4 for the profile of the morning period). We have chosen to keep this profile as it is chosen by the assumption that a sharp increase/decrease of the fee might result in undesirable effects. Alternatively, it is possible to let the time profile of the pricing scheme be variable, as is done in [18]. The computation of the social surplus is however only based on the peak-hour, due to limitations in the VisumDUE software.

Table 4: Morning period time profile for the congestion pricing scheme in Stockholm (2007-2015).

Time period	Fee as % of maximum fee
06:30-07:00	50%
07:00-07:30	75%
07:30-08:30	100%
08:30-09:00	75%

Figure 9 shows the central part of the Stockholm network in VisumDUE, with charged links marked in purple. The Lidingö exemption [4] has not been implemented in the model. Charged links belong to one of the six cordons included in Scenario 1, previously presented in Section 4.2 and illustrated in Figure 10.

The computation of the change in social surplus follows (1), where the change in revenues, ΔR , is the sum of the collected tolls during the morning peak-hour. For the commuting trips, the change in consumer surplus is given by the logsum from Regent and for other trips and commercial traffic the change in consumer surplus is given by the change in total user costs during the morning peak-hour for each user class, with travel time converted to SEK by the VOT.



Figure 9: VisumDUE.

5.2 Experimental setup

The best performing experimental design with the STA model was ED-7. We will, however, not use this experimental design as we believe there are differences between the static and dynamic models which should be taken into account in this choice. The ED-7 will produce toll sets to sample on the boundary of the feasible set. For the STA model, high objective function values are obtained for solutions on the boundary. For the Regent/VisumDUE model sampled sets of toll levels on the boundary instead return low objective function values (all sampled toll levels with ED-7 returned negative objective function values). Therefore ED-7U will instead be used, augmented with two additional sets of toll levels which were already available. One is the current pricing scheme in Stockholm, which is to set the peak toll to 0 for all cordons but the ones corresponding

to the current pricing scheme, which in turn are set to 20 SEK. The other one consists of a pricing scheme suggested by one of the experts involved in designing the Stockholm congestion pricing scheme.

Table 5 presents the 14 sets of initially sampled toll levels. Set 1-7 are sampled using ED-7, set 8 and 9 are sampled from the two corner points given by minimum and maximum toll levels, set 10-12 are the three additional interior points with uniform toll levels and set 13 and 14 are the two additional points.

Based on the experiments with the STA model, a Kriging model with Gaussian correlation function has been chosen as surrogate model and EI sampling as infill strategy. Good performances with the STA model were also achieved with both the exponential and generalized exponential correlation functions. Due to the computational burden of running Regent/VisumDUE we can, however, only afford evaluating one combination of response surface, experimental design and infill strategy. Our choice of the Gaussian correlation function is mainly based on the high objective function values returned, together with the smallest confidence intervals (considering only ED-7/ED-7U) after 20 samples. For comparison, Kriging models with exponential and generalized exponential functions, as well as the RBF interpolation function, have been constructed and evaluated as response surfaces fitted to the finally sampled sets of toll levels. This will, however, not give a fair comparison since the infill samples are generated based on the Kriging model with Gaussian correlation function.

Experiments with the STA model have been evaluated after 20, 30 and 40 iterations. With ED-7U this corresponds to eight additionally sampled sets of toll levels. From these experiments we can also see that infill points with EI sampling is important. For a fair comparison, we will therefore evaluate the experiment with Regent/VisumDUE after eight infill points as well, resulting in a total of 22 samples.

5.3 Results

The eight generated toll sets (numbered 15 to 22) are displayed in Table 6 together with their corresponding costly objective function value. Optimal toll levels obtained with the four different response surfaces, fitted to the final 22 toll set samples are also shown in Table 6. For the final sets of sampled toll levels, Table 7 presents PCC, NRMSE and NMAE values with the four different response surfaces.

Table 5: Toll levels (in SEK) from the 14 initially sampled sets of toll levels.

Toll set	Bypass highway		Inner city bridges		Toll ring		$\Delta S(\hat{\tau})$
	Southbound	Northbound	Southbound	Northbound	In	Out	
1	10	30	0	20	50	0	-536 424
2	60	10	10	30	40	40	-7 402
3	0	0	20	0	0	30	-16 648
4	0	0	30	60	30	20	-416 996
5	20	50	40	40	60	50	-438 640
6	30	20	60	10	20	60	-271 827
7	50	40	50	50	10	10	-174 695
8	0	0	0	0	0	0	0
9	60	60	60	60	60	60	-817 633
10	30	30	30	30	30	30	45 047
11	15	15	15	15	15	15	349 560
12	45	45	45	45	45	45	-335 677
13	20	50	10	20	30	10	-54 157
14	0	0	0	0	20	20	232 562

5.4 Analysis

While it is difficult to evaluate the performance of the surrogate-based optimization framework without knowing the global optimal solution, it is still clear that the framework is successful in generating infill samples and fitting a response surface which can be used to generate toll levels which produce higher social surplus, in comparison to the initial samples. Comparing the best set of toll levels among the initial samples with the best found solution an improvement of 16.2% is obtained.

Comparing the static and dynamic models, we can see the PCC and NMAE values are in the same range for both type of models. The NRMSE value is, however, more than doubled when using the dynamic model. This suggests that the DTA used within this work, results in a costly function which is much more difficult to approximate with the evaluated response surfaces, with a similar number of total samples.

Finally comparing the four different interpolation functions applied to the final set of sampled toll levels (both initial and infill samples), they all show PCC, NRMSE and NMAE values in the same region. The Kriging model with exponential correlation function ($p = 1$), has lowest PCC value and highest NRMSE value. This is also the interpolation function which produce the lowest estimation (s^*) of the optimal costly objective function value, as well as the lowest actual best obtained costly objective function value ($\Delta S(\hat{\tau})$). The best performing response surface is the Kriging model with Gaussian correlation function ($p = 2$). It has a PCC value in the upper region of obtained PCC values, and NRMSE value in the region of obtained NRMSE values, and the lowest obtained NMAE value of all evaluated response surfaces.

Table 6: Infill samples and optimal solutions with their corresponding costly objective function value.

Toll set	Bypass highway		Inner city bridges		Toll ring		$\Delta S(\hat{\tau})$
	Southbound	Northbound	Southbound	Northbound	In	Out	
	Infill samples using EI sampling						
15	28.74	8.12	0.00	11.22	19.23	15.19	234 804
16	6.62	19.37	0.00	8.95	15.49	25.93	123 848
17	12.43	4.82	18.25	11.83	17.92	8.00	262 760
18	37.08	24.94	23.05	11.06	12.74	12.25	89 116
19	21.13	3.25	12.17	13.69	20.26	28.29	234 617
20	19.68	13.13	10.34	32.66	10.86	9.78	181 917
21	2.20	9.93	10.83	7.54	19.52	16.77	346 589
22	14.63	11.38	16.11	4.81	18.94	19.03	262 519
	Optimal solutions						
$p = 1$	15.00	9.93	10.83	11.83	17.92	15.00	353 927
$p = 2$	6.07	10.15	10.23	14.93	16.72	14.16	406 300
$1 \leq p \leq 2$	10.40	8.63	8.78	13.68	16.53	15.92	357 588
RBF	5.45	13.14	14.97	14.61	16.62	15.21	396 307

Table 7: PCC, NRMSE, NMAE and s^* values for the four fitted response surfaces.

Response surface	PCC	NRMSE	NMAE	s^*	Evaluated
$p = 1$	0.72	0.71	1.94	355 948	353 927
$p = 2$	0.77	0.65	1.76	411 230	406 300
$1 \leq p \leq 2$	0.78	0.64	2.12	400 786	357 588
RBF	0.78	0.68	2.13	369 608	396 307

6 Discussion

For both the static and dynamic model it is clear that the surrogate-based optimization framework, with evaluated infill strategies, is able to find solutions with higher costly objective function value, in comparison to what were known after the initial sampling. There is, however, a question whether the thorough evaluation conducted in Section 4 with the STA model can contribute when choosing response surface, experimental design and infill strategy.

There are two alternative approaches, which have not been used within this work:

- i) To use the same network layout, the same definition of user classes and OD pairs, and the same demand model, for both static and dynamic assignment. This would facilitate the isolation of the effects of introducing dynamic assignment.
- ii) To use a less computational demanding dynamic model (for instance by considering smaller network). This would allow comparisons with global optimal solutions also for the case of a DTA model.

For (i) the main difficulty to overcome is that even with a STA model, the computational time using the more detailed network from our Regent/VisumDUE model, would be too long in order to carry out the evaluation in Section 4. The main difficulty for (ii) is whether a smaller network contains the complexity of a network of real size, and thus whether any conclusions drawn from experiments on this smaller network would be valid for a case with a larger network. Our approach is similar to the one used in [18], but with the difference that the evaluations with the STA model is done on a network representing an actual traffic system, rather than a toy network. Also, the structure of the demand models used within both the DTA and STA models are similar (logit choice models).

In terms of applicability of the surrogate-based optimization framework, it is only relying on the computation of the costly objective function value. Thus it can easily be applied to any traffic modeling system capable of returning this information for a given set of toll levels. For the small number of samples considered in this paper, 20–40 samples, it might not even be worth the trouble to integrate the computation of the costly function with the surrogate-based optimization framework, especially if one will only evaluate one cordon design.

7 Conclusions and further research

This paper has focused on the performance of a surrogate-based optimization framework for optimizing toll levels in cordon pricing schemes, when the number of possible evaluations of the costly objective function is limited to a small number. We have for the STA model shown that for a case with six different cordon tolls to optimize, close to global optimal solutions can be computed with as few as 20 evaluations of the costly

objective function. The approach is, however, sensitive to the problem dimension, and might show a more local optimizing behavior when the problem dimension is increased, if the number of initial samples is not increased. For the DTA model the surrogate-based optimization framework has been successfully applied. The experiments, however, suggest that the approximation of the costly objective function using any of the evaluated response surfaces, becomes more difficult for the DTA in comparison with the STA.

While the Regent/VisumDUE DTA model is based on macroscopic deterministic assignment, further research is needed in order to evaluate the approach for mesoscopic and microscopic models based on simulation. Using simulation-based models will introduce stochasticity to the costly objective function, why other response surfaces, better equipped for handling noisy functions, might be more suitable. Also, allowing variations of toll levels over time will affect the performance of the optimization approach, and the inclusion of time profile as a variable is an important area for further evaluation.

The introduction of more dynamic traffic assignment models for larger urban areas makes decision support tools for infrastructure investment and pricing policies, based on DTA models, an important area for future research. How to combine information from the less computational demanding STA with information from the more computational demanding DTA model then becomes an interesting question for further research.

References

- [1] A. C. Pigou, *Welth and Welfare*. London: MacMillan, 1920.
- [2] W. S. Vickrey, “Congestion theory and transport investment,” *The American Economic Review*, vol. 59, no. 2, pp. 251–260, 1969.
- [3] M. Fosgerau, “Congestion in the bathtub,” *The Munich Personal RePEc Archive (MPRA)*, 2015.
- [4] J. Eliasson, “A cost-benefit analysis of the Stockholm congestion charging system,” *Transportation Research Part A*, vol. 43, no. 4, pp. 468 – 480, 2009.
- [5] J. Eliasson, L. Hultkrantz, L. Nerhagen, and L. S. Rosqvist, “The Stockholm congestion - charging trial 2006: Overview of effects,” *Transportation Research Part A*, vol. 43, no. 3, pp. 240 – 250, 2009.
- [6] J. Leape, “The London Congestion Charge,” *Journal of Economic Perspectives*, vol. 20, no. 4, pp. 157–176, 2006.
- [7] L. Engelson and D. van Amelsfort, “The role of volume-delay functions in forecast and evaluation of congestion charging schemes, application to Stockholm,” in *Proceedings of the Kuhmo Nectar Conference*, (Stockholm), 2011.
- [8] J. Eliasson, M. Börjesson, D. van Amelsfort, K. Brundell-Freij, and L. Engelson, “Accuracy of congestion pricing forecasts,” *Transportation Research Part A: Policy and Practice*, vol. 52, pp. 34 – 46, 2013.
- [9] M. Börjesson and I. Kristoffersson, “Assessing the welfare effects of congestion charges in a real world setting,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 70, pp. 339 – 355, 2014.
- [10] M. Beckmann, C. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*. New Haven: Yale University Press, 1956.
- [11] A. May, R. Liu, S. Shepherd, and A. Sumalee, “The impact of cordon design on the performance of road pricing schemes,” *Transport Policy*, vol. 9, no. 3, pp. 209 – 220, 2002.
- [12] A. Sumalee, “Optimal implementation path road pricing schemes time dependent model,” *Journal of the Eastern Asia Society for Transportation Studies*, vol. 6, pp. 624–639, 2005.
- [13] T. Akiyama and M. Okushima, “Implementation of cordon pricing on urban network with practical approach,” *Journal of Advanced Transportation*, vol. 40, no. 2, pp. 221–248, 2006.

- [14] G. Schuitema, L. Steg, and S. Forward, “Explaining differences in acceptability before and acceptance after the implementation of a congestion charge in Stockholm,” *Transportation Research Part A: Policy and Practice*, vol. 44, no. 2, pp. 99 – 109, 2010.
- [15] M. Börjesson, J. Eliasson, M. B. Hugosson, and K. Brundell-Freij, “The Stockholm congestion charges -5 years on. effects, acceptability and lessons learnt,” *Transport Policy*, vol. 20, no. 0, pp. 1 – 12, 2012.
- [16] J. Ekström, *Optimization Approaches for Design of Congestion Pricing Schemes*. PhD thesis, Linköping University, 2012.
- [17] D. Joksimovic, M. C. J. Bliemer, and P. H. L. Bovy, “Article title - optimal toll design problem in dynamic traffic networks with joint route and departure time choice,” *Transportation Research Record*, vol. 1923, pp. 61–72, 2005.
- [18] X. M. Chen, L. Zhang, X. He, C. Xiong, and Z. Li, “Surrogate-based optimization of expensive-to-evaluate objective for optimal highway toll charges in transportation network,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 29, no. 5, pp. 359–381, 2014.
- [19] I. Kristoffersson and L. Engelson, “Alternative road pricing schemes and their equity effects: Results of simulations for Stockholm,” in *Proceeding of the 90th Annual Meeting of the Transportation Research Board*, (Washington DC), 2011.
- [20] J. Ekström, L. Engelson, and C. Rydergren, “Optimal toll locations and levels in congestion pricing schemes: A case study of Stockholm,” *Transportation Planning and Technology*, vol. 37, pp. 333–353, 2014.
- [21] I. Kristoffersson, “Impacts of time-varying cordon pricing: Validation and application of mesoscopic model for Stockholm,” *Transport Policy*, pp. 1–10, 2011. 10.1016/j.tranpol.2011.06.006.
- [22] B.-W. Wie and R. L. Tobin, “Dynamic congestion pricing models for general traffic networks,” *Transportation Research Part B: Methodological*, vol. 32, no. 5, pp. 313 – 327, 1998.
- [23] R. Zhong, A. Sumalee, and T. Maruyama, “Dynamic marginal cost, access control, and pollution charge: a comparison of bottleneck and whole link models,” *Journal of Advanced Transportation*, vol. 46, no. 3, pp. 191–221, 2012.
- [24] A. Karoonsoontawong, S. Ukkusuri, and a. K. M. K. S. Travis Waller, “A simulation-based approximation algorithm for dynamic marginal cost pricing,” *Journal of the Transportation Research Forum*, vol. 47, pp. 81–99, 2008.

- [25] A. de Palma, M. Kilani, and R. Lindsey, “Congestion pricing on a road network: A study using the dynamic equilibrium simulator metropolis,” *Transportation Research Part A*, vol. 39, no. 7-9, pp. 588–611, 2005.
- [26] J. Chow and A. Regan, “A surrogate-based multiobjective metaheuristic and network degradation simulation model for robust toll pricing,” *Optimization and Engineering*, vol. 15, no. 1, pp. 137–165, 2014.
- [27] M. Börjesson and I. Kristoffersson, “The Gothenburg congestion charge. Effects, design and politics,” *Transportation Research Part A: Policy and Practice*, vol. 75, pp. 134 – 146, 2015.
- [28] A. Almroth, S. Berglund, O. Canella, L. Engelson, G. Flötteröd, D. Jonsson, I. Kristoffersson, and J. West, “Further development of SAMPERS and modeling of urban congestion,” Working papers in Transport Economics 2014:10, CTS - Centre for Transport Studies Stockholm (KTH and VTI), May 2014.
- [29] A. de Palma and R. Lindsey, “Modelling and evaluation of road pricing in Paris,” *Transport Policy*, vol. 13, no. 2, pp. 115–126, 2006.
- [30] H. C. W. L. Williams, “On the formation of travel demand models and economic evaluation measures of user benefit,” *Environment and Planning A*, vol. 9, pp. 285–344, 1977.
- [31] N.-H. Quttineh, *Models and Methods for Costly Global Optimization and Military Decision Support Systems*. PhD thesis, Linköping University, 2012.
- [32] H.-M. Gutmann, “A radial basis function method for global optimization,” *Journal of Global Optimization*, vol. 19, no. 3, pp. 201–227, 2001.
- [33] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, “Design and analysis of computer experiments,” *Statist. Sci.*, vol. 4, pp. 409–423, 11 1989.
- [34] D. R. Jones, M. Schonlau, and W. J. Welch, “Efficient global optimization of expensive black-box functions,” *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.
- [35] L. Zuluaga and T. Terlaky, *Modeling and Optimization: Theory and Applications: Selected Contributions from the MOPTA 2012 Conference*. Springer Proceedings in Mathematics & Statistics, Springer, 2013.
- [36] J. Müller, “User guide for modularized surrogate model toolbox,” tech. rep., Tampere University of Technology, 2012.
- [37] M. J. Powell, “The theory of radial basis function approximation in 1990,” in *Advances in Numerical Analysis, Volume 2, Wavelets, Subdivision Algorithms and Radial Basis Functions* (W. Light, ed.), pp. 105–210, Oxford University Press, 1992.

- [38] M. Locatelli and F. Schoen, *Global Optimization: Theory, Algorithms, and Applications*. MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 2013.
- [39] R. G. Regis and C. A. Shoemaker, “A stochastic radial basis function method for the global optimization of expensive functions,” *INFORMS Journal on Computing*, vol. 19, no. 4, pp. 497–509, 2007.
- [40] D. Jones, “A taxonomy of global optimization methods based on response surfaces,” *Journal of Global Optimization*, vol. 21, pp. 345–383, 2002.
- [41] N.-H. Quttineh and M. Edvall, “An adaptive radial basis algorithm (ARBF) for expensive black-box mixed-integer constrained global optimization,” *Optimization and Engineering*, vol. 9, pp. 311–339, 2008.
- [42] D. R. Jones, “Improved quantification of prediction error for kriging response surfaces.” Presented at IMA Hot Topics Workshop Uncertainty Quantification in Industrial and Energy Applications: Experiences and Challenges, June 2 2011.
- [43] A. Forrester, A. Sobester, and A. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*. Wiley, 2008.
- [44] N.-H. Quttineh and K. Holmström, “The influence of experimental designs on the performance of surrogate model based costly global optimization solvers,” *Studies in Informatics and Control*, vol. 18, no. 1, pp. 87–95, 2009.
- [45] T. J. Mitchell and M. D. Morris, “Bayesian design and analysis of computer experiments: two examples,” *Statistica Sinica*, vol. 2, pp. 359–379, 1992.
- [46] G. G. Wang and S. Shan, “Review of metamodeling techniques in support of engineering design optimization,” *Journal of Mechanical Design*, vol. 129, no. 4, pp. 370–380, 2007.
- [47] J. Müller, *Surrogate Model A Algorithm for CoComputational Expensive Black-Box Global Optimization Problems*. PhD thesis, Tampere University of Technology, Tampere, 2012.
- [48] J. Ekström, “Finding second-best toll locations and levels by relaxing the set of first-best feasible toll vectors,” *European Journal of Transport and Infrastructure Research*, vol. 14, no. 1, pp. 7–29, 2014.
- [49] Trafikverket, “Förändrade trängselskatter i Stockholm -Underlag för 2013 års Stockholmsförhandling,” Rapport 2013:110, Trafikverket i samarbete med CTS och KTH, 2013.

A Cordon layouts

Figure 10 and 11 illustrate the cordon layouts in Scenario 1 and 2, with red marks indicating a tolled link (direction of flow given by the arrow) and toll level variable indicated by τ_c .

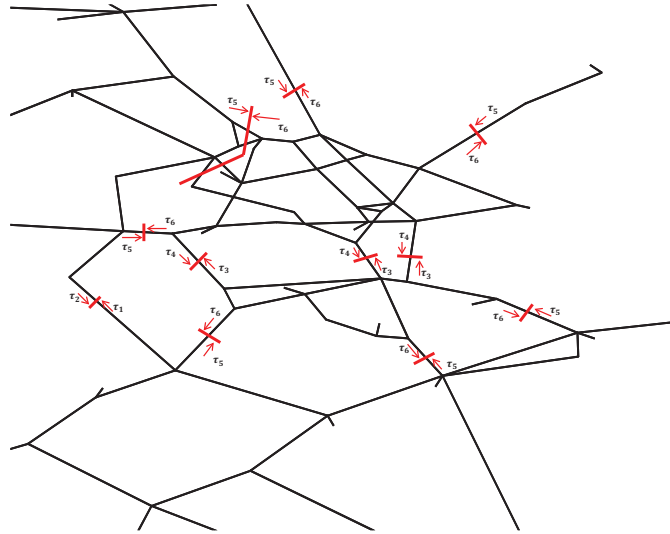


Figure 10: Cordon layout for Scenario 1. This scenario includes 6 toll levels.

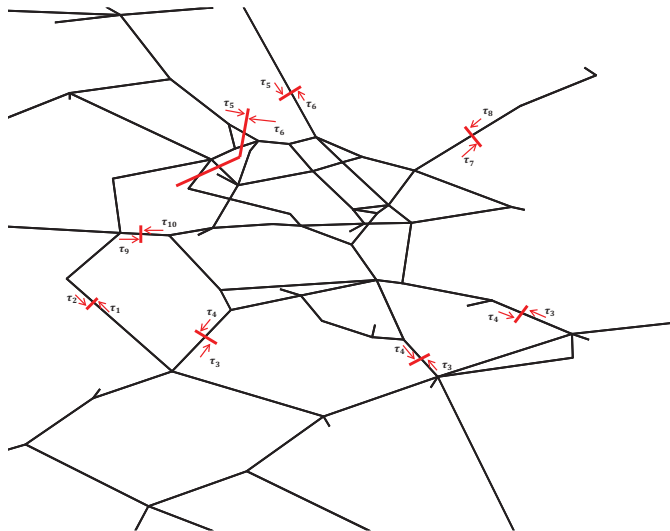


Figure 11: Cordon layout for Scenario 2. This scenario includes 10 toll levels.