

1N-34
185532
56 P

NASA Contractor Report 191510
ICASE Report No. 93-50

ICASE



SURROGATES FOR NUMERICAL SIMULATIONS; OPTIMIZATION OF EDDY-PROMOTER HEAT EXCHANGERS

Serhat Yesilyurt
Anthony T. Patera

N94-13074

Unclas

G3/34 0185532

NASA Contract No. NAS1-19480
August 1993

Institute for Computer Applications in Science and Engineering
NASA Langley Research Center
Hampton, Virginia 23681-0001

Operated by the Universities Space Research Association



National Aeronautics and
Space Administration
Langley Research Center
Hampton, Virginia 23681-0001

(NASA-CR-191510) SURROGATES FOR
NUMERICAL SIMULATIONS; OPTIMIZATION
OF EDDY-PROMOTER HEAT EXCHANGERS
Final Report (ICASE) 56 P



ICASE Fluid Mechanics

Due to increasing research being conducted at ICASE in the field of fluid mechanics, future ICASE reports in this area of research will be printed with a green cover. Applied and numerical mathematics reports will have the familiar blue cover, while computer science reports will have yellow covers. In all other aspects the reports will remain the same; in particular, they will continue to be submitted to the appropriate journals or conferences for formal publication.



SURROGATES FOR NUMERICAL SIMULATIONS; OPTIMIZATION OF EDDY-PROMOTER HEAT EXCHANGERS

Serhat Yesilyurt

Department of Nuclear Engineering
Massachusetts Institute of Technology
Cambridge, MA

and

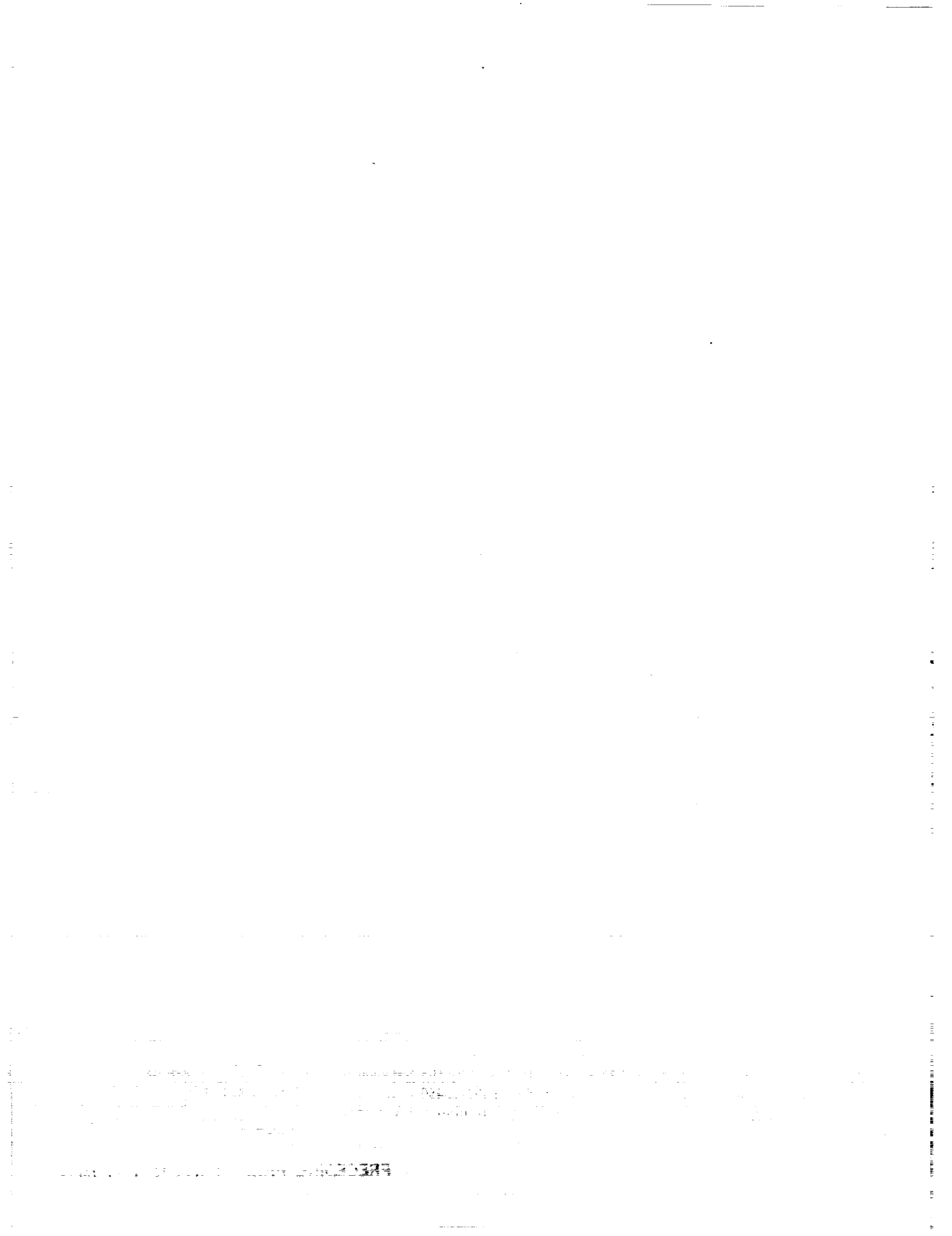
Anthony Patera¹

Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA

ABSTRACT

Although the advent of fast and inexpensive parallel computers has rendered numerous previously intractable calculations feasible, many numerical simulations remain too resource-intensive to be directly inserted in engineering optimization efforts. An attractive alternative to direct insertion considers models for computational systems: the expensive simulation is evoked only to construct and validate a simplified input-output model; this simplified input-output model then serves as a simulation *surrogate* in subsequent engineering optimization studies. We present here a simple "Bayesian-validated" statistical framework for the construction, validation, and purposive application of static computer simulation surrogates. As an example, we consider dissipation-transport optimization of laminar-flow eddy-promoter heat exchangers: parallel spectral element Navier-Stokes calculations serve to construct and validate surrogates for the flowrate and Nusselt number; these surrogates then represent the originating Navier-Stokes equations in the ensuing design process.

¹This work was supported by DARPA Grant N00014-91-J-1889; ONR Grant N00014-90-J-4124; ONR Grant N00014-89-J-1610; and by the National Aeronautics and Space Administration under NASA Contract No. NAS1-19480 while the author was in Engineering (ICASE), NASA Langley Research Center, Hampton, VA 23681.



1 Introduction

Large-scale numerical calculation, such as fluid flow simulation, is an increasingly significant component of engineering and scientific *analysis*. However, despite recent advances in both algorithms and architectures, many relevant individual calculations still require many hours of expensive supercomputer time. As a result, direct insertion of these resource-intensive simulations as “subroutine” calls in particular design and optimization studies is typically not viable. First, the number of objective-function evaluations required to find an optimal, or even reasonable, solution to an optimization problem will not be known *a priori*. Direct insertion of expensive simulations may exhaust allocated resources before interesting — or even feasible — solutions are obtained. Second, effective engineering design and optimization processes are evolutionary, with goals and constraints continually modified to reflect newly available information or specifications. Direct insertion of large-scale calculations strongly *inhibits* adaptability: with each revision of objectives, previous computations must be discarded, and a new sequence of expensive simulations must be initiated. Third, the value of expensive numerical simulations can be greatly enhanced by proper incorporation of prior information derived from collateral analytical, experimental, or heuristic investigations. Direct insertion of simulation results renders model fusion and validation difficult. Fourth, most design and optimization exercises are multidisciplinary in nature [1], involving numerous relatively distinct fields of physical inquiry (e.g., fluid mechanics, solid mechanics, physical chemistry). Direct insertion of diverse simulations affords little opportunity to accommodate — or exploit — differing degrees of complexity and sensitivity. In summary, if large-scale computation is to graduate from analysis to *synthesis*, new paradigms are required.

One attractive solution to the simulation-integration impasse considers *models for computational systems*: the expensive, large-scale simulation, denoted \mathcal{M}^0 , is evoked only to construct and validate a simplified computational model, denoted \mathcal{M} ; this simplified model, \mathcal{M} , then serves as an inexpensive *surrogate* for \mathcal{M}^0 in subsequent engineering applications. The simplified model \mathcal{M} can be evaluated effectively *ad infinitum*, can support a large class of objective functions, can readily accommodate extra-simulation information, and can be easily incorporated into multidisciplinary design studies. The application of models for computational systems does, however,

raise new questions, in particular as regards purposiveness: to what extent is the design procedure misdirected, or proposed designs mispredicted, by the introduction of approximate simulation surrogates?

In this paper we develop — and apply — a complete surrogate framework for optimization based on the simple validation concepts presented in [2]. More broadly, the work is founded upon several related streams of inquiry. From system identification (control) theory [3–6] we borrow the notion of algorithmic logical empiricism, in which available data is systematically incorporated into the model construction and validation processes; from the design of experiments [7] we appreciate the need for sampling heuristics and response surfaces; from statistical prediction rules and artificial neural networks [8–11] we adopt the concept of “construct and validate” — or “train and test” — data partitions; from the theory of machine learning [12,13] we appropriate the “probably approximately correct” framework; from Monte Carlo methods [14] and the classical equivalence of measure and probability [15] we derive our sampling procedures; from nonparametric statistical theory [16] we deduce our statistical error estimates; from scattered-data methodology [17] we derive our model-construction procedures; and from statistical quality-control theory [18,19] we adapt relevant *a posteriori* reliability concepts. Lastly, our work, in philosophy, is most closely aligned to earlier seminal efforts in statistical simulation surrogates [20–23], in which, first, the need for surrogates is motivated, second, the special role of statistical statements is recognized, and third, the idiosyncrasies of (largely deterministic) *computer* experiments are identified; other “non-surrogate” statistically motivated approaches to the incorporation of expensive simulations into optimization studies [24] are also relevant to our study.

The paper is organized as follows. In Section 2, we describe the optimization framework in which surrogates will ultimately be applied, discuss the *general* class of subproblems for which our surrogate methods are most appropriate, and introduce the *particular* laminar-flow eddy-promoter heat exchanger optimization study that will serve as our detailed illustration. Finally, we reiterate the motivation for the surrogate approach, formally define the surrogate problem, and describe the broad methodological guidelines that effective surrogate procedures should honor. In Section 3, we present our modelling methodology, treating both validation and construction-validation; the algorithms and error estimates are described, and results for the eddy-promoter heat exchanger are presented. In Section

4, we consider the incorporation of surrogate techniques into the full optimization framework, with particular focus on purposiveness and *a posteriori* analysis; the surrogate-based optimization approach is illustrated for the eddy-promoter heat exchanger problem. In Section 5, we consider several extensions to the surrogate methodology: classification maps; databoards; and multiple-output estimates. Lastly, in Section 6, we briefly state our conclusions. (For clarity and self-containedness, we include here some material already discussed in [2], in particular as regards the validation procedure; however, the optimization framework is new, as is the treatment of a “real” application, that is, a problem which truly requires a surrogate approach.)

2 General Problem Statement

2.1 Optimization Framework

In this paper we develop simulation surrogate techniques designed to function as part of a larger optimization study: we therefore require a general optimization framework in which to interpret our results. We emphasize that this paper is *not* concerned with more classical optimization issues such as optimality conditions and mathematical programming techniques [25]; we assume that our optimization problems are well posed, and that procedures exist to find at least local, and preferably global, optima.

We first introduce a bounded, lower semi-continuous objective function,

$$\Phi(\mathbf{p}; \underline{\lambda}) : \Omega \times \Lambda \rightarrow \mathbb{R}, \quad (1)$$

where \mathbf{p} is the optimization design M -vector, $\Omega \subset \mathbb{R}^M$ is the admissible (closed) domain for \mathbf{p} , or “design space,” $\underline{\lambda}$ is the optimization definition N -vector, and $\Lambda \subset \mathbb{R}^N$ is the admissible domain for $\underline{\lambda}$. The $\underline{\lambda}$ vector comprises coefficients which, as regards the optimization process, may be treated as parameters. Our minimization problem is thus: Find $\Phi_{\min}(\underline{\lambda}), \mathbf{p}^*(\underline{\lambda})$ such that

$$\Phi_{\min}(\underline{\lambda}) = \Phi(\mathbf{p}^*(\underline{\lambda}); \underline{\lambda}) \leq \Phi(\mathbf{p}; \underline{\lambda}) \quad \forall \mathbf{p} \in \Omega, \quad (2)$$

where $\Phi_{\min}(\underline{\lambda})$ and $\mathbf{p}^*(\underline{\lambda})$ are the minimum and minimizer, respectively. We explicitly introduce the dependence of the minimum and minimizer on $\underline{\lambda}$ to underscore that our optimization problem is, in fact, a *family* of optimization

problems parametrized by $\underline{\lambda}$. We define $\Phi_{\min}(\underline{\lambda})$ as a global minimum to emphasize that our surrogate techniques are intended to serve not only final, local optimization studies, but also initial, exploratory design efforts.

We are interested in a particular class of optimization problems in which the objective function can be written in terms of a *subproblem*,

$$\Phi(\underline{p}; \underline{\lambda}) = \phi(\underline{\mathcal{S}}(\underline{p}); \underline{p}; \underline{\lambda}) . \quad (3)$$

More explicitly, we can think of evaluating $\Phi(\underline{p}; \underline{\lambda})$ as

$$\underline{p} \in \Omega \xrightarrow{\underline{\mathcal{S}}(\underline{p}) \in L^\infty(\Omega)^K} \underline{s} \in \mathbb{R}^K$$

$$\Phi(\underline{p}; \underline{\lambda}) = \phi(\underline{s}; \underline{p}; \underline{\lambda}) ,$$

where \underline{p} is the subproblem input M -vector, Ω is the subproblem input domain, $\underline{s} \in \mathbb{R}^K$ is the subproblem output vector, $\underline{\mathcal{S}}(\underline{p})$ is the subproblem input-output function, and $L^\infty(\Omega)$ is the space of bounded measurable functions over the domain Ω . (We believe that most of our results require only that $\underline{\mathcal{S}}(\underline{p})$ be in $L^\infty(\Omega)^K$; however, all mathematics in this paper must be considered purely formal pending complete hypotheses and serious proofs.)

We shall further assume that the *deterministic* subproblem input-output function, $\underline{\mathcal{S}}(\underline{p}): \mathbb{R}^M \rightarrow \mathbb{R}^K$, is expressed as a functional, \underline{J} , applied to a field $\mathbf{U}(\mathbf{x}, t; \underline{p})$,

$$\underline{\mathcal{S}}(\underline{p}) = \underline{J}(\mathbf{U}(\cdot, \cdot; \underline{p}); \underline{p}) , \quad (4)$$

where $\mathbf{U}(\mathbf{x}, t; \underline{p})$ satisfies the initial-boundary-value *field subproblem*,

$$\mathbf{M}_{\underline{p}} \frac{\partial \mathbf{U}}{\partial t} = \mathbf{A}_{\underline{p}}(\mathbf{U}) \quad (5)$$

$$\mathbf{U}(\mathbf{x}, t = 0; \underline{p}) = \mathbf{U}^0(\mathbf{x}; \underline{p}) . \quad (6)$$

Here $\underline{J}: \mathbf{X} \times \Omega \rightarrow \mathbb{R}^K$ is the output functional; \mathbf{X} is the function space in which the field subproblem solution $\mathbf{U}(\mathbf{x}, t; \underline{p})$ resides; \mathbf{x} and t are space and time, respectively; $\mathbf{M}_{\underline{p}}$ and $\mathbf{A}_{\underline{p}}$ are deterministic spatial differential operators (and associated boundary conditions) parametrized by the input vector \underline{p} ; and $\mathbf{U}^0(\mathbf{x}; \underline{p})$ is the initial condition on $\mathbf{U}(\mathbf{x}, t; \underline{p})$.

We give here a very simple illustration from incompressible fluid dynamics intended to render the abstract subproblem framework more comprehensible.

To wit, we consider the drag coefficient for flow past a cylinder, in which we identify: p as a single input ($M = 1$), the Reynolds number; Ω , the input domain, as the Reynolds-number interval of interest; \underline{z} as a single output ($K = 1$), the drag coefficient; $\underline{\mathcal{S}}(p)$ as the drag coefficient-Reynolds number relationship; $\underline{\mathcal{J}}$ as the time-averaged streamwise component of the integral of the stress-normal product over the cylinder surface; $\mathbf{U}(\mathbf{x}, t; p)$ as the {velocity, pressure} pair; \mathbf{M}_p and \mathbf{A}_p as the incompressible Navier-Stokes system, in which the Reynolds number enters as a parameter. (Note that, for reasons described in Section 2.2, we will typically not be interested in either time or space as an input: all temporal and spatial dependence is eliminated by $\underline{\mathcal{J}}$, either by evaluating the subproblem field at a particular point, by averaging over time or space, or by considering properties of asymptotic, steady, or stationary solutions. For this simple example, we perform temporal and spatial averages of temporally stationary solutions.)

We make four final remarks. First, we have equated the input variables and the design variables (and hence the input space and the design space); more generally, the input variables may comprise only a subset of the design variables, p , but may also include certain definition variables, $\underline{\lambda}$. The former would be fortunate, reducing the size of the subproblem; the latter would be unfortunate, reducing the flexibility of a single surrogate to readily address several different optimization problems. Second, we presume that the field subproblem must be solved numerically, but that, for the purposes of this paper, all numerical errors are sufficiently small that we may *equate* the numerical and exact solutions. Third, as shall be discussed in Section 2.2, we shall be particularly interested in subproblems for which $\underline{\mathcal{S}}(p)$ is computationally expensive to evaluate (that is, the field subproblem is difficult); we shall assume, however, that, once $\underline{z} = \underline{\mathcal{S}}(p)$ is known, the objective function $\Phi(p; \underline{\lambda})$ can be inexpensively evaluated as $\phi(\underline{z}; p; \underline{\lambda})$. Fourth, for the optimization problems we consider, the computational complexity — and the greatest opportunity for improved efficiency — resides not in the search process, but in the objective function evaluation.

2.2 Class of Subproblems of Interest

The surrogate approach is particularly appropriate for optimization problems in which the subproblem satisfies the following three “complexity conditions.”

Condition C1: The $\underline{\mathcal{S}}(p)$ are expensive to evaluate in terms of computer costs, elapsed time, or human effort. This condition may appear to be transitory, given the continual and rapid decrease in computational times and costs [26]. We claim, however, that as computational capacity increases, problem size will also grow to accommodate: increased (perhaps finally adequate) resolution; higher fidelity mathematical models; increased physical complexity of new technologies. \square

Condition C2: Sharp regularity information on the $\underline{\mathcal{S}}(p)$, such as a Lipschitz condition,

$$|\underline{\mathcal{S}}(p_2) - \underline{\mathcal{S}}(p_1)| \leq c_L |p_2 - p_1|$$

is difficult to obtain and problem-specific. This implies, in effect, that very little regularity can be assumed of the input-output function $\underline{\mathcal{S}}(p)$. \square

Condition C3: Knowledge of $\underline{\mathcal{S}}(p)$ at one input value, p_1 , is of minimal computational value in evaluation of $\underline{\mathcal{S}}(p)$ at a second input value, p_2 : subproblem evaluation enjoys no computational economy of scales. This condition precludes certain {subproblem; input; diagnostic} triples, such as {·; time; ·}, and {steady Navier-Stokes; Reynolds number; Newton continuation}. In both these cases — assuming sufficient regularity — later calculations can exploit earlier results in order to reduce computational effort. (As we are not considering time as an input, our surrogates are “static;” this does not imply, of course, that the field subproblem involves only steady phenomena, nor that the outputs may not contribute to a time-dependent model.) \square

In colloquial terms, for a subproblem which satisfies these three conditions: we can not afford to generate subproblem solutions for *many* different input values (C1,C3); we can not consider the subproblem solutions at a *few* input values to be representative of the entire input space (C2); we can not simplify subproblem evaluations by exploiting special features (such as locality) of the parent optimization problem (C2,C3). We believe that, unfortunately, there are many problems which approximately satisfy these conditions.

Remark on Physical Experiments. It is of interest to ask why our methods are not as appropriate for experimental investigations as for computational systems, and, conversely, why experimental data analysis techniques

are not directly appropriate for computational systems. Considering the former, first, quite apart from noise, physical (in particular continuum mechanics) experiments often do not satisfy condition C3; for example, once an expensive flow apparatus is configured, there are great economies of scale in (in fact, opportunity costs incurred in *not*) obtaining data for a large number of flowrates, rather than just a few. This is because much experimental equipment is problem-specific, amortized over only a particular class of inquiries, and because elapsed time is not a serious consideration in many (though not all) laboratory environments. Second, despite recent advances in transducers and imaging, experimental data at a particular input value, p' , is already greatly reduced with respect to analogous numerical data; whereas raw *simulation* data resident on a databoard (see [2] and Section 5) can be subsequently processed to produce a wide range of different outputs, *experimental* data can serve only those applications requiring the few outputs selected in the originating investigation.

Turning now to why well-developed experimental data analysis techniques are not directly applicable to simulations, first, many experimental inquiries assume significant noise levels. In contrast, although computational inquiries do contain difficult-to-quantify factors (e.g., resolution, incomplete iteration) that may perhaps be gainfully interpreted as noise, these factors tend to be both relatively small and largely controllable. Many experiment-design techniques developed to reduce or understand uncontrolled factors (e.g., blocking and randomization [7]) are thus largely irrelevant in the computational arena [21]. Second, many experimental surrogate (e.g., response surface) methods are premised upon assumptions of both smoothness and locality (e.g., linear models, fractional factorial designs [7]); although these assumptions may be necessary for noisy experiments, deterministic simulations can benefit from less restrictive hypotheses. \square

2.3 Eddy-Promoter Heat Exchanger Example

2.3.1 The Optimization Problem

As our physical problem we consider two-dimensional laminar flow and convective heat transfer in the eddy-promoter heat exchanger shown in Figure 1. In overview, the eddy-promoter channel comprises a two-dimensional (infinite in \hat{x}_3) plane channel with plate separation (in \hat{x}_2) $2\hat{h}$, geometrically

interrupted by an infinite periodic array of insulating cylindrical eddy promoters of bottom wall-to-cylinder spacing \hat{a} , radius \hat{R} , and pitch (cylinder-to-cylinder \hat{x}_1 -separation) \hat{L} . Heat enters the channel through an isothermal bottom wall maintained at temperature \hat{T}_1 (representing, for example, a highly conducting plate housing electronic components) and leaves the channel through an isothermal "cold" top wall maintained at temperature \hat{T}_0 . A fluid flow driven by an imposed pressure gradient, $d\hat{p}/d\hat{x}_1\hat{e}_1$ — and excited to significant cross-stream transport by the eddy-promoters — serves to reduce the bottom wall-to-top wall thermal resistance. We are interested in determining that eddy-promoter placement and radius which minimizes pumping power (e.g., operating cost) and eddy promoter volume (e.g., materials cost) while simultaneously maintaining a temporally and spatially averaged bottom-wall heat flux (e.g., electronic component density), $\langle \hat{F} \rangle$, not much less than $\langle \hat{F} \rangle_{nom}$. In more quantitative terms, we wish to: minimize $\hat{\beta}_1 \times$ the pumping power + $\hat{\beta}_2 \times \hat{R}^2$ + $\hat{\beta}_3 \times$ a penalty if $(\langle \hat{F} \rangle_{nom} - \langle \hat{F} \rangle) / \langle \hat{F} \rangle_{nom} > 0$; with respect to eddy-promoter placement \hat{a} and radius \hat{R} ; for various objective-function weights $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ and thermal loads $\langle \hat{F} \rangle_{nom}$.

(Notational aside: dimensional variables shall carry carats; length, velocity, and time will be nondimensionalized with respect to \hat{h} , $d\hat{p}/d\hat{x}_1\hat{h}^2/2\hat{\rho}\hat{\nu}$, and $2\hat{\rho}\hat{\nu}/d\hat{p}/d\hat{x}_1\hat{h}$, respectively; temperature will be measured relative to \hat{T}_0 and nondimensionalized with respect to $\widehat{\Delta T} = \hat{T}_1 - \hat{T}_0$. The incompressible working fluid is characterized by a constant density, $\hat{\rho}$, kinematic viscosity, $\hat{\nu}$, thermal conductivity, \hat{k} , and thermal diffusivity, $\hat{\alpha}$. The domain associated with one periodicity length of the channel (arbitrarily positioned as shown in Figure 1) will be denoted \hat{D} , with the bottom and top walls denoted \hat{B}_1 and \hat{B}_0 , respectively, and the eddy-promoter surface denoted \hat{B}_C . A generic point in \hat{D} is (\hat{x}_1, \hat{x}_2) ; the fluid velocity, pressure (perturbation from the imposed linear field), and temperature are written as $\hat{u} = \hat{u}_1\hat{e}_1 + \hat{u}_2\hat{e}_2$, \hat{p}' , and \hat{T} , respectively, where (\hat{e}_1, \hat{e}_2) are the unit vectors associated with the two coordinate directions. Angular brackets, $\langle \cdot \rangle$, refer to time average with respect to a temporally stationary state.)

We can pose the (nondimensional) eddy-promoter heat exchanger optimization problem as an instantiation of the general framework of Section 2.1: $\Phi, \phi \mapsto \Phi^{EP}, \phi^{EP}$; $M \mapsto M^{EP} = (5) 2$, the number of design variables; $p \mapsto p^{EP} \equiv \{(Re), (Pr), a, R, (L)\}$; $\Omega \mapsto \Omega^{EP} = \{.1 \leq a \leq 1, .05 \leq R \leq a - .05\}$ (see Figure 2); $N \mapsto N^{EP} = 4$, the number of definition variables;

$\Delta \mapsto \Delta^{\text{EP}} \equiv \{\beta_1, \beta_2, \beta_3, \kappa\}$; $\Lambda \mapsto \Lambda^{\text{EP}} = \mathcal{R}_+^4$. Here Re and Pr denote the Reynolds number and Prandtl number, defined as $Re = -\overline{d\hat{p}/d\hat{x}}\hat{h}^3/2\hat{\rho}\hat{\nu}^2$ and $Pr = \hat{\nu}/\hat{\alpha}$, respectively, and $\kappa = \langle \hat{F} \rangle_{nom} 2\hat{h}/\hat{k}\widehat{\Delta T}$ is the nondimensional thermal load. Note that for the purposes of our optimization problem the Reynolds number, Prandtl number, and eddy-promoter pitch are fixed at $Re = 300$, $Pr = 1$, and $L = 6.666$, respectively, leaving only two design variables, the nondimensional eddy promoter placement, a , and radius, R ; we have indicated parenthetically that a more general optimization problem might involve five design variables, in which Re , Pr , and L are also free to vary.

As our objective function we take

$$\Phi^{\text{EP}}(a, R; \Delta^{\text{EP}}) = \phi^{\text{EP}}(\mathcal{G}(a, R), \mathcal{Q}(a, R); R, (Re); \Delta^{\text{EP}}) \quad (7)$$

where

$$\phi^{\text{EP}}(g, q; R, (Re); \beta_1, \beta_2, \beta_3, \kappa) = \beta_1(Re)^2 g + \beta_2 R^2 + \beta_3 \mathcal{H}(1 - q/\kappa). \quad (8)$$

Here $g = \mathcal{G}(a, R)$ is the time-averaged flowrate through the channel,

$$\mathcal{G}(a, R) = \frac{1}{2L} \int_D \langle u_1 \rangle dx, \quad (9)$$

and $q = \mathcal{Q}(a, R)$ is the time-averaged Nusselt number,

$$\mathcal{Q}(a, R) = \langle \hat{F} \rangle 2\hat{h}/\hat{k}\widehat{\Delta T}. \quad (10)$$

From $\langle \hat{F} \rangle = \frac{1}{L} \int_{B_1} \langle -k \frac{\partial \hat{T}}{\partial x_2} \rangle dx_1$, it follows that

$$\mathcal{Q}(a, R) = \frac{2}{L} \int_{B_1} \langle -\frac{\partial T}{\partial x_2} \rangle dx_1. \quad (11)$$

It is readily shown that, as required, $(Re)^2 g$ is the pumping power per periodicity length per unit depth (nondimensionalized by $4\hat{\rho}\hat{\nu}^3/\hat{h}^3$), and $1 - q/\kappa$ is $(\langle \hat{F} \rangle_{nom} - \langle \hat{F} \rangle) / \langle \hat{F} \rangle_{nom}$. The penalty function $\mathcal{H}(z)$ is chosen to be $\mathcal{H}(z) = H(z)z^2$, where $H(z)$ is the Heaviside distribution. It is critical to note that, even in a real (not just illustrative) design exercise, the particular, initial, form for the objective function is not overly important, as surrogate techniques are designed to permit significant variation in the objective function at low marginal cost. Indeed, it is often only through observing optima and varying the objective function that design-goal intuition is clearly articulated.

2.3.2 The Eddy-Promoter Subproblem

We next identify our eddy-promoter subproblem with the general subproblem described in Section 2.1: $M \mapsto M^{\text{EP}} = (5) 2$, the number of inputs; $\underline{p} \mapsto \underline{p}^{\text{EP}} \equiv \{(Re), (Pr), a, R, (L)\}$; $\Omega \mapsto \Omega^{\text{EP}} = \{.1 \leq a \leq 1, .05 \leq R \leq a - .05\}$; $K \mapsto K^{\text{EP}} = 2$, the number of outputs; $\underline{z} \mapsto \underline{z}^{\text{EP}} \equiv \{g, q\}$; $\underline{\mathcal{L}}(\underline{p}) \mapsto \underline{\mathcal{L}}^{\text{EP}}(\underline{p}) \equiv \{\mathcal{G}(\underline{p}), \mathcal{Q}(\underline{p})\}$; $\underline{\mathcal{L}} \mapsto \{\frac{1}{2L} \int_D \langle u_1 \rangle dx, \frac{2}{L} \int_{B_1} \langle -\frac{\partial T}{\partial x_2} \rangle dx_1\}$; $\mathbf{U} \mapsto \{\mathbf{u}, p', T\}$; $\mathbf{M}_2, \mathbf{A}_2 \mapsto$ Navier-Stokes and forced convection. More explicitly, the Navier-Stokes field subproblem for $(\mathbf{u}(x_1, x_2, t), p'(x_1, x_2, t))$ is given by,

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} = -\frac{\partial p'}{\partial x_i} + \frac{1}{Re} \frac{\partial^2 u_i}{\partial x_j \partial x_j} + \frac{2}{Re} \delta_{i1} \quad \text{in } D \quad (12)$$

$$\frac{\partial u_i}{\partial x_i} = 0 \quad \text{in } D \quad (13)$$

$$\mathbf{u} = 0 \quad \text{on } B_0 \cup B_1 \cup B_C \quad (14)$$

$$(\mathbf{u}, p')(x_1 + mL, x_2, t) = (\mathbf{u}, p')(x_1, x_2, t) \quad \forall m \in \mathbf{Z}, \quad (15)$$

and the forced-convection energy equation for $T(x_1, x_2, t)$ is given by,

$$\frac{\partial T}{\partial t} + u_j \frac{\partial T}{\partial x_j} = \frac{1}{Re Pr} \frac{\partial^2 T}{\partial x_j \partial x_j} \quad \text{in } D \quad (16)$$

$$T = 0 \quad (1) \quad \text{on } B_0 (B_1) \quad (17)$$

$$\frac{\partial T}{\partial n} = 0 \quad \text{on } B_C \quad (18)$$

$$T(x_1 + mL, x_2, t) = T(x_1, x_2, t) \quad \forall m \in \mathbf{Z}. \quad (19)$$

Here δ_{ij} is the Kronecker-delta symbol, $\partial/\partial n$ denotes differentiation in the direction of the boundary normal, \mathbf{Z} is the set of integers, $Re = 300$ and $Pr = 1$, free indices range over $\{1, 2\}$, and summation (\sum_1^2) over repeated indices is assumed. Initial conditions are not important since, at this Reynolds number, only a single attractor appears to exist for all $\{a, R\} \in \Omega^{\text{EP}}$.

We solve the Navier-Stokes and energy equations with the NEKTON code on the Intel iPSC/860 multiprocessor. The numerical method comprises: fractional timestepping schemes [26–28]; spectral element spatial discretizations [29]; and parallel [26,30] deflated [31] multilevel conjugate gradient [28] iterative solution procedures. A typical calculation proceeds by: specification of $\{a, R\}$; automatic spectral element mesh generation from several skeletal

templates; automatic parallel partition; integration in time until a temporally stationary state is achieved; evaluation of the requisite output functionals. The spectral element templates are constructed to permit relatively undistorted meshes and adequate resolution even for extreme cylinder placements and sizes. We have confirmed the mesh independence of our calculations [32], and have verified our results, where possible, with other numerical predictions and experiment [33–37]. For a typical $\{a, R\} \in \Omega^{\text{EP}}$, roughly \$75 and 6 16-processor hours are needed to reach the steady or steady-periodic state required to evaluate the flowrate and Nusselt number; note that each subproblem evaluation would cost as much as \$750 on a *single*-processor supercomputer [26].

2.3.3 Flow Phenomena

We aim to find $\Phi_{\min}^{\text{EP}}(\underline{\lambda})$ and $\underline{p}^{\text{EP}^*}(\underline{\lambda}^{\text{EP}}) \equiv \{a^*(\underline{\lambda}^{\text{EP}}), R^*(\underline{\lambda}^{\text{EP}})\}$ such that, $\forall \{a, R\} \in \Omega^{\text{EP}}, \Phi_{\min}^{\text{EP}}(\underline{\lambda}^{\text{EP}}) = \Phi^{\text{EP}}(a^*(\underline{\lambda}^{\text{EP}}), R^*(\underline{\lambda}^{\text{EP}}); \underline{\lambda}^{\text{EP}}) \leq \Phi^{\text{EP}}(a, R; \underline{\lambda}^{\text{EP}})$. We emphasize that this design problem is not local: the triangular admissible design space, Ω^{EP} , shown in Figure 2, admits virtually all geometrically possible cylinder placements and radii. More importantly, different flow phenomena occur in different regions of the design space. Figure 3 depicts the isotherms for three representative $\{a, R\}$ points in Ω^{EP} ; for $a \approx .5$ unsteady steady-periodic supercritical Tollmien-Schlichting wall-mode channel waves obtain (Figure 3a); for large R steady wavy flows predominate (Figure 3b); for very large R the flow is effectively blocked (Figure 3c) [32]. (Recall that the Navier-Stokes calculation is at fixed pressure gradient, not fixed flowrate.)

This paper is not concerned with the details of eddy-promoter flows except to the extent that these details illustrate essential aspects of the surrogate procedure. Readers interested in more details on *dissipation-transport optimization* of eddy-promoter systems are referred to [32,34–36]. These studies treat more realistic boundary conditions (e.g., in which the heat is carried away by the fluid flow) and optimization objective functions, and address a wider range of both laminar flows (numerically and experimentally) and transitional and turbulent flows (experimentally). However, extensive optimization with respect to geometric inputs (e.g., a and R) has not been undertaken, due to the expense (in this case, in both the numerical and experimental contexts) associated with system modification and re-analysis. Readers interested in more details on the *flow physics*, in particular the hydro-

dynamic stability, of eddy-promoter (and related) flows may consult [32,35–38]; these papers interpret eddy-promoter bifurcations as the interaction of simpler-geometry shear, cylinder, and channel instabilities. Recent quiet experiments and inflow-outflow numerical simulations [37] indicate that the initial instability is convective, not absolute; however, noisier experiments [36] agree quite well with periodic calculations, suggesting that in engineering applications the assumption of spatial periodicity may be acceptable for sufficiently long channels.

We claim the eddy-promoter subproblem satisfies the three conditions of Section 2.2: the computation is expensive and time-consuming (C1); bifurcations preclude sharp regularity estimates (C2); complex time-dependence and geometry variation precludes continuation methods (C3).

2.4 The Surrogate Approach

In the “direct insertion” approach to simulation-based optimization, the objective function $\Phi(\underline{p}, \underline{\lambda})$ is evaluated, for each candidate \underline{p} , as $\phi(\underline{s}; \underline{p}; \underline{\lambda})$, where

$$\underline{p} \in \Omega \xrightarrow{M_{\underline{p}}, A_{\underline{p}}} \underbrace{U(\mathbf{x}, t; \underline{p})}_{\mathcal{S}(\underline{p})} \xrightarrow{J} \underline{s} \in \mathbb{R}^K .$$

The disadvantages of this approach are described in the Introduction: the number of evaluations of $\Phi(\underline{p}, \underline{\lambda})$ is not known *a priori* — resources may be exhausted before a sensible design is proposed; simulations evoked in a first optimization study with objective function $\Phi(\underline{p}, \underline{\lambda}^1)$ will be of limited use in a second optimization study with objective function $\Phi(\underline{p}, \underline{\lambda}^2)$ — design adaptability is frustrated; and systematic fusion of simulation results with prior analytical, heuristic, or experimental information is, at best, difficult. In short, it is difficult to perceive of a day-long thousand-dollar Navier-Stokes simulation as a function call from a mathematical programming routine.

In the surrogate alternative, the subproblem simulation is evoked only to construct and validate a simplified input-output model, $\tilde{\mathcal{S}}(\underline{p}): \Omega \rightarrow \mathbb{R}^K$, which is intended to approximate $\mathcal{S}(\underline{p})$ over the input domain Ω . This simplified input-output model then serves as a simulation surrogate in subsequent optimization studies, that is, $\Phi(\underline{p}, \underline{\lambda}) = \phi(\mathcal{S}(\underline{p}); \underline{p}; \underline{\lambda})$ is replaced with $\phi(\tilde{\mathcal{S}}(\underline{p}); \underline{p}; \underline{\lambda}) = \tilde{\Phi}(\underline{p}, \underline{\lambda})$: $\tilde{\Phi}(\underline{p}, \underline{\lambda})$, not $\Phi(\underline{p}, \underline{\lambda})$, is minimized. Given the assumptions of Section 2.1, $\tilde{\Phi}(\underline{p}, \underline{\lambda})$ can be inexpensively evaluated for any

candidate p as $\phi(\tilde{z}; p; \lambda)$, where

$$p \in \Omega \xrightarrow{\tilde{z}(p)} \tilde{z} \in R^K .$$

The surrogate problem can thus be stated as follows: *Given* a limited (or even fixed) number of appeals (recall conditions C1 and C3) to a largely uncharacterized (recall condition C2) but deterministic function, $\mathcal{Q}(p)$; *Find* a simple but validated approximation to $\mathcal{Q}(p)$ over Ω , $\tilde{\mathcal{Q}}(p)$, which a) conservatively but effectively exploits prior information, and b) can be gainfully incorporated into design and optimization studies.

The advantages of surrogates are manifold: surrogates are, by construction, inexpensive, and can be evaluated *ad infinitum* — premature termination of the optimization procedure will not be required; a single surrogate can support a large λ -family of related objective functions — adaptive, non-incremental modification of design criteria and specifications is encouraged; surrogates can readily incorporate prior extra-numerical information concerning not only regularity, but also form — thereby reducing the computational burden. The primary disadvantage of surrogates is the introduction of new errors into the optimization process due to the additional level of approximation, $\phi(\mathcal{Q}(p); p; \lambda) \approx \phi(\tilde{\mathcal{Q}}(p); p; \lambda)$; this “purposiveness” issue is discussed in depth in Section 4. The many advantages (and significant disadvantage) of surrogates have long been recognized: computational scientists typically search for “insight not numbers”; engineers often exploit reduced-order models. However, with the exception of relatively recent work [20–23]: the surrogate concept is rarely explicitly articulated; application of the surrogate concept is typically *ad hoc*; and surrogates are not usually accompanied by useful error estimates. It is the latter shortcomings that we aim to partially mitigate.

The surrogate problem statement and the complexity conditions C1, C2, and C3 suggest several broad methodological guidelines. First, from conditions C1 and C3, we require error estimates for a *fixed* number of function evaluations; this implicates a statistical approach, in which uncertainty can be precisely accommodated. Second, from condition C2, we must presume that, in the general case, we know relatively little about our input–output function; this implicates a nonparametric statistical approach. Third, from conditions C1 and C2, design sensitivity derivatives, though a powerful tool for both gradient-based minimization and post-optimization sensitivity analyses

[1,39,40], may not be sufficient: more global, general models for objective-function approximation must be admitted. Fourth, from conditions C1, C2, and C3, we deduce that neither *a priori* nor *a posteriori* regularity-based approximation and estimation techniques are effective: we can not hope to be asymptotic (C1); we will have very little insight into the proper norm, form, or constants of approximation errors (C2); *a posteriori* error analysis based on local subproblems or extrapolation will not be computationally viable (C3). These considerations suggest that new error norms are required. Fifth, from condition C3, we can assume at least partial decoupling of the parent optimization problem and the expensive subproblem; although results of the former will certainly affect the region in which we choose to examine the latter, the two tasks remain computationally relatively independent. Armed with this methodological outline, we now proceed to discuss the particular algorithms developed.

Remark on Modelling. We assume here that our *mathematical* model, \mathcal{M}^0 , accurately reflects the *physical* problem of interest, denoted \mathcal{M}^{00} . Our computational surrogate approach implicitly considers the modelling process in two stages: $\mathcal{M}^{00} \rightarrow \mathcal{M}^0 \rightarrow \mathcal{M}$. An alternative, one-stage, approach, $\mathcal{M}^{00} \rightarrow \mathcal{M}$, proceeds directly from the physical problem, \mathcal{M}^{00} , to a computationally simple engineering model, \mathcal{M} , without passing through a large-scale-simulation intermediary, \mathcal{M}^0 (or physical experiment). We believe the two-stage approach is preferable, as the more difficult problem of physical-to-mathematical translation is conducted at a level which accommodates greater complexity. This greater flexibility should not, of course, serve as an excuse for *less* discriminating modelling practices. \square

3 Modelling Methodology

In this section we consider both validation procedures, in which we assess a given $\tilde{\mathcal{G}}(p)$, and construction-validation procedures, in which we both propose and assess $\tilde{\mathcal{G}}(p)$. As much of [2] is focussed on the motivation, analysis, and empirical verification of the validation and construction-validation algorithms, we confine ourselves here to a brief summary of the major points. We restrict ourselves initially to a single output, $K = 1$; extension of the theory to multiple outputs is described in Section 5.

3.1 Validation

3.1.1 Algorithm

The validation algorithm takes as given: (i) A subproblem, with an input M -vector, \underline{p} , an input domain, $\Omega \in \mathbb{R}^M$, a (single) output, $s \in \mathbb{R}$, and an input-output function, $\mathcal{S}(\underline{p}) \in L^\infty(\Omega)$. We prefer, but do not require, that the subproblem satisfy the three complexity conditions C1, C2, and C3, of Section 2.2. (ii) A proposed surrogate, $\tilde{\mathcal{S}}(\underline{p}) : \Omega \rightarrow \mathbb{R}$. (iii) A strictly positive Bayesian importance function,

$$\rho(\underline{p}) : \Omega \rightarrow \mathbb{R}_+, \quad \int_{\Omega} \rho(\underline{p}) d\underline{p} = 1, \quad (20)$$

which describes the *a priori* relative importance ascribed to different points within the input domain Ω . As will be discussed in Section 4, this importance function is best interpreted as a prior “density” for $\underline{p}^*(\underline{\lambda})$. We note that, for the error statements developed in Section 3.1.2, the importance function is *required* to ensure input-transformation objectivity. (iv) The maximum number of $\mathcal{S}(\underline{p})$ evaluations permitted, N^{ev} . This parameter describes the resource limitation associated with the validation exercise. (v) Two validation error tolerances, $\varepsilon_1, \varepsilon_2 \in [0, 1]^2$, the significance of which will become clear in the validation error statement of Section 3.1.2.

We now summarize the simple Monte-Carlo [14] Model Validation (MV) Algorithm of [2]. We note that this algorithm is, in effect, nothing more than randomization of obvious parameter-space exploration procedures; however, the introduction of a probabilistic framework permits an error statement which marries well with subsequent optimization analyses.

Algorithm $MV(\mathcal{S}(\underline{p}), \tilde{\mathcal{S}}(\underline{p}), \rho(\underline{p}), N^{ev}, \varepsilon_1, \varepsilon_2)$

1. SET $N^{va} = \mathcal{N}(\varepsilon_1, \varepsilon_2)$,

$$\mathcal{N}(\varepsilon_1, \varepsilon_2) = \frac{\ln \varepsilon_2}{\ln(1 - \varepsilon_1)}. \quad (21)$$

2. IF $N^{va} > N^{ev}$, validation is not possible.
3. FOR $j = 1, \dots, N^{va}$
 - S1. DRAW $\underline{P}_j \sim \rho(\underline{p})$
 - S2. COMPUTE $S_j = \mathcal{S}(\underline{P}_j)$.

4. SET $E_{\max} = \max_{j \in \{1, \dots, N^{\text{va}}\}} E_j$, where $E_j = |S_j - \tilde{S}(\underline{P}_j)|$. \square

Here $\underline{X} \sim f(\underline{x})$ refers to a random vector \underline{X} with probability density function $f(\underline{x})$; we shall indicate a particular realization of a random quantity Y (variable, vector, or domain) as $\Re Y$. In the MV Algorithm, the \underline{P}_j are random vectors, and the S_j , E_j , and E_{\max} are random variables.

3.1.2 Error Analysis

The output of the MV Algorithm, the model prediction error estimator, E_{\max} , is related to the model validation error estimate, ε_1 , and the validation statement uncertainty, ε_2 , by a precise probabilistic statement,

$$\Pr\left\{\int_{\{\underline{p} \in \Omega \mid |\mathcal{S}(\underline{p}) - \tilde{\mathcal{S}}(\underline{p})| \leq E_{\max}\}} \rho(\underline{p}) d\underline{p} \geq 1 - \varepsilon_1\right\} \geq 1 - \varepsilon_2, \quad (22)$$

where $\Pr\{\text{event}\}$ is the probability that *event* occurs. In words, (22) states that, with probability greater than or equal to $1 - \varepsilon_2$, $|\mathcal{S}(\underline{p}) - \tilde{\mathcal{S}}(\underline{p})| \leq E_{\max}$ over a region of Ω of relative weighted volume greater than or equal to $1 - \varepsilon_1$; equivalently, with probability greater than or equal to $1 - \varepsilon_2$, $|\mathcal{S}(\underline{p}) - \tilde{\mathcal{S}}(\underline{p})| > E_{\max}$ over a region of Ω of relative weighted volume no greater than ε_1 . (The probability ensemble here is defined with respect to repetition of the algorithm: we expect that, in greater than $1 - \varepsilon_2$ of all realizations of the algorithm, $|\mathcal{S}(\underline{p}) - \tilde{\mathcal{S}}(\underline{p})| \leq E_{\max}$ over a region of Ω of weighted relative volume greater than $1 - \varepsilon_1$.)

The critical aspects of the validation procedure are: first, a precise (albeit probabilistic) error statement, (22), can be made for a *fixed* number of evaluations of $\mathcal{S}(\underline{p})$; second, the sample-size requirement, (21), and resulting error estimate, (22), are nonparametric, valid for any functions $\mathcal{S}(\underline{p})$ and $\tilde{\mathcal{S}}(\underline{p})$; third, the validation statement, (22), requires *no assumptions* on $\mathcal{S}(\underline{p})$ or $\tilde{\mathcal{S}}(\underline{p})$ as to regularity or functional form. Perhaps most importantly, the error estimate will also prove amenable to *a posteriori* analysis in the optimization context (see Section 4). The error statement (22) is readily interpreted in the probably approximately correct framework developed for classification problems in the theory of learning [12,13]; in the probably-approximately-correct context, finite uncertainty — in our case represented by ε_1 and ε_2 — permits a precise statement for a fixed sample size.

The derivation of (22) is given in [2] in terms of order-statistic tolerance limits [16,41,42]. We indicate here an alternative derivation, based on binomial considerations, that has the advantage of direct (multinomial) extension to the multiple-output case (see Section 5). We define F_E to be the (perforce increasing, though not necessarily strictly increasing, nor continuous) cumulative distribution function of the random variable $E = |\mathcal{S}(\underline{P}) - \tilde{\mathcal{S}}(\underline{P})|$, where \underline{P} is a random vector with probability density $\rho(\underline{p})$. The $1 - \varepsilon_1$ quantile of E , $e_{1-\varepsilon_1}$, is then defined by $F_E(e_{1-\varepsilon_1}) = 1 - \varepsilon_1$ (more precisely, $e_{1-\varepsilon_1}$ is the minimum z such that $F_E(z) \geq 1 - \varepsilon_1$). Lastly, j_{\max} is any $j \in \{1, \dots, N^{va}\}$ for which $E_j = E_{\max}$. Then, if $\mathcal{T}_{1-\varepsilon_1} = \{\underline{p} \in \Omega \mid |\mathcal{S}(\underline{p}) - \tilde{\mathcal{S}}(\underline{p})| < e_{1-\varepsilon_1}\}$,

$$\Pr\{\forall j \in \{1, \dots, N^{va}\}, \underline{P}_j \in \mathcal{T}_{1-\varepsilon_1}\} \leq (1 - \varepsilon_1)^{N^{va}}. \quad (23)$$

It follows that *at least one* \underline{P}_j will lie in $\Omega \setminus \mathcal{T}_{1-\varepsilon_1} = \{\underline{p} \in \Omega \mid |\mathcal{S}(\underline{p}) - \tilde{\mathcal{S}}(\underline{p})| \geq e_{1-\varepsilon_1}\}$ with probability greater than or equal to $1 - (1 - \varepsilon_1)^{N^{va}}$. Furthermore, if at least one \underline{P}_j lies in $\Omega \setminus \mathcal{T}_{1-\varepsilon_1}$, then, since $E_{\max} \geq E_j, \forall j \in \{1, \dots, N^{va}\}$, $\underline{P}_{j_{\max}}$ in particular will lie in $\Omega \setminus \mathcal{T}_{1-\varepsilon_1}$, and thus $F_E(E_{\max}) \geq 1 - \varepsilon_1$. Finally, recognizing $F_E(e) = \int_{\{\underline{p} \in \Omega \mid |\mathcal{S}(\underline{p}) - \tilde{\mathcal{S}}(\underline{p})| \leq e\}} \rho(\underline{p}) d\underline{p}$, and substituting from (21) $(1 - \varepsilon_1)^{N^{va}} = \varepsilon_2$, we obtain (22). This binomial derivation of (22) is essentially a classification argument; not surprisingly, the sample-size requirement (21) also appears in [13].

The origin of uncertainty in (22) is a random region

$$\mathcal{U} = \{\underline{p} \in \Omega \mid |\mathcal{S}(\underline{p}) - \tilde{\mathcal{S}}(\underline{p})| > E_{\max}\},$$

$1 - \varepsilon_2$ probably of relative weighted volume less than or equal to ε_1 , of *undetermined* location and shape, over which the surrogate misfit, $|\mathcal{S}(\underline{p}) - \tilde{\mathcal{S}}(\underline{p})|$, is *unknown*. The usual confidence interval-confidence level balance inherent in (21) has an interesting interpretation: if we consider $-\ln \varepsilon_2$ to be *how well* we know the simulation behavior, and $-1/\ln(1 - \varepsilon_1)$ to be *how much* of the simulation behavior we know, then, for a fixed number of appeals to $\mathcal{S}(\underline{p})$, N^{va} , (21) implies that the product of *how well* and *how much* is fixed (in fact, equal to N^{va}). If we choose the deterministic limit, in which we tolerate no uncertainty ($\varepsilon_2 \rightarrow 0$), then *how well* we know the simulation behavior tends to infinity, but *how much* of the simulation behavior we know tends to zero: we know the simulation behavior only at the points sampled, which is a set of measure zero. By permitting finite but controlled uncertainty in *how well*,

surrogates accommodate a finite *how much*: surrogates – probable but global — bridge the gap between direct computation, which is sure but pointwise, and analytical methods, which are sure and global.

The balance between ε_1 and ε_2 is not “symmetric,” however. In particular, $\varepsilon_1 \sim -\ln \varepsilon_2 / N^{va}$ as $N^{va} \rightarrow \infty$ for ε_2 fixed, corresponding to rather slow algebraic decay. For example, for $N^{va} = 22$, we can choose $\varepsilon_1 = .1, \varepsilon_2 = .1$; doubling the number of evaluations to $N^{va} = 44$ reduces ε_1 by only a factor of two, $\varepsilon_1 = .05, \varepsilon_2 = .1$. In contrast, $\varepsilon_2 \sim e^{-\varepsilon_1 N^{va}}$ as $N^{va} \rightarrow \infty$ for ε_1 fixed, corresponding to rapid exponential decay. For example, again beginning with $N^{va} = 22$ and $\varepsilon_1 = .1, \varepsilon_2 = .1$, doubling the number of evaluations to $N^{va} = 44$ permits a tenfold decrease in ε_2 , $\varepsilon_1 = .1, \varepsilon_2 = .01$. It follows that, with only a modest number of evaluations of $\mathcal{S}(\underline{p})$, ε_2 will be sufficiently small that we can assume with near certainty that \mathcal{U} is, indeed, of relative weighted volume less than or equal to ε_1 ; the remaining uncertainties are the location of \mathcal{U} , and the surrogate misfit, $|\mathcal{S}(\underline{p}) - \tilde{\mathcal{S}}(\underline{p})|$, over \mathcal{U} .

Remark on Dimensionality. Our algorithm, sampling requirement (21), and error estimate (22) apply *independent* of input-vector dimensionality, M . However, this generality is deceptive; as M increases, although the relative volume of \mathcal{U} remains invariant, \mathcal{U} will reflect increasingly significant excursions in individual components of the input vector (e.g., compare the side length of a square and cube of the same volume). This implies that our technique will not be viable for too many subproblem inputs (although the number of design variables may be large). Surrogate techniques should, however, be extensible to problems of shape optimization [40], which are essentially infinite-dimensional, if geometrically motivated correlations between the inputs are introduced in order to reduce — through $\rho(\underline{p})$ — the effective volume of the input domain. \square

3.1.3 Eddy-Promoter Example

We apply the validation procedure to the eddy-promoter Nusselt number, $Q(a, R)$. In order to evoke the MV Algorithm, the prerequisites listed in Section 3.1.1 must be supplied. The necessary quantities are, in fact, all defined in Section 2.3, save the proposed surrogate, $\tilde{Q}(a, R)$, the importance function, $\rho(\underline{p})$, the maximum number of evaluations, N^{ev} , and the uncertainty tolerances, ε_1 and ε_2 . For our purposes here we simply select for the surro-

gate $\tilde{Q}(a, R) = 1$, which corresponds to the Nusselt number for conduction in the channel in the absence of the eddy-promoters and any flow. (This simple surrogate is chosen for lack of a better heuristic; however, this example also illustrates application of surrogate procedures to test *global* “stability,” or sensitivity, of a solution — in this case the conduction solution — to variations in the design variables.) We take the importance function, $\rho(\mathbf{p})$, to be uniform over the triangular input domain, reflecting no prior knowledge as to which parts of the domain will prove more interesting in the ultimate optimization application. Lastly, we set $N^{ev} = 44, \varepsilon_1 = .1, \varepsilon_2 = .01$ (by construction, $N^{va} = \mathcal{N}(\varepsilon_1 = .1, \varepsilon_2 = .01) = 44 = N^{ev}$).

Implementation of the MV Algorithm is now straightforward. First, we employ a standard acceptance-rejection Monte-Carlo method [14] and a congruential pseudorandom number generator [43] to produce $N^{va} = 44$ input points which are randomly and uniformly distributed over the input domain Ω^{EP} ; the input points, $\{a_j, R_j\}, j = 1, \dots, N^{va}$, resulting from one realization of this process are shown in Figure 4. Next, the Nusselt number is computed at each of the input points, $q_j = Q(a_j, R_j), j = 1, \dots, N^{va}$, following the field subproblem evaluation procedure described in Section 2.3. Lastly, we compute $\mathcal{RE}_{\max}^q = \max_{j \in \{1, \dots, N^{va}\}} |q_j - \tilde{Q}(a_j, R_j)| = \max_{j \in \{1, \dots, N^{va}\}} |q_j - 1|$; for the realization shown in Figure 4, we find $\mathcal{RE}_{\max}^q = .236$. We thus conclude that, with confidence level greater than .99, the discrepancy between $Q(a, R)$ (the Navier-Stokes solution) and $\tilde{Q} = 1$ (our simple surrogate) is less than .236 over more than 90% of Ω^{EP} . The error in the surrogate is, expectedly, rather large, confirming that the flow departs significantly from conduction within the design space Ω^{EP} . To capture this departure in greater detail, we need to consider construction-validation procedures.

3.2 Construction-Validation

3.2.1 Algorithm

The construction-validation algorithm takes as given: (i) A subproblem, with an input M -vector, \mathbf{p} , an input domain, $\Omega \in \mathbb{R}^M$, a (single) output, $s \in \mathbb{R}$, and an input-output function, $\mathcal{S}(\mathbf{p}) \in L^\infty(\Omega)$. (ii) A modelling (approximation) procedure,

$$\mathcal{A} : (\Omega \times \mathbb{R})^{\mathcal{P}} \rightarrow L^\infty(\Omega), \quad (24)$$

which, given \mathcal{P} input-output pairs, $\Xi_{\mathcal{S}}^{\mathcal{P}} = \{(\underline{p}_1, \mathcal{S}(\underline{p}_1)), \dots, (\underline{p}_{\mathcal{P}}, \mathcal{S}(\underline{p}_{\mathcal{P}}))\}$, generates the surrogate rule, $\tilde{\mathcal{S}}(\underline{p})$. (iii) A Bayesian importance function, $\rho(\underline{p})$, satisfying (20). (iv) The maximum number of $\mathcal{S}(\underline{p})$ evaluations permitted, N^{ev} . (v) Two validation error tolerances, $\varepsilon_1, \varepsilon_2 \in [0, 1]^2$. (vi) The Model Validation Algorithm of Section 3.1.1.

We now summarize a simple Monte-Carlo Model Construction-Validation (MCV) Algorithm based on random datasets [2].

Algorithm MCV($\mathcal{S}(\underline{p}), \rho(\underline{p}), N^{ev}, \varepsilon_1, \varepsilon_2$)

1. COMPUTE $N^{va} = \mathcal{N}(\varepsilon_1, \varepsilon_2)$ from (21).
2. IF $N^{va} \geq N^{ev}$, QUIT; ELSE SET $N^{co(nstruction)} = N^{ev} - N^{va}$.
3. FOR $j = 1, \dots, N^{co}$ (random dataset):
 - S1. DRAW $\underline{P}_j \sim \rho(\underline{p})$
 - S2. COMPUTE $S_j = \mathcal{S}(\underline{P}_j)$.
4. SET $\tilde{\mathcal{S}}(\underline{p}) = \mathcal{A}(\{(\underline{P}_1, S_1), \dots, (\underline{P}_{N^{co}}, S_{N^{co}})\})$.
5. CALL MV($\mathcal{S}(\underline{p}), \tilde{\mathcal{S}}(\underline{p}), \rho(\underline{p}), N^{va}, \varepsilon_1, \varepsilon_2$) $\rightarrow E_{\max}$. \square

(For simplicity of presentation we indicate that the first N^{co} input points serve for construction and the last N^{va} input points serve for validation; in practice, a sample of N^{ev} input points is drawn and then randomly partitioned into construction and validation subsets.) The constructed model and model prediction error estimator, E_{\max} , satisfy our probabilistic validation statement, (22). Although the $\tilde{\mathcal{S}}(\underline{p})$ are, in fact, random, for the purposes of this paper we shall condition all results on a *given* model $\tilde{\mathcal{S}}(\underline{p})$.

The MCV Algorithm presented is rather crude and inefficient. First, we would prefer to compare and select amongst different surrogates, choosing that model which incurs the smallest model prediction error estimate or which is computationally least expensive [5,6]. Second, we would like to adapt to information generated during the construction-validation process; a sequential approach offers clear advantages, permitting the algorithm — and the appeals to the expensive $\mathcal{S}(\underline{p})$ — to terminate when the (or a) model prediction error estimate is sufficiently small. Both of these improvements are made possible by the multiple-output extension described in Section 5.

3.2.2 Classes of Models

Models can be characterized in several ways: by dataset $\Xi^{\mathcal{P}}$, deterministic or random; or by procedure, \mathcal{A} , “graybox” or “blackbox.” Although the modelling problem would appear to be a routine exercise, several factors complicate the process. First, the domain Ω will often be irregular, precluding simple tensor-product techniques. Second, the input vector and domain Ω may be of high dimension, M : most complex-geometry interpolation procedures developed for partial-differential-equation applications in two or three space dimensions are increasingly cumbersome or computationally intensive with increasing space dimension; local, linearized models commonly used for multivariate response surfaces are inappropriate for our (global) purposes. Third, random datasets offer certain advantages within the surrogate context: scattered-data approximation procedures [17] are considerably more problematic than ordered datasets.

We begin by comparing the relative advantages (marked with a +) and disadvantages (marked with a -) of deterministic and random datasets. First, deterministic datasets: (+) ensure the anticipated distribution is realized; (+) can exploit existing datasets; (+) permit a range of well-developed approximation procedures (tensor-product, “finite-element”); (+) permit *a priori* regularity-based approximation-error estimates; (-) extend with some difficulty to complex Ω , in particular for larger M ; (-) preclude recycling of datapoints for (perforce random) validation (see [2] and Section 5). Random datasets: (-) exhibit fickle “distribution”; (-) disqualify existing (deterministic) data; (-) permit only scattered-data approximation methods, such as Voronoi methods [2,44], modified Shepard techniques [17,45], and radial-basis-function approaches [11]; (-) permit only limited *a priori* regularity-based approximation error estimates; (+) extend readily to complex Ω ; (+) extend readily to larger M ; (+) permit recycling for validation.

We also briefly compare graybox and blackbox modelling approaches. Graybox models are intended to reflect prior information as to the anticipated form of the phenomenon under consideration. The resulting model is “parametric,” involving a finite number of to-be-determined basis coefficients representing a fixed-dimensional approximation space. Blackbox models are intended to be largely unbiased as to possible functional form, though clearly some minimal regularity assumptions are required. Blackbox models are preferably “nonparametric,” permitting the approximation

of arbitrary functions to arbitrary accuracy by consideration of a family of approximation spaces of increasingly large dimension. Graybox and blackbox approaches can be gainfully combined as blackbox-corrected graybox models.

In [2] we develop a random-dataset Voronoi-based piecewise-constant blackbox approximation procedure for complex domains which extends directly and efficiently (linearly in complexity) with increasing input dimension, M ; the technique is applied to a problem of stress concentration in linear elasticity. Unfortunately, although the Voronoi method, which is effectively a piecewise constant finite element approximation over convex tiles, does enjoy certain approximation properties, the method is too low-order to make effective use of the perforce limited datasets available for surrogate construction. In the current paper we choose for our construction procedure the two-dimensional implementation of the scattered-data (random dataset) modified Shepard method, $\mathcal{A} \mapsto \text{QSHEP2D}$ (ACM Algorithm 660, [45]). At present both the Voronoi and Shepard methods are “Lagrangian,” based only on function values; as numerical and automatic differentiation techniques [46,47] become better developed, “Hermitian” approximations incorporating sensitivity derivatives may prove more efficient. General multivariate (large- M) approximation theory remains an open research area.

3.2.3 Eddy-Promoter Example

We now apply the construction-validation procedure to the eddy-promoter flowrate and Nusselt number, $\mathcal{G}(a, R)$ and $\mathcal{Q}(a, R)$, respectively. In order to evoke the MCV Algorithm, the prerequisites listed in Section 3.1.1 must be supplied. The necessary quantities are all defined in Section 2.3, save the approximation procedure \mathcal{A} , the importance function, $\rho(\underline{p})$, the maximum number of evaluations, N^{ev} , and the uncertainty tolerances, ε_1 and ε_2 . As described in Section 3.2.2, we use the Renka [45] implementation of the modified Shepard method as our construction procedure. We take the importance function, $\rho(\underline{p})$, to be uniform over the triangular input domain, reflecting no prior prejudice as to areas of potentially higher interest. Lastly, we set $N^{ev} = 44$, $\varepsilon_1 = .1$, and $\varepsilon_2 = .1$ ($N^{va} = \mathcal{N}(\varepsilon_1 = .1, \varepsilon_2 = .1) = 22$, and thus $N^{co} = N^{va} = 22$).

Implementation of the MCV Algorithm is now straightforward. First, as in the MV Algorithm, we employ a standard acceptance-rejection Monte-

Carlo method and a congruential pseudorandom number generator to produce $N^{ev} = 44$ points which are randomly and uniformly distributed over the input domain Ω^{EP} . Then, the flowrate and Nusselt number are computed at each of these input points, $\{g_j, q_j\} = \{\mathcal{G}(a_j, R_j), \mathcal{Q}(a_j, R_j)\}, j = 1, \dots, N^{ev}$, following the field subproblem evaluation procedure described in Section 2.3. Next, these $N^{ev} = 44$ input-output pairs are randomly partitioned into two subsets, a construction set of size $N^{co} = 22$ for Step 4 of the MCV Algorithm, and a validation set of size $N^{va} = 22$ for Step 5 of the MCV Algorithm. Finally, the surrogates are formed and tested. Our construction method is slightly modified to include prior information: QSHEP2D is evoked in Step 4 not with N^{co} points, but with $N^{co} + 2$ points. The two additional contributions comprise a “plane-Poiseuille-flow” input-output pair, $(\{a, R\} = \{0, 0\}, g = 2/3)$, $(\{a, R\} = \{0, 0\}, q = 1)$, and a prior-work eddy-promoter Tollmien-Schlichting input-output pair, $(\{a, R\} = \{.5, .2\}, g = .311)$, $(\{a, R\} = \{.5, .2\}, q = 1.12)$. (Although the Nusselt number for the extensively studied $\{a, R\} = \{.5, .2\}$ geometry [35–37] can be estimated from published data [35], we prefer to exactly recompute g and q for the precise boundary conditions of the current paper.)

For the particular (single) train-test realization of Figure 5, we obtain the flowrate and Nusselt number surrogates shown in Figure 6, and model prediction error estimates for the flowrate and Nusselt number of $\mathfrak{RE}_{\max}^g = .035$ and $\mathfrak{RE}_{\max}^q = .092$, respectively. We thus conclude that, with confidence level greater than .90, the discrepancy between $\mathcal{G}(a, R)$ (the Navier-Stokes solution) and $\tilde{\mathcal{G}}$ (our surrogate) is less than .035 over more than 90% of Ω^{EP} ; similarly (but not jointly, see below), with confidence level greater than .90, the discrepancy between $\mathcal{Q}(a, R)$ (the Navier-Stokes solution) and $\tilde{\mathcal{Q}}$ (our surrogate) is less than .092 over more than 90% of Ω^{EP} . Discussing first the flowrate, we see from Figure 6 that, not surprisingly, for our *fixed* pressure gradient, the flowrate decreases for cylinders either farther away from the wall or of larger radius: both of these variations increase the drag on the eddy-promoter. The flowrate surrogate is rather accurate, with a model prediction error estimate of only .035 over an observed flowrate range of $0 < g < .667$. Turning now to the Nusselt number, we see that the largest Nusselt number obtains for the larger-cylinder *steady* wavy mechanism, but that significant transport also occurs for the *unsteady* Tollmien-Schlichting mode. The Nusselt number surrogate is less accurate than the flowrate surrogate, with a model prediction error estimate of .092 over an observed Nusselt

number range of $.764 \leq q \leq 1.186$.

The reader will notice that the input points are the same for the validation example of Section 3.1.3 (see Figure 4) and construction-validation example (see Figure 5) of the current section; furthermore, the flowrate and Nusselt number are both validated on the same set of input points. In essence, we are exploiting the databoard concept (see [2] and Section 5), in which a single set of input points is recycled for different models, outputs, or optimization studies. Note however, that, in the current single-output context ($K = 1$), each example must be treated as a separate problem — the sample-size requirement (21) and associated error estimates (22) are *not* joint. In Section 5 we describe the simple modification which permits us to state joint error estimates for multiple outputs validated over a common input set.

4 Optimization Purposiveness

4.1 Surrogate-Based Optimization

We recall from Section 2.4 that the essential aspect of surrogate optimization is the replacement of the actual objective function, $\Phi(\mathbf{p}, \lambda) = \phi(\mathcal{Q}(\mathbf{p}); \mathbf{p}; \lambda)$, with a surrogate objective function, $\tilde{\Phi}(\mathbf{p}, \lambda) = \phi(\tilde{\mathcal{Q}}(\mathbf{p}); \mathbf{p}; \lambda)$. Within this broad framework, however, several different approaches are possible. First, one can proceed in an “unvalidated mode,” in which one tests surrogate predictions only at surrogate-proposed design points. This approach has the advantage that all points are dedicated to construction, but the disadvantages that: first, even if the surrogate is accurate at the proposed design point, one has no assurances as to the accuracy of the surrogate at other input points upon which selection of the design point may be conditioned (e.g., through gradient information or simple rejection); second, if the surrogate is *not* sufficiently accurate at the proposed design point, an appropriate course of action is not clear. A second approach, “learn-by-doing” (e.g., [48]), performs construction-validation *during* the optimization process; that is, the surrogate model input sample reflects the local structure of the objective function. The learn-by-doing approach has the advantage that the importance function is, perforce, relevant to the local search process, but the disadvantage that the surrogate developed for one objective function may be inappropriate for a subsequent optimization study in which the objective

function has been modified.

We pursue here a third approach, a “Bayesian validated” approach, in which the validation importance function, $\rho(\underline{p})$, serves to indicate the anticipated relative relevance of points within the feasible design space Ω ; ideally, in a parent optimization project involving numerous optimization studies parametrized by $\underline{\lambda}$, $\underline{p}^*(\underline{\lambda})$ would be “distributed” according to $\rho(\underline{p}^*)$. It is critical to note that we rely on this definition to motivate, but not to justify, the choice of $\rho(\underline{p})$; techniques for determining the *de facto* influence of $\rho(\underline{p})$ on any *single* optimization study (that is, particular $\underline{\lambda}$) are presented in Section 4.3. The Bayesian validated approach permits better *a posteriori* error estimates than the “unvalidated mode,” though at the expense of fewer points for construction. (In fact, validation points can, subsequent to validation in Step 5 of the MCV Algorithm, be included in a revised construction, however the resulting model no longer satisfies a rigorous error statement.) The Bayesian validated approach ensures greater flexibility than the “learn-by-doing” approach, though at the cost of lower relevance for any particular study. In summary, the Bayesian validated approach is probably better suited for initial, global studies than for final, local designs.

4.2 Monte Carlo Algorithm

Our surrogate-based optimization algorithm takes as given: (i) A subproblem, with an input M -vector, \underline{p} , an input domain, $\Omega \in \mathbb{R}^M$, a (now possibly multiple) output, $\underline{g} \in \mathbb{R}^K$, and an input-output function, $\underline{g}(\underline{p}) \in L^\infty(\Omega)^K$. (ii) A modelling (approximation) procedure, $\underline{A}(\Xi_{\underline{g}}^{\underline{p}})$, as described in Section 3.2. (iii) An optimization evaluation procedure, $\phi(\underline{g}; \underline{p}; \underline{\lambda}) : \mathbb{R}^K \times \mathbb{R}^M \times \Lambda \rightarrow \mathbb{R}$. (iv) A Bayesian importance function, $\rho(\underline{p})$, satisfying (20). (v) The maximum number of $\underline{g}(\underline{p})$ evaluations permitted, N^{ev} . (vi) Two validation error tolerances, $\varepsilon_1, \varepsilon_2 \in [0, 1]^2$.

We now summarize the Surrogate-Based Optimization (SBO) Algorithm.

Algorithm SBO($\phi(\underline{g}; \underline{p}; \underline{\lambda}), \rho(\underline{p}), N^{ev}, \varepsilon_1, \varepsilon_2$)

1. COMPUTE $N^{va} = \mathcal{N}(\varepsilon_1, \varepsilon_2)$ from (21).
2. IF $N^{va} \geq N^{ev}$ QUIT; ELSE SET $N^{co} = N^{ev} - N^{va}$.
3. FOR $j = 1, \dots, N^{ev}$:
 - S1. DRAW $\underline{p}_j \sim \rho(\underline{p})$

- S2. COMPUTE $\underline{\mathcal{S}}_j = \underline{\mathcal{S}}(\underline{\mathcal{P}}_j)$.
4. SET $\tilde{\mathcal{S}}(\underline{p}) = \mathcal{A}(\{(\underline{\mathcal{P}}_1, \underline{\mathcal{S}}_1), \dots, (\underline{\mathcal{P}}_{N^{co}}, \underline{\mathcal{S}}_{N^{co}})\})$.
5. FOR $\underline{\lambda} = \underline{\lambda}^1, \underline{\lambda}^2, \underline{\lambda}^3, \dots$
- { 6. FIND $\tilde{\Phi}_{\min}(\underline{\lambda})$ and $\tilde{\underline{p}}^*(\underline{\lambda})$, where

$$\tilde{\Phi}_{\min}(\underline{\lambda}) = \phi(\tilde{\mathcal{S}}(\tilde{\underline{p}}^*(\underline{\lambda})); \tilde{\underline{p}}^*(\underline{\lambda}); \underline{\lambda}) \leq \phi(\tilde{\mathcal{S}}(\underline{p}); \underline{p}; \underline{\lambda}) \quad \forall \underline{p} \in \Omega. \quad (25)$$

7. SET $E_{\max}^{\Phi} = \max_{j \in \{N^{co}+1, \dots, N^{ev}\}} E_j^{\Phi}$, where

$$E_j^{\Phi} = |\phi(\underline{\mathcal{S}}_j; \underline{\mathcal{P}}_j; \underline{\lambda}) - \phi(\tilde{\mathcal{S}}(\underline{\mathcal{P}}_j); \underline{\mathcal{P}}_j; \underline{\lambda})|.$$

8. COMPUTE *a posteriori* estimates (see Section 4.3).
9. CONSIDER adaptive refinement (see Section 4.4). } □

(For simplicity of presentation we indicate that the first N^{co} input points serve for construction and the last N^{ev} input points serve for validation; in practice, the sample of N^{ev} input points is randomly partitioned into construction and validation subsets.) The $\tilde{\Phi}_{\min}(\underline{\lambda})$ and $\tilde{\underline{p}}^*(\underline{\lambda})$ of (25) will be denoted the “surrogate minimum” (more precisely, the global minimum of the surrogate objective function) and the (or a) “surrogate minimizer,” respectively; $\Phi_{\min}(\underline{\lambda})$ and $\underline{p}^*(\underline{\lambda})$ of (2) will be referred to as the “actual minimum” (more properly, the minimum of the actual objective function, $\Phi(\underline{p}, \underline{\lambda})$) and “actual minimizer,” respectively.

We make several comments concerning the SBO Algorithm. First, the $\underline{\lambda}^m$ of Step 5 are selected based on currently available information, including, perhaps, the results of previous optimization studies (that is, for $\underline{\lambda}^n, n < m$). Second, we remark that the surrogate objective function is constructed only indirectly; that is, rather than directly construct the objective function output from \mathcal{A} applied to $(\underline{p}, \phi(\underline{\mathcal{S}}(\underline{p}); \underline{p}; \underline{\lambda}))$ pairs, we first construct surrogates, $\tilde{\mathcal{S}}(\underline{p})$, for the “intermediate physical outputs,” $\underline{\mathcal{S}}(\underline{p})$, in Step 4, and then simply *evaluate* $\phi(\tilde{\mathcal{S}}(\underline{p}); \underline{p}; \underline{\lambda})$ in Step 6 and Step 7. For example, for the eddy-promoter heat exchanger, we construct (7) not directly from the sampled data, but indirectly from the intermediate surrogates for flowrate and Nusselt number described in Section 3.2.3. We prefer this two-stage approach to direct construction of the objective function because: each new $\underline{\lambda}^m$ does not require re-appeal to a construction procedure; we are more likely to have prior information for the physical quantities than for an artificially

synthesized objective function. Third, we note that, despite the indirect construction-cum-evaluation of the surrogate objective function, we *directly* validate the surrogate objective function in Step 7, that is, we directly compute the errors in the objective function rather than (less precisely) infer these errors from the errors in $\tilde{\mathcal{Q}}(p)$. From Step 7 and Section 3.1.2 we know

$$\Pr\left\{\int_{\{p \in \Omega \mid |\Phi(p) - \tilde{\Phi}(p)| \leq E_{\max}^{\Phi}\}} \rho(p) dp \geq 1 - \varepsilon_1\right\} \geq 1 - \varepsilon_2, \quad (26)$$

or, equivalently,

$$\Pr\left\{\int_{\mathcal{U}^{\Phi}} \rho(p) dp \leq \varepsilon_1\right\} \geq 1 - \varepsilon_2, \quad (27)$$

where $\mathcal{U}^{\Phi} = \{p \in \Omega \mid |\Phi(p; \lambda) - \tilde{\Phi}(p; \lambda)| > E_{\max}^{\Phi}\}$. These (effectively single-output) objective function prediction error estimates are required for the *a posteriori* analysis described in Section 4.3.

We close this section by remarking that our algorithm is related to, but significantly different from, several other stochastic optimization procedures [14]. First, as compared to the simplest random search procedure, in which the minimum is approximated as the minimum of a random sample, our approach offers two advantages: by constructing and subsequently minimizing a surrogate, we can exploit whatever prior information may be available for, and whatever continuity may be present in, the objective function; the surrogate reveals internal error and sensitivity estimates not apparent from the bare “nodal values” used in the random search procedure. Second, as compared to multistart techniques, in which a random sample serves as the starting point for many parallel local (e.g., gradient-based) minimization problems: the multistart technique shares the disadvantages (but also advantages) of direct insertion as regards each local search; the probabilistic estimates for the multistart technique, although ostensibly similar to our validation statements, are in fact expressed in terms of the (unknown) size of the basin of attraction associated with the global minimum.

4.3 A Posteriori Estimates

We discuss in this subsection what can be said of a definitive (or almost definitive) nature following surrogate-based optimization. Our goal is to understand, for any given *single* realization, the influence of the selected importance function, $\rho(p)$, on the reliability of the surrogate-based optimization

process; our notion of success is thus very different than that adopted in [12], in which the quality of the surrogate is quantified only for the case of *repeated* trials according to the *initially* prescribed importance function. The *a posteriori* estimates of Step 8 serve in Step 9 first, to ascertain if the surrogate minimum is acceptable, and second, if the surrogate minimum is not acceptable, to guide subsequent adaptive improvement efforts. It is critical to note that, consistent with the notion of an expensive subproblem, all *a posteriori* estimates in Step 8 require no new appeals to $\underline{S}(p)$.

4.3.1 General Case

We state our result, present a brief formal derivation, and discuss the theoretical and practical implications. To begin, for any $r \in (1, 1/\varepsilon_1)$, we define χ_r to be the set of all closed domains, \mathcal{R} , in Ω for which $\int_{\mathcal{R}} \rho(p) dp = r\varepsilon_1$. We then set

$$\delta = \min_{\mathcal{R} \in \chi_r} \max_{p \in \mathcal{R}} \{ \tilde{\Phi}(p) - \tilde{\Phi}_{\min} \}, \quad (28)$$

$$\mathcal{R}_\delta = \arg \min_{\mathcal{R} \in \chi_r} [\max_{p \in \mathcal{R}} \{ \tilde{\Phi}(p) - \tilde{\Phi}_{\min} \}]. \quad (29)$$

In essence, the sensitivity region \mathcal{R}_δ is that (or a) region of relative weighted volume $r\varepsilon_1$ for which the deviation of the surrogate objective function from $\tilde{\Phi}_{\min}$ is minimal; this minimal deviation, δ , reflects the sensitivity of the surrogate objective function to variations in the design variables in the vicinity of the surrogate minimizer. We can now express a form of “lower semi-continuity:” with probability greater than $1 - \varepsilon_2$,

$$\exists p \in \mathcal{R}_\delta, \underline{p}_\diamond, \text{ such that } \Phi(\underline{p}_\diamond, \lambda) \leq \tilde{\Phi}_{\min} + \varpi, \quad (30)$$

where ϖ , the “predictability gap,” is the random variable

$$\varpi = E_{\max}^\Phi + \delta. \quad (31)$$

Note that (30) says nothing concerning the discrepancy between the surrogate minimum and actual minimum or the surrogate minimizer and actual minimizer; without further hypotheses on $\Phi(p; \lambda)$ and $\tilde{\Phi}(p; \lambda)$, no such statement can be made. Condition (30) does, however, say something concerning the *reliability* of the surrogate prediction: within a constructable region, \mathcal{R}_δ , there exists a point (in fact, many points) at which *actual* system performance

is within ϖ of the surrogate-predicted optimum. Our underlying strategy is to *strive* for a global minimum of $\Phi(\underline{p}; \underline{\lambda})$, but to *require* reliability in the surrogate prediction; our approach is thus, at least philosophically, related to the Taguchi approach to quality control [18,19].

To derive (30), we first note that, from (27), with probability greater than or equal to $1 - \varepsilon_2$, \mathcal{U}^Φ is of relative weighted volume less than or equal to ε_1 . If \mathcal{U}^Φ is of relative weighted volume less than or equal to ε_1 , then $\mathcal{R}_\delta \setminus \mathcal{U}^\Phi$ must be nonempty, since \mathcal{R}_δ is of relative weighted volume *strictly greater* than ε_1 (recall $r > 1$). Then, for any point \underline{p}' in $\mathcal{R}_\delta \setminus \mathcal{U}^\Phi$,

$$\begin{aligned} |\Phi(\underline{p}'; \underline{\lambda}) - \tilde{\Phi}(\tilde{\underline{p}}^*; \underline{\lambda})| &= |\Phi(\underline{p}'; \underline{\lambda}) - \tilde{\Phi}(\underline{p}'; \underline{\lambda}) + \tilde{\Phi}(\underline{p}'; \underline{\lambda}) - \tilde{\Phi}(\tilde{\underline{p}}^*; \underline{\lambda})| \\ &\leq |\Phi(\underline{p}'; \underline{\lambda}) - \tilde{\Phi}(\underline{p}'; \underline{\lambda})| + |\tilde{\Phi}(\underline{p}'; \underline{\lambda}) - \tilde{\Phi}(\tilde{\underline{p}}^*; \underline{\lambda})|. \end{aligned}$$

As $|\Phi(\underline{p}'; \underline{\lambda}) - \tilde{\Phi}(\underline{p}'; \underline{\lambda})| \leq E_{\max}^\Phi$ (recall $\underline{p}' \notin \mathcal{U}^\Phi$), and $|\tilde{\Phi}(\underline{p}'; \underline{\lambda}) - \tilde{\Phi}(\tilde{\underline{p}}^*; \underline{\lambda})| \leq \delta$, (30–31) directly follows, with $\underline{p}_\circ = \underline{p}'$. From this derivation it is clear that (30) applies to each $\underline{\lambda}^m$ of Step 5 separately, not jointly; however, the SBO Algorithm is readily extended such that (30) is jointly valid (see Section 5).

In order to illustrate (30), we consider a simple model problem with $M = 2$, $\underline{p} \in \Omega = [-1, 1] \times [-1, 1]$, and $\rho(\underline{p}) = 1/4$. We presume that the result of the SBO Algorithm is a surrogate objective function,

$$\tilde{\Phi}(\underline{p}) = p_1^2 + \omega^2 p_2^2 + 1 \quad (\omega \geq 1), \quad (32)$$

with minimum and minimizer $\tilde{\Phi}_{\min} = 1$ and $\tilde{\underline{p}}^* = \{0, 0\}$, respectively, and prediction error estimate $\Re E_{\max}^\Phi$. It is then readily computed that $\delta = 4r\varepsilon_1\omega/\pi$, with \mathcal{R}_δ given by the elliptical region centered at the origin with major (p_1) and minor (p_2) axes of $\sqrt{\delta}$ and $\sqrt{\delta}/\omega$, respectively. The predictability gap is thus $\Re\varpi = \Re E_{\max}^\Phi + 4r\varepsilon_1\omega/\pi$. It is clear, since we have not even defined the actual objective function, $\Phi(\underline{p}; \underline{\lambda})$, that this analysis is based entirely upon appeals to the *surrogate*. This model problem is not, of course, completely arbitrary; the M -dimensional generalization of the objective function (32) is a *local* representation of any sufficiently differentiable objective function near an interior minimizer.

Turning now to the implications of (30), we remark, first, that (30) quantifies the effect of a poorly selected $\rho(\underline{p})$: we expect that δ , $|\mathcal{R}_\delta|$ and ϖ will be inversely proportional to $\rho(\tilde{\underline{p}}^*(\underline{\lambda}))$. Second, we understand the origin of the predictability gap, ϖ , as distinguishable construction, E_{\max}^Φ , and validation, δ , contributions. We expect that, as N^{co} and $N^{va} \rightarrow \infty$ (ε_2 fixed),

E_{\max} , δ and hence ϖ will tend to zero; furthermore, for any particular study at finite N^{ev} , the ratio $\zeta = \Re E_{\max}^{\Phi} / \delta$ provides valuable “construction *versus* validation” guidance for adaptive improvement (see Section 4.4). Third, the continuity statement (30) includes a notion of “ \underline{p} -sensitivity,” which we define as the sensitivity of $\Phi(\underline{p}; \underline{\lambda})$ to variations in \underline{p} . (We contrast \underline{p} -sensitivity to “ $\underline{\lambda}$ -sensitivity,” which we define as the sensitivity of design points, such as $\Phi_{\min}(\underline{\lambda})$ and $\underline{p}^*(\underline{\lambda})$, to variations in $\underline{\lambda}$.) Our “ \underline{p} -sensitivity” result should prove quite useful in preliminary optimization studies: if ϖ is acceptably small, and \mathcal{R}_{δ} is acceptably located, the precise design point need not be specified, thereby maintaining maximal flexibility in the ensuing design process. This flexibility is particularly important when the optimization problem, (2), reflects only one subsystem of a larger, more complex endeavor [49].

Remark: Random Search Revisited. It is readily shown that, with probability greater than $1 - \varepsilon_2^r$ (as $\varepsilon_1 \rightarrow 0$), a validation input point, \underline{P}_j , resides in \mathcal{R}_{δ} (and hence $\mathcal{R}_{\delta} \setminus \mathcal{U}^{\Phi}$). Thus, even in the worst case, in which we simply set $\underline{p}_{\mathcal{O}}$ of (30) to $\Re \underline{P}_j$, the surrogate approach reproduces the simple random-search result, and, additionally, provides: valuable \underline{p} -sensitivity information through condition (30); convergence guidance through the quasi-convex analysis of the next section. \square

4.3.2 Quasi-Convex Case

We consider here the case in which Ω is convex, and $\Phi(\underline{p}; \underline{\lambda})$ and $\tilde{\Phi}(\underline{p}; \underline{\lambda})$ are quasi-convex in the first argument. (A function $f(\underline{p}): \Omega \rightarrow \mathbb{R}$ is quasi-convex if: $\forall \alpha, \underline{p}_1, \underline{p}_2 \in ([0, 1], \Omega, \Omega)$, $f(\alpha \underline{p}_1 + (1 - \alpha) \underline{p}_2) \leq \max[f(\underline{p}_1), f(\underline{p}_2)]$; or, equivalently, the level sets, $\{\underline{p} \in \Omega \mid f(\underline{p}) \leq \varphi\}$, are convex [25].) We first state our main result: given a “separating value” random variable $\Delta = 2E_{\max}^{\Phi} + \delta$ and an associated random “buffer zone,”

$$\mathcal{B}_{\Delta} = \{\underline{p} \in \Omega \mid \tilde{\Phi} - \tilde{\Phi}_{\min} > \Delta\},$$

a random “containing region,” \mathcal{K} , can be constructed in which, with probability greater than $1 - \varepsilon_2$, an actual minimizer, \underline{p}^* , must reside; furthermore, as E_{\max}^{Φ} and ε_1 tend to zero, the region \mathcal{K} shrinks to $\tilde{\underline{p}}^*$, the surrogate minimizer. The \mathcal{K} -construction depends only on $\rho(\underline{p}), r$, and the geometric properties of \mathcal{R}_{δ} and \mathcal{B}_{Δ} . In this section we: formally derive the \mathcal{K} -construction for a general one-dimensional ($M = 1$) optimization problem;

state the \mathcal{K} -construction for a particular two-dimensional ($M = 2$) model optimization problem; and discuss the implications. The derivation of the general $M = 2$ \mathcal{K} -construction is given in [32]; development of the general $M > 2$ \mathcal{K} -construction, though tractable, is rather involved and not yet complete. All results presented are for the case of uniform $\rho(p)$.

We now proceed with the \mathcal{K} -construction for the general one-dimensional ($M = 1$) quasi-convex optimization problem. We introduce the (perforce convex) regions $\Omega = [p_{\Omega}^-, p_{\Omega}^+]$, $\mathcal{R}_{\delta} = [p_{\delta}^-, p_{\delta}^+]$, $\mathcal{B}_{\Delta} = [p_{\Omega}^-, P_{\Delta}^-] \cup (P_{\Delta}^+, p_{\Omega}^+]$, and $\mathcal{K} = [P_{\Delta}^- - \varepsilon_1, P_{\Delta}^+ + \varepsilon_1]$ shown in Figure 7 (note P_{Δ}^{\pm} are random variables); for clarity of exposition, we assume that $p_{\delta}^- > P_{\Delta}^-$, $p_{\delta}^+ < P_{\Delta}^+$, and $P_{\Delta}^- > p_{\Omega}^- + \varepsilon_1$, $P_{\Delta}^+ < p_{\Omega}^+ - \varepsilon_1$. First, from (27) we know that, with probability greater than or equal to $1 - \varepsilon_2$, \mathcal{U}^{Φ} is of total length less than or equal to ε_1 ; note that \mathcal{U}^{Φ} need *not* be convex. If \mathcal{U}^{Φ} is of total length less than or equal to ε_1 , then: there exists a point p' in $\mathcal{R}_{\delta} \setminus \mathcal{U}^{\Phi}$; for *any* point p''' in $\Omega \setminus \mathcal{K}$, there exists a point p'' in $\mathcal{B}_{\Delta} \cap \mathcal{K} \setminus \mathcal{U}^{\Phi}$ for which $p''' < p'' < p'$ or $p' < p'' < p'''$. For example, to prove the latter, take (say) $p''' > P_{\Delta}^+ + \varepsilon_1$, and assume that no point p'' in $\mathcal{B}_{\Delta} \cap \mathcal{K} \setminus \mathcal{U}^{\Phi}$ exists such that $p' < p'' < p'''$; but then, $(P_{\Delta}^+, p''') \subset \mathcal{U}^{\Phi}$, and since $p''' > P_{\Delta}^+ + \varepsilon_1$, we arrive at a contradiction. We next claim that, for p' in $\mathcal{R}_{\delta} \setminus \mathcal{U}^{\Phi}$ and p'' in $\mathcal{B}_{\Delta} \cap \mathcal{K} \setminus \mathcal{U}^{\Phi}$,

$$\Phi(p''; \lambda) > \tilde{\Phi}_{\min} + E_{\max}^{\Phi} + \delta \quad (33)$$

and

$$\Phi(p''; \lambda) > \Phi(p'; \lambda). \quad (34)$$

Inequality (33) follows from

$$\begin{aligned} |\tilde{\Phi}(\tilde{p}''; \lambda) - \Phi(p''; \lambda)| &= |\tilde{\Phi}(\tilde{p}''; \lambda) - \tilde{\Phi}(p''; \lambda) + \tilde{\Phi}(p''; \lambda) - \Phi(p''; \lambda)| \\ &\geq |\tilde{\Phi}(\tilde{p}''; \lambda) - \tilde{\Phi}(p''; \lambda)| - |\tilde{\Phi}(p''; \lambda) - \Phi(p''; \lambda)|, \end{aligned}$$

and $|\tilde{\Phi}(\tilde{p}''; \lambda) - \tilde{\Phi}(p''; \lambda)| > 2E_{\max}^{\Phi} + \delta$, $|\tilde{\Phi}(p''; \lambda) - \Phi(p''; \lambda)| \leq E_{\max}^{\Phi}$ (recall $p'' \notin \mathcal{U}^{\Phi}$). Inequality (34) follows from (33) and the results of Section 4.3.1. Lastly, the event (say) $p' < p'' < p'''$ (p' in $\mathcal{R}_{\delta} \setminus \mathcal{U}^{\Phi}$, p'' in $\mathcal{B}_{\Delta} \cap \mathcal{K} \setminus \mathcal{U}^{\Phi}$, p''' in $\Omega \setminus \mathcal{K}$) implies, from the inequality (34) and quasi-convexity, $\Phi(p''; \lambda) \leq \max[\Phi(p'; \lambda), \Phi(p'''; \lambda)]$, that $\Phi(p''; \lambda)$ can not be greater than $\Phi(p'''; \lambda)$; thus, there exists a minimizer of $\Phi(p; \lambda)$ within \mathcal{K} . This proves the desired result, and constructs the requisite region, \mathcal{K} .

We next pass to the two-dimensional case, and present the \mathcal{K} -construction for the particular model problem discussed in Section 4.3.1: $M = 2$; $p \in \Omega =$

$[-1, 1] \times [-1, 1]$; $\rho(p) = 1/4$; surrogate objective function (32); and prediction error estimate, $\Re E_{\max}^{\Phi}$. As described in Section 4.3.1, $\delta = 4r\varepsilon_1\omega/\pi$, with \mathcal{R}_δ given by an elliptical region with (major(p_1),minor(p_2)) axes $\sqrt{\delta}(1, 1/\omega)$. Following an analysis conceptually similar to — though technically more complicated than — our analysis for the one-dimensional optimization problem, we find that, as $\delta \rightarrow 0$, $\Re \mathcal{B}_\Delta$ is the exterior of an ellipse-like (though not exactly elliptical) region of (major(p_1),minor(p_2)) axes $\sqrt{\delta}\sqrt{1+2\zeta}(1, 1/\omega)$, and $\Re \mathcal{K}$ is an ellipse-like region of (major(p_1),minor(p_2)) axes

$$\sqrt{\delta}\{\sqrt{1+2\zeta} + 2/r[1 + \sqrt{1+r\sqrt{1+2\zeta}}]\}(1, 1/\omega),$$

where $\zeta = \Re E_{\max}^{\Phi}/\delta$. Note that, as $\Re E_{\max}^{\Phi}$ and $\delta \rightarrow 0$, $\Re \mathcal{K}$ shrinks to \tilde{p}^* .

The implications of our results are clear. First, from the theoretical perspective, we obtain convergence of the surrogate minimizer (and probably, with convexity, the surrogate minimum) to the actual minimizer (actual minimum) as $N^{ev} \rightarrow \infty$. Second, from the practical perspective, we provide a natural framework in which to pursue search-domain reduction strategies [50,51]: an initial surrogate-based optimization study over Ω provides the design space, $\Re \mathcal{K}$, for subsequent adaptive improvement; with high confidence, $\Re \mathcal{K}$ contains the requisite global minimizer. This search-reduction approach is, in practice, hampered by the pessimistically large regions \mathcal{K} that result from the rather minimal assumptions placed on $\Phi(p; \lambda)$ and $\tilde{\Phi}(p; \lambda)$.

4.4 Adaptive Improvement

IF (i) on the basis of the surrogate minimum $\tilde{\Phi}_{\min}(\lambda)$, the surrogate minimizer, $\tilde{p}^*(\lambda)$, the surrogate prediction error estimate, $\Re E_{\max}^{\Phi}$, the predictability gap, $\Re \omega$, the sensitivity region, \mathcal{R}_δ , and the containing region, $\Re \mathcal{K}$, the surrogate minimum and minimizer of Step 6 are deemed *unacceptable*, AND IF (ii) further appeals to $\mathcal{S}(p)$ are permitted (e.g., N^{ev} reflects only part of the total resource allocation for the entire project, or additional resources can be renegotiated given the optimization results to date), THEN a *posteriori* estimate-based adaptive improvement can be pursued. (Note that if (i) is false, then we simply declare success and return to Step 5 of the SBO algorithm; however, if (i) is true but (ii) is false, we must admit defeat. In the latter case, the *a posteriori* estimates serve the unpopular but valuable function of *qualifying* the surrogate minimum and minimizer.)

Two adaptive improvement branches (from the local bindings associated with the parent SBO procedure) can be envisioned. In the first branch, we simply evoke one additional appeal to $\underline{\mathcal{S}}(\underline{p})$ to determine actual system performance at $\underline{p} = \tilde{\underline{p}}^*$, $\phi(\underline{\mathcal{S}}(\tilde{\underline{p}}^*); \tilde{\underline{p}}^*; \underline{\lambda})$; all existing information implicates $\tilde{\underline{p}}^*$ as a viable design point, and this possibility therefore merits investigation before proceeding further. In the second branch, we recursively evoke a second instantiation of the SBO Algorithm, in which we re-appeal to the databoard (see Section 5) or expensive simulation $\underline{\mathcal{S}}(\underline{p})$ in order to: if $\zeta = \Re E_{\max}^{\Phi} / \delta \geq 1$, devote additional input-output pairs to the validation of new models (see Section 5) or to the further construction of existing models; if $\zeta = \Re E_{\max}^{\Phi} / \delta < 1$, devote additional input-output pairs to the validation of existing models. Re-appeal to the databoard or $\underline{\mathcal{S}}(\underline{p})$ will typically be accompanied by search-domain modification, in which we, say, focus $\rho(\underline{p})$ in, or relocate Ω to, the $\Re\mathcal{K}$ -neighborhood of the current surrogate minimizer.

Adaptive, or multipass, strategies, in which optimization information feeds back to the surrogate hypotheses, are clearly a necessity. As in experimental data collection procedures [7], all diagnostic (here simulation) resources should not be expended in the first salvo; unfortunately, as for experimental inquiries, we can hope to proffer only “rules of thumb” as to when to commit resources. More explicit strategies for, and examples of, multipass interaction [49] will be addressed in future papers.

4.5 Eddy-Promoter Heat Exchanger

We consider now the eddy-promoter heat exchanger example. The objective function is given by (8); we choose $\rho(\underline{p})$ to be uniform over Ω^{EP} ; we set $N^{\text{ev}} = 44$ and $\varepsilon_1 = \varepsilon_2 = .1$. The flowrate and Nusselt number intermediate physical surrogates are constructed as in Section 3.2.3 based on the construction sample shown in Figure 5. For the definition vector we take $\underline{\Delta}^{\text{EP}} = \underline{\Delta}^{\text{EP},1} = \{\beta_1 = 10^{-7}, \beta_2 = .20, \beta_3 = 1.1, \kappa = 2.0\}$. Proceeding to Step 6 of the SBO Algorithm, we find (for this two-dimensional case by a simple search) $\tilde{\Phi}_{\min}^{\text{EP}}(\underline{\Delta}^{\text{EP},1}) = .223$, $\tilde{\underline{p}}^{\text{EP}*}(\underline{\Delta}^{\text{EP},1}) = (.40, .14)$ for the surrogate and surrogate minimizer, respectively; a contour plot of the surrogate objective function is shown in Figure 8. From Step 7 of the SBO Algorithm we find, for the particular validation sample shown in Figure 5, $\Re E_{\max}^{\Phi} = .0605$. Then, from the *a posteriori* analysis of Step 8, we calculate (for $r = 2$) $\delta = .029$, and thus $\Re\omega = .090$ and $\zeta = 2.14$; we show in Figure 8 the region \mathcal{R}_{δ} in which,

from (30), with confidence level greater than $1 - \varepsilon_2$, we can find a $\underline{p}^{\text{EP}}, \underline{p}_{\diamond}^{\text{EP}}$, such that $\Phi^{\text{EP}}(\underline{p}_{\diamond}^{\text{EP}}, \underline{\lambda}^{\text{EP},1})$ (actual system performance) $\leq \tilde{\Phi}_{\min}^{\text{EP}}(\underline{\lambda}^{\text{EP},1}) + .090$. In Step 9, we re-appeal to the Navier-Stokes subproblem for $\underline{p}^{\text{EP}} = \tilde{\underline{p}}^{\text{EP}*}(\underline{\lambda}^{\text{EP},1})$ to obtain $\Phi^{\text{EP}}(\tilde{\underline{p}}^{\text{EP}*}(\underline{\lambda}^{\text{EP},1}); \underline{\lambda}^{\text{EP},1}) = .253$, quite close to the value predicted by the surrogate objective function.

We now return to Step 5, and reselect $\underline{\lambda}^{\text{EP},2} = \{\beta_1 = 10^{-6}, \beta_2 = .01, \beta_3 = 1.1, \kappa = 2.0\}$, because (say) the optimum predicted in Figure 8 corresponds to a pump size which is unexpectedly large. Exploiting the same samples and intermediate physical surrogates as for the first study (recall the resulting estimates are not *joint* pending the revised multiple-output sample-size requirement of Section 5), we find, in Step 6 and Step 7, $\tilde{\Phi}_{\min}^{\text{EP}}(\underline{\lambda}^{\text{EP},2}) = .190$, $\tilde{\underline{p}}^{\text{EP}*}(\underline{\lambda}^{\text{EP},2}) = (.80, .74)$, and $\mathcal{RE}_{\max}^{\Phi} = .0611$ for the surrogate, surrogate minimizer, and prediction error estimate, respectively; a contour plot of the surrogate objective function is shown in Figure 9. Pursuing in Step 8 our *a posteriori* analysis, we calculate (for $r = 2$) $\delta = .022$, and thus $\mathcal{R}\omega = .083$ and $\zeta = 2.77$; we show in Figure 9 the region \mathcal{R}_{δ} in which, from (30), with confidence level greater than $1 - \varepsilon_2$, we can find a $\underline{p}^{\text{EP}}, \underline{p}_{\diamond}^{\text{EP}}$, such that $\Phi^{\text{EP}}(\underline{p}_{\diamond}^{\text{EP}}, \underline{\lambda}^{\text{EP},2})$ (actual system performance) $\leq \tilde{\Phi}_{\min}^{\text{EP}}(\underline{\lambda}^{\text{EP},2}) + .083$. Note that, even if we falsely presume global quasi-convexity, the region \mathcal{RK} in which the actual minimizer must lie is, disappointingly, essentially Ω^{EP} . In Step 9 we re-appeal to the Navier-Stokes subproblem for $\underline{p}^{\text{EP}} = \tilde{\underline{p}}^{\text{EP}*}(\underline{\lambda}^{\text{EP},2})$ to obtain $\Phi^{\text{EP}}(\tilde{\underline{p}}^{\text{EP}*}(\underline{\lambda}^{\text{EP},2}); \underline{\lambda}^{\text{EP},2}) = .194$, again quite close to the value predicted by the surrogate objective function. It is not surprising that, with the increased (decreased) penalty on pumping power (materials cost), optimal performance now occurs at a lower flowrate in the vicinity of the global maximum in heat transfer.

5 Extensions

Classification Procedures. In classification problems we search for a region of input space in which certain conditions are satisfied; for example, in the eddy promoter problem, we might be interested in that region of Ω^{EP} in which the heat transfer rate, $Q(a, R)$, is greater than a prescribed threshold. In such situations, it is clearly advantageous to replace (say) the Navier-Stokes equations with a less expensive — surrogate — $\{0, 1\}$ char-

acteristic function. For noisy classification problems, statistical prediction rules [9] or neural network approaches [11] prove effective; for deterministic simulation-based classification problems, we develop in [2] a Boolean Voronoi construction method and a binomial-tail-statistic [16] validation technique. Modelling and classification algorithms share much in common as regards both motivation and formulation, and, in the future, must be combined in a single procedure, in which a model-based surrogate objective function is minimized over a classification-based feasible domain.

The Databoard. The databoard concept, developed in [2], permits investigations defined over different input domains to share data. The technique, based on simple conditional sampling procedures, is best illustrated by an example. Consider the second eddy-promoter heat exchanger optimization problem considered in Section 4.4, in which the minimum is near the $\{a = 1.0, R = .95\}$ vertex of Ω^{EP} . Assume, however, that the original design space is defined to be not the triangular Ω^{EP} , but, rather, the square domain $\Omega_{i(\text{initial})}^{\text{EP}}$ shown in Figure 10. Upon minimization, the minimizer will no doubt reside on the boundary of Ω_i^{EP} , perhaps prompting the investigator to expand the design space to (say) the full triangle, Ω^{EP} . In performing the subsequent adaptive improvement over Ω^{EP} , the researcher can: recycle existing data from Ω_i^{EP} ; evoke new simulations only when Ω_i^{EP} is depleted, or when the requested input point lies in $\Omega^{\text{EP}} \setminus \Omega_i^{\text{EP}}$; post new simulations to the databoard for the benefit of future investigations. The concept is readily expanded to permit rather general input domains and both modelling and classification studies; furthermore, if raw, rather than processed, simulation data is posted to the databoard, significant output flexibility can be achieved.

Multiple-Output Validation. We discuss here the generalization of the Model Validation Algorithm (and, by obvious extension, the MCV and SBO Algorithms) to the case of multiple outputs, $K > 1$. In particular, we consider the situation in which we wish to validate $K > 1$ outputs (e.g., for the eddy-promoter problem, the flowrate and the Nusselt number) at the *same* sample input points, $\underline{P}_1, \dots, \underline{P}_{N^{\text{va}}}$,

$$E_{\max}^k = \max_{j \in \{1, \dots, N^{\text{va}}\}} E_j^k, \quad k = 1, \dots, K, \quad (35)$$

$$E_j^k = |\mathcal{S}^k(\underline{P}_j) - \tilde{\mathcal{S}}^k(\underline{P}_j)|, \quad j = 1, \dots, N^{\text{va}}, \quad (36)$$

where $\mathcal{S}^k(\mathbf{p})$ and $\tilde{\mathcal{S}}^k(\mathbf{p})$ refer to the k^{th} component of the $\underline{\mathcal{S}}(\mathbf{p})$ and $\tilde{\underline{\mathcal{S}}}(\mathbf{p})$ vectors, respectively. It is readily shown [32] by multinomial extension of the binomial arguments described for the single-output case that, if we simply replace the sample-size requirement of Step 1 of the MV Algorithm with

$$N^{va} = \frac{\ln(\varepsilon_2/K)}{\ln(1 - \varepsilon_1)}, \quad (37)$$

then the K outputs *jointly* satisfy a validation statement,

$$\Pr\left\{\int_{\{\mathbf{p} \in \Omega \mid |\mathcal{S}^k(\mathbf{p}) - \tilde{\mathcal{S}}^k(\mathbf{p})| \leq E_{\max}^k\}} \rho(\mathbf{p}) d\mathbf{p} \geq 1 - \varepsilon_1, k = 1, \dots, K\right\} \geq 1 - \varepsilon_2. \quad (38)$$

Note to arrive at the simple expression (37), certain quantities in the multinomial expansion are bounded. However, the requirement (37) is often pessimistic in practice not because of the bounds, which are rather sharp for $K < 1/\varepsilon_1$, but because most actual output vectors are better correlated than the worst-case assumption inherent in (37). The remarkable conclusion from (37) is that only logarithmically more simulations must be performed in order to jointly validate multiple outputs over a common input set; indeed, if (say) $\varepsilon_2 = .01$, $K = 100$ outputs require only twice the sample size as a single output ($K = 1$). This logarithmic dependence further justifies the surrogate concept; if, to obtain joint estimates, the sample size grew linearly with K , the surrogate approach would be not too different from direct insertion procedures as regards adaptability.

We mention three applications of the multiple-output theory. First, multicriteria optimization frameworks can now be addressed. Second, multiple-optimization studies ($\underline{\lambda}^1, \underline{\lambda}^2 \dots$) can be jointly validated so that confidence can be assured not only in the final study, but in all earlier studies on which the final study is conditioned: by replacing Step 1 of the SBO Algorithm with (37), the $\underline{\lambda}$ variation in Step 5 is now jointly justified. Third, if we interpret the K outputs as the K errors associated with a single physical output approximated by K different models, (37) permits efficient model-optimal construction procedures: we replace Step 1 of the MCV Algorithm with (37); we test several candidate models according to (35); we choose the best model based on accuracy or cost criteria [5,6]; we are assured, from (38), that our rank ordering is significant, and that our validation statement applies to the particular model selected. Sequential procedures can also be pursued.

6 Conclusions

Surrogate techniques constitute an attractive alternative to “direct insertion” for the incorporation of large-scale simulations into engineering optimization studies. Simulation surrogates provide direct resource control, support flexibility in design objectives and specifications, and gainfully accommodate prior information. Furthermore, the particular construction-validation procedures proposed here enjoy *a posteriori* error estimates that permit both qualification of surrogate results and guidance for subsequent adaptive improvement. Much additional work is required, however, if the simulation surrogate framework is to prove useful in engineering design. In particular, multipass adaptive refinement strategies must be articulated for both single- and multiple-optimization-study projects, with emphasis on construction-validation refinement, search domain reduction and relocation, and graduated deployment of resources.

7 Acknowledgements

We would like to thank Nektonics, Inc., and Fluent, Inc., for making the NEKTON code available for this work; we would like to thank Mr. Mariano Gurfinkel of M.I.T., Drs. Einar Rønquist and Lee-Wing Ho of Nektonics, Inc., and Professor Paul Fischer of Brown University for their help in integrating the parallel NEKTON code into the surrogate framework. We would also like to acknowledge fruitful discussions with Drs. Jaroslaw Sobieski and Thomas Zang of NASA Langley Research Center. This work was supported by DARPA Grant N00014-91-J-1889, ONR Grant N00014-90-J-4124, ONR Grant N00014-89-J-1610, and NASA Contract NAS1-19480.

References

- [1] J. Sobieszczanski-Sobieski. Sensitivity of complex, internally coupled systems. *AIAA Journal*, 28(1):153-160, 1990.
- [2] S. Yesilyurt and A. T. Patera. Statistical modelling methods for deterministic computational systems. *J. Comput. Phys.*, 1992. Submitted.

- [3] L. Ljung. *System identification: Theory for the user*. Prentice-Hall, New Jersey, 1987.
- [4] K. H. Reckhow. Validation of simulation models: Philosophy and statistical methods of confirmation. In M. G. Singh, editor, *Systems and Control Encyclopedia*, volume 7, pages 5011–5015. Pergamon Press, New York, 1987.
- [5] C. Heij. *Deterministic identification of dynamical systems*. Number 127 in Lecture Notes in Control and Information Sciences. Springer-Verlag, Berlin, 1989.
- [6] T. Bohlin. *Interactive system identification: Prospects and pitfalls*. Springer-Verlag, Berlin, 1991.
- [7] G. E. P. Box, W. G. Hunter, and J. S. Hunter. *Statistics for experimenters*. John Wiley & Sons, New York, 1978.
- [8] M. Stone. Cross-validatory choice and assessment of statistical predictions. *J. Roy. Stat. Soc. Ser. B*, 36:111–147, 1974.
- [9] B. Efron. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J. Amer. Stat. Assoc.*, 78(382):316–331, 1983.
- [10] S. M. Weiss and C. A. Kulikowski. *Computer systems that learn*. Morgan Kaufmann, San Mateo, 1991.
- [11] J. A. Leonard, M. A. Kramer, and L. H. Ungar. A neural network architecture that computes its own reliability. *Computers and Chemical Engineering*, 1991. Submitted.
- [12] L. G. Valiant. A theory for the learnable. *CACM*, 27(11):1134–1142, 1984.
- [13] S. I. Gallant. A connectionist learning algorithm with provable generalization and scaling bounds. *Neural Networks*, 3:191–201, 1990.
- [14] R. Y. Rubinstein. *Simulation and the Monte-Carlo method*. John Wiley & Sons, New York, 1981.

- [15] M. Adams and V. Guillemin. *Measure theory and probability*. Wadsworth & Brooks, Monterey, California, 1986.
- [16] J. W. Pratt and J. D. Gibbons. *Concepts of nonparametric theory*. Springer-Verlag, New York, 1981.
- [17] R. Franke and G. M. Nielson. Scattered data interpolation and applications: A tutorial and study. In H. Hagen and D. Roller, editors, *Geometric Modelling: Methods and Applications*, pages 131–160. Springer-Verlag, Berlin, 1990.
- [18] G. Taguchi. *Introduction to quality engineering*. Asian Productivity Organization, Tokyo, 1986.
- [19] V. N. Nair, editor. Taguchi's parameter design: A panel discussion. *Technometrics*, 34(2):127–161, 1992.
- [20] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.
- [21] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435, 1989.
- [22] J. Sacks, S. B. Schiller, and W. J. Welch. Design for computer experiments. *Technometrics*, 31(1):41–47, 1989.
- [23] D. D. Cox and S. John. A statistical method for global optimization. Technical Report 67, Department of Statistics, University of Illinois, Champagne-Urbana, 1992.
- [24] R. W. Lewis and Y. Zheng. Coarse optimization for complex systems: An application of orthogonal arrays. *Comp. Meth. Appl. Mech. Eng.*, 94:63–92, 1992.
- [25] M. Avriel. *Nonlinear programming: Analysis and methods*. Prentice-Hall, Englewood Cliffs, New Jersey, 1976.
- [26] P. F. Fischer and A. T. Patera. Parallel simulation of viscous incompressible flows. *Annual Review of Fluid Mechanics*, 26, 1994.

- [27] Y. Maday, A. T. Patera, and E. M. Rønquist. An operator-integration-factor splitting method for time dependent problems: Application to incompressible fluid flow. *J. Sci. Comput.*, 5(4):263–292, 1990.
- [28] E. Rønquist. A domain decomposition method for elliptic boundary value problems: Application to unsteady incompressible fluid flow. In D. E. Keyes, T. F. Chan, G. A. Meurant, J. S. Scroggs, and R. G. Voigt, editors, *Proc. Int. Conf. on Domain Decomposition Methods for Partial Differential Equations, 5th*, pages 545–557. SIAM, Philadelphia, 1992.
- [29] Y. Maday and Patera A. T. Spectral element methods for the Navier-Stokes equations. In A. K. Noor, editor, *State-of-the-art surveys in computational mechanics*. ASME, New York, 1989.
- [30] P. F. Fischer and A. T. Patera. Parallel spectral element solution of the Stokes problem. *J. Comput. Phys.*, 92(2):380–421, 1991.
- [31] R. A. Nicolaides. Deflation of conjugate gradients with applications to boundary value problems. *SIAM J. Num. Anal.*, 51(2):355–365, 1987.
- [32] S. Yesilyurt. PhD Thesis. Department of Nuclear Engineering, M.I.T.. In progress.
- [33] O. Pironneau, W. Rodi, I. L. Ryhming, A. M. Savill, and T. V. Truong, editors. *Numerical Simulation of Unsteady Flows and Transition to Turbulence*. ERCOFTAC, Cambridge University Press, Cambridge, 1992.
- [34] M. S. Isaacson and A. A. Sonin. Sherwood number and friction factor correlations for electro dialysis systems, with application to process optimization. *I & EC Process Des. Dev.*, 15:313, 1976.
- [35] G. E. Karniadakis, B. B. Mikic, and A. T. Patera. Minimum-dissipation transport enhancement by flow destabilization: Reynolds' analogy revisited. *J. Fluid Mech.*, 192:365–391, 1988.
- [36] H. Kozlu, B. B. Mikic, and A. T. Patera. Minimum dissipation heat removal by scale-matched flow destabilization. *Int. J. Heat Mass Transfer*, 31(10):981–991, 1988.

- [37] M. F. Schatz, R. P. Tagg, H. L. Swinney, P. F. Fischer, and A. T. Patera. Supercritical transition in plane channel flow with spatially periodic perturbations. *Phys. Rev. Lett.*, 66(12):1579–1582, 1991.
- [38] E. P. L. Roberts. Numerical and experimental study of transition processes in obstructed channel flow. *J. Fluid Mech.*, 1993. To appear.
- [39] J. F. M. Barthelemy and J. Sobieszczanski-Sobieski. Optimum sensitivity derivatives of objective functions in nonlinear programming. *AIAA Journal*, 21(5), 1983.
- [40] J. A. Bennet and M. E. Botkin, editors. *The optimum shape: Automated structural design*. General Motors, Plenum Press, New York, 1986.
- [41] H. A. David. *Order statistics*. John Wiley & Sons, New York, 1970.
- [42] A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the theory of statistics*. McGraw-Hill, New York, 1974.
- [43] CRAY Research, Inc., Mendota Heights, Minnesota. *CF77 compiling systems, volume 1: Fortran reference manual*, 5th edition, 1991. SR-3071, 5.0.
- [44] A. Okabe, B. Boots, and K. Sugihara. *Spatial tessellations*. John Wiley & Sons, New York, 1992.
- [45] R. L. Renka. Algorithm 660: QSHEP2D: Quadratic Shepard method for bivariate interpolation of scattered data. *ACM TOMS*, 14:149–150, 1988.
- [46] P. A. Newman, G. J.-W. Hou, H. E. Jones, A. C. Taylor III, and V. M. Korivi. Observations on computational methodologies for use in large-scale, gradient-based, multidisciplinary design (AIAA Paper 92-4753). In *Symposium on Multidisciplinary Analysis and Optimization*, 1992.
- [47] C. Bischof, G. Corliss, L. Green, A. Griewank, K. Haigler, and P. Newman. Automatic differentiation of advanced CFD codes for multidisciplinary design. *Journal on Computing Systems in Engineering*, 1992. Submitted.

- [48] J. L. Rogers and W. J. LaMarsh II. Application of a neural network to simulate analysis in an optimization process. *Artificial Intelligence in Design*, 1992.
- [49] J. Sobieszczanski-Sobieski. Aircraft optimization by a system approach: Achievements and trends. Technical Memorandum 107622, NASA, 1992.
- [50] M. F. Berger and H. F. Silverman. Microphone array optimization by stochastic region contraction. *IEEE Trans. Signal Proc.*, 39(11):2377-2386, 1991.
- [51] R. Spaans and R. Luus. Importance of search-domain reduction in random optimization. *J. Opt. Theor. Appl.*, 75(3):635-638, 1992.

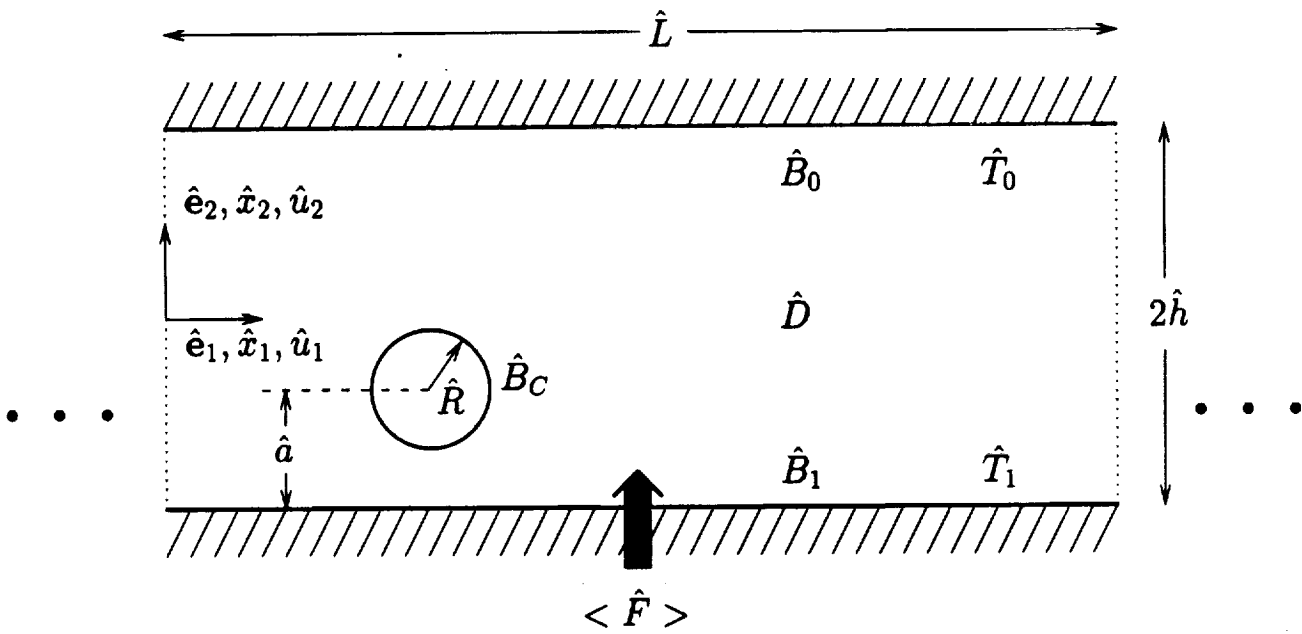


Figure 1: One periodicity cell of the eddy-promoter heat exchanger.

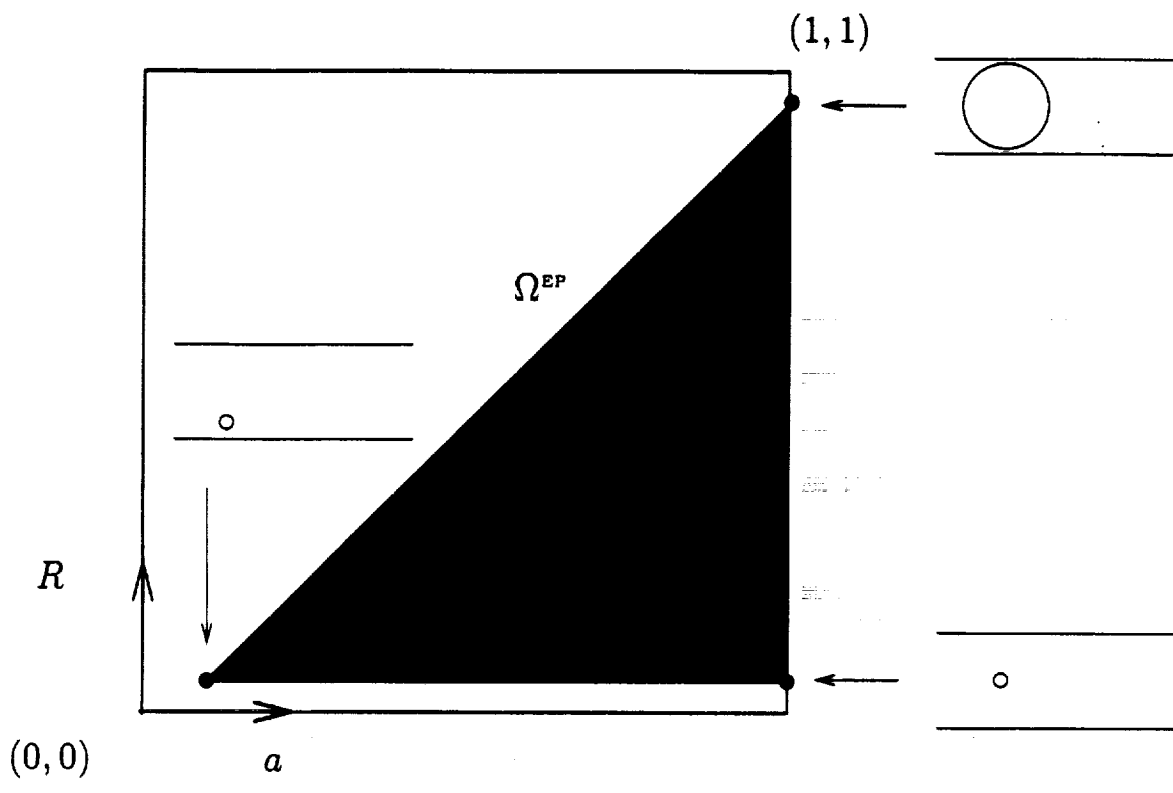


Figure 2: The eddy-promoter heat exchanger placement-radius design space, Ω^{EP} .

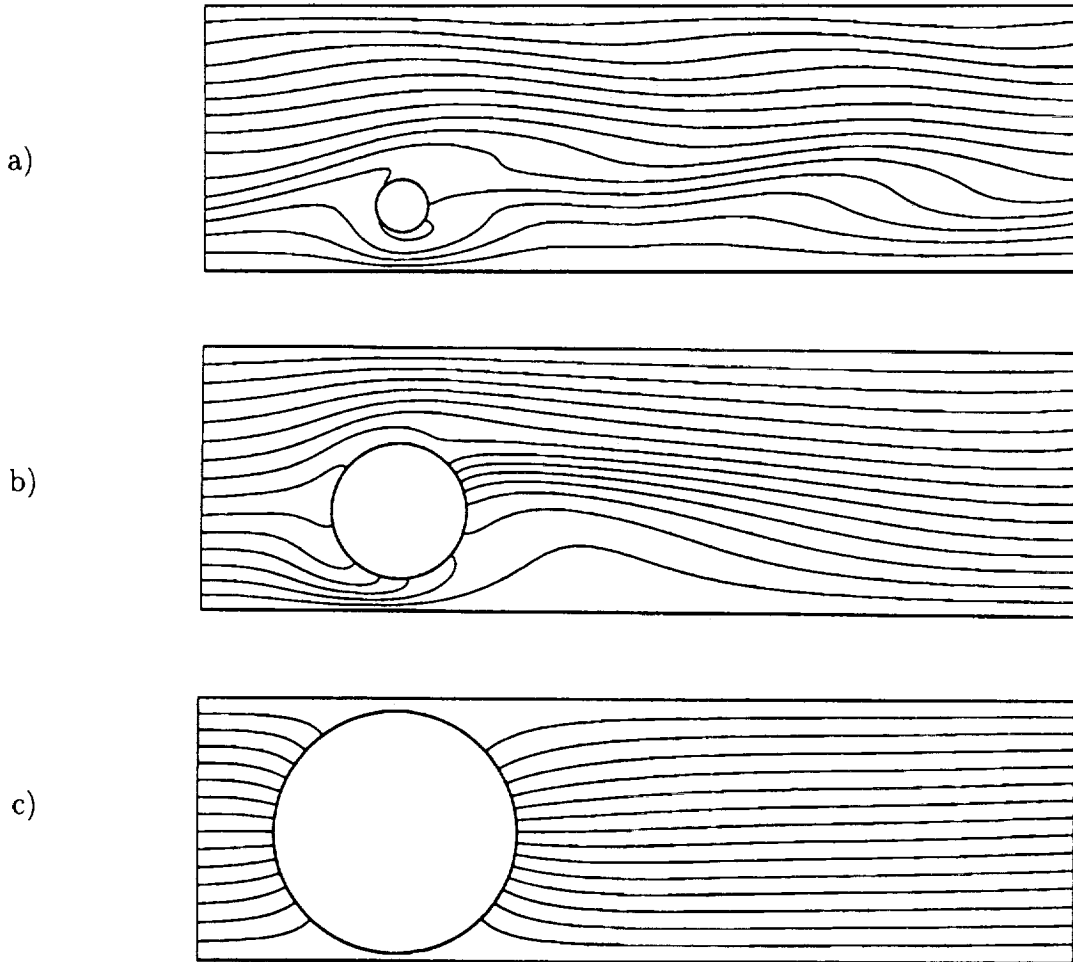


Figure 3: Eddy-promoter flow isotherms at one instant in time for (a) $\{a, R\} = \{.50, .20\}$, (b) $\{a, R\} = \{.81, .75\}$, and (c) $\{a, R\} = \{.98, .92\}$; in all cases, $Re = 300$, $Pr = 1$, $L = 6.666$.

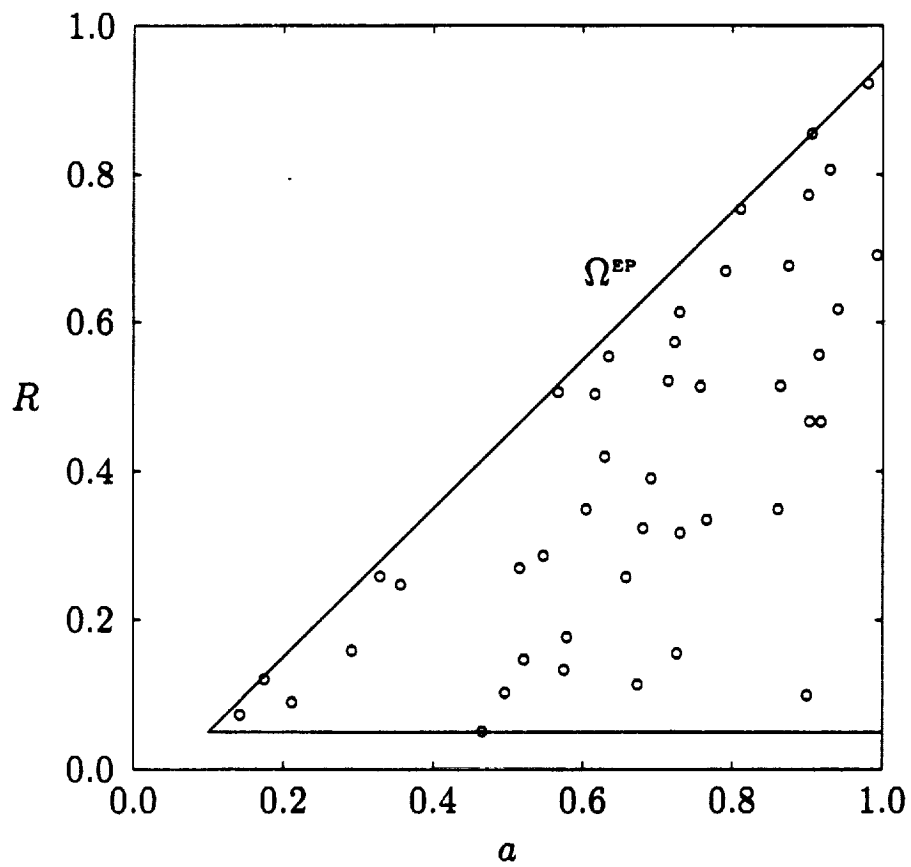


Figure 4: Validation sample (\circ) for the the simple Nusselt number surrogate $\tilde{Q}(a, R) = 1$.

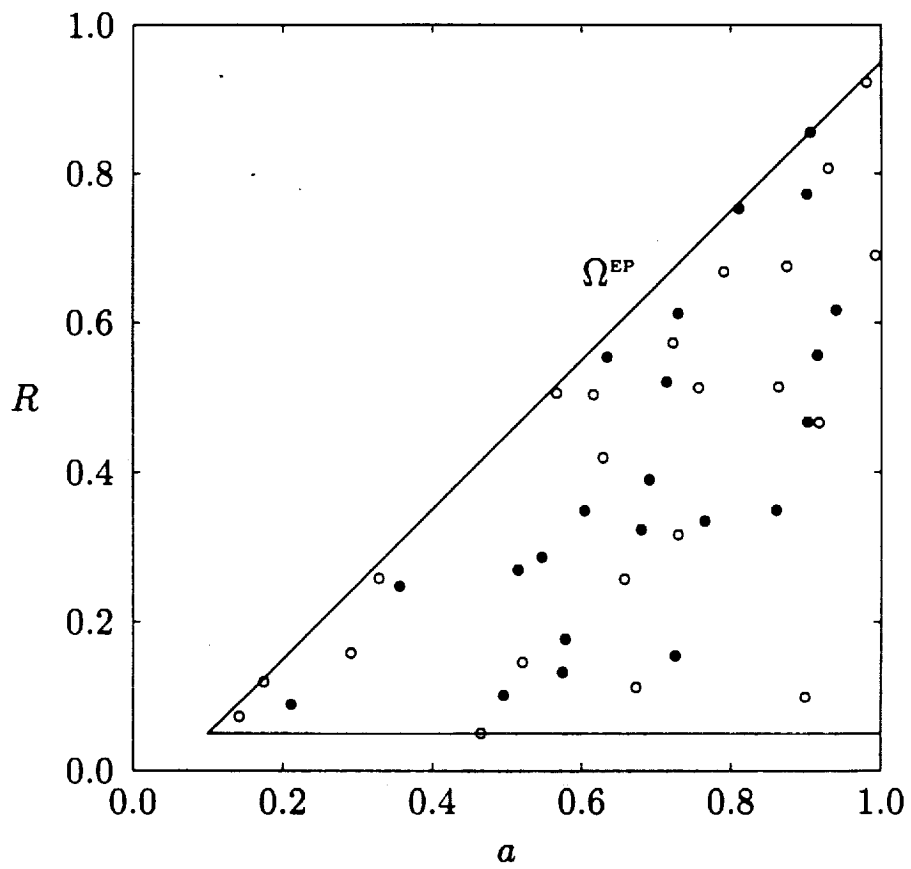


Figure 5: Construction (\bullet) and validation (\circ) samples for the flowrate and Nusselt number surrogates.

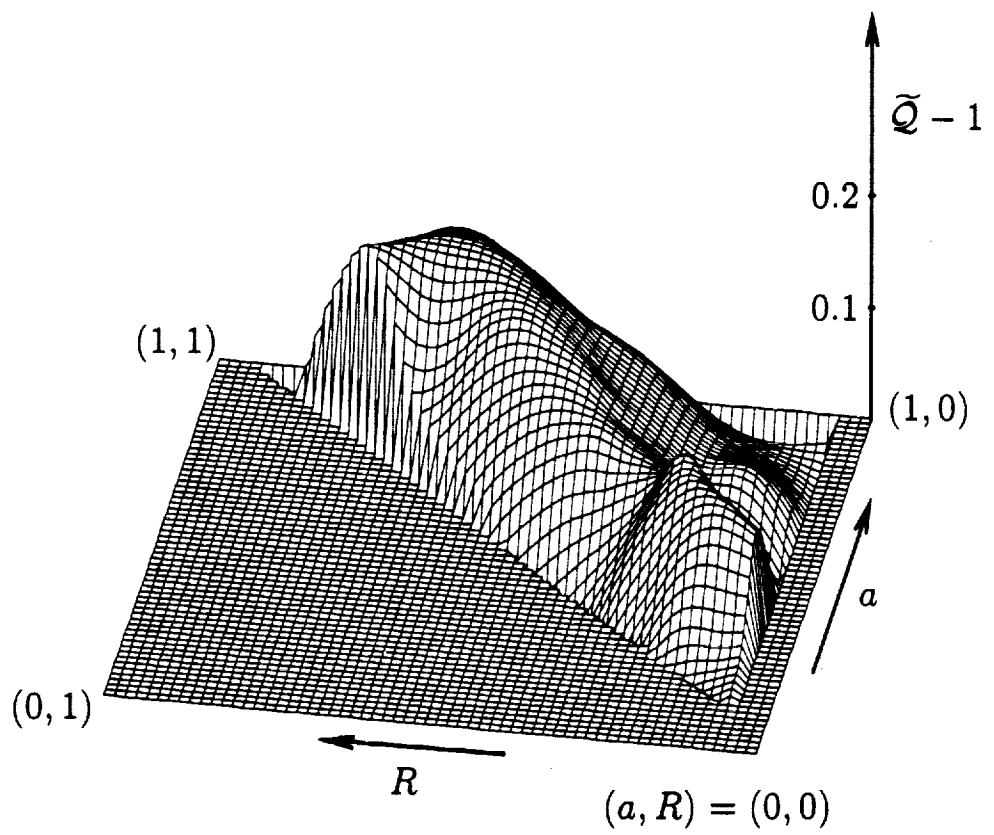
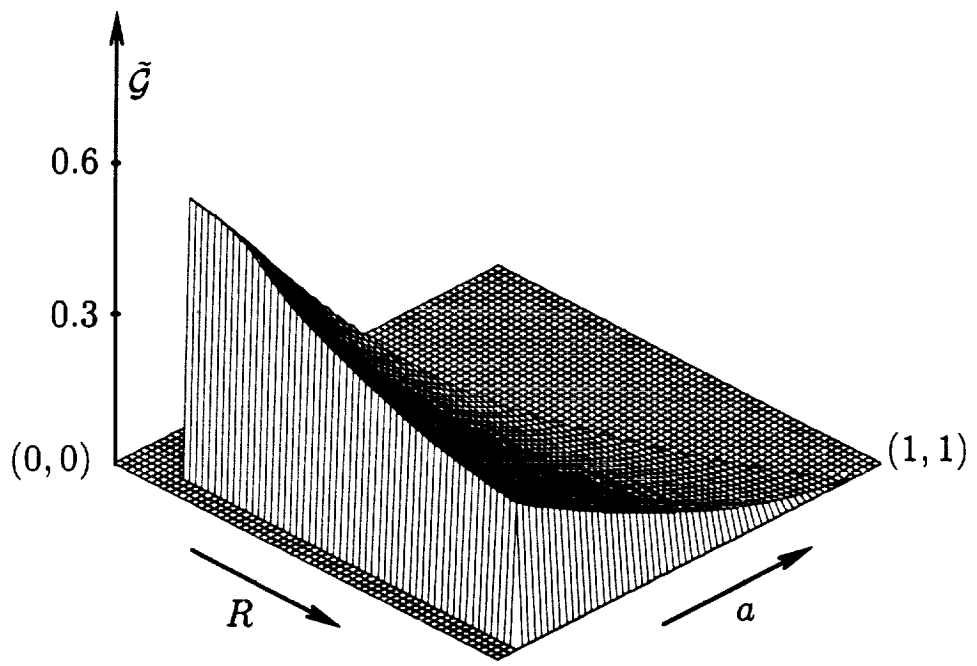


Figure 6: (a) Flowrate surrogate, $\tilde{G}(a, R)$. (b) Nusselt number surrogate, $\tilde{Q}(a, R)$.

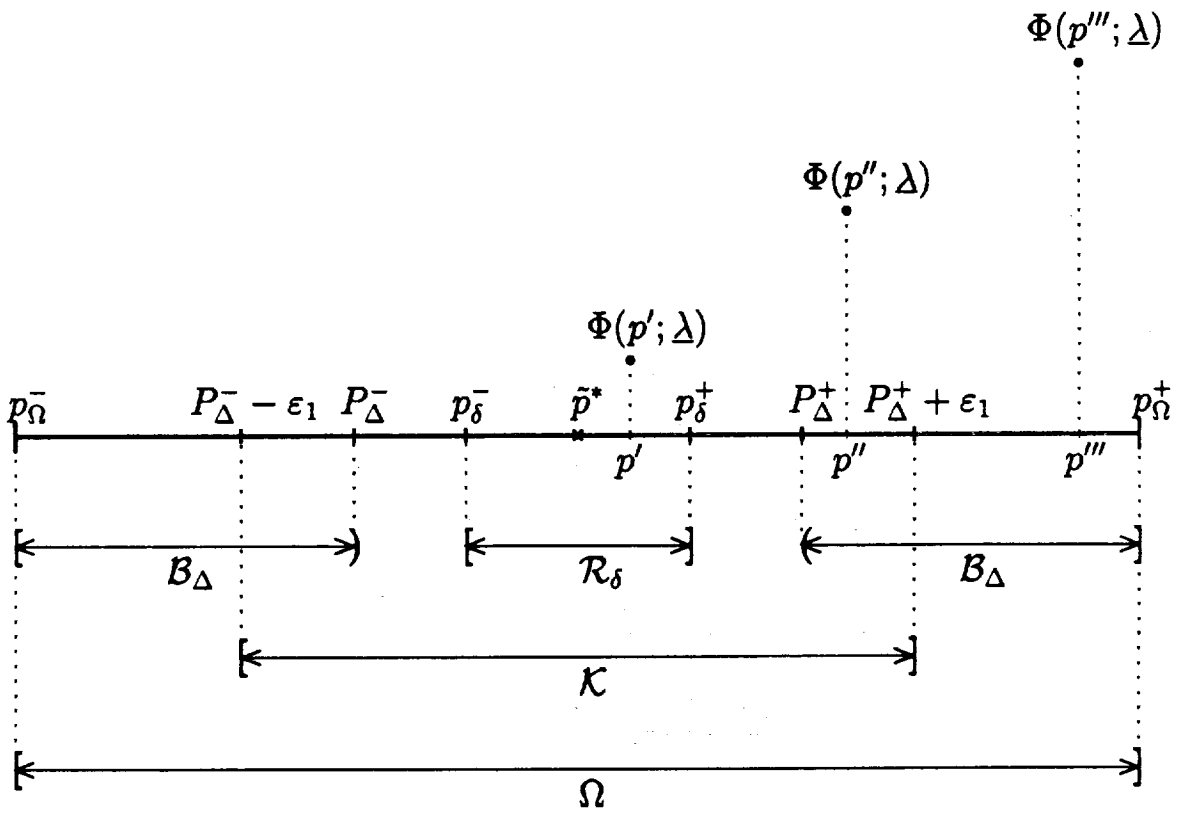


Figure 7: Regions Ω , \mathcal{R}_{δ} , \mathcal{B}_{Δ} , and \mathcal{K} associated with one-dimensional quasi-convex optimization problem.

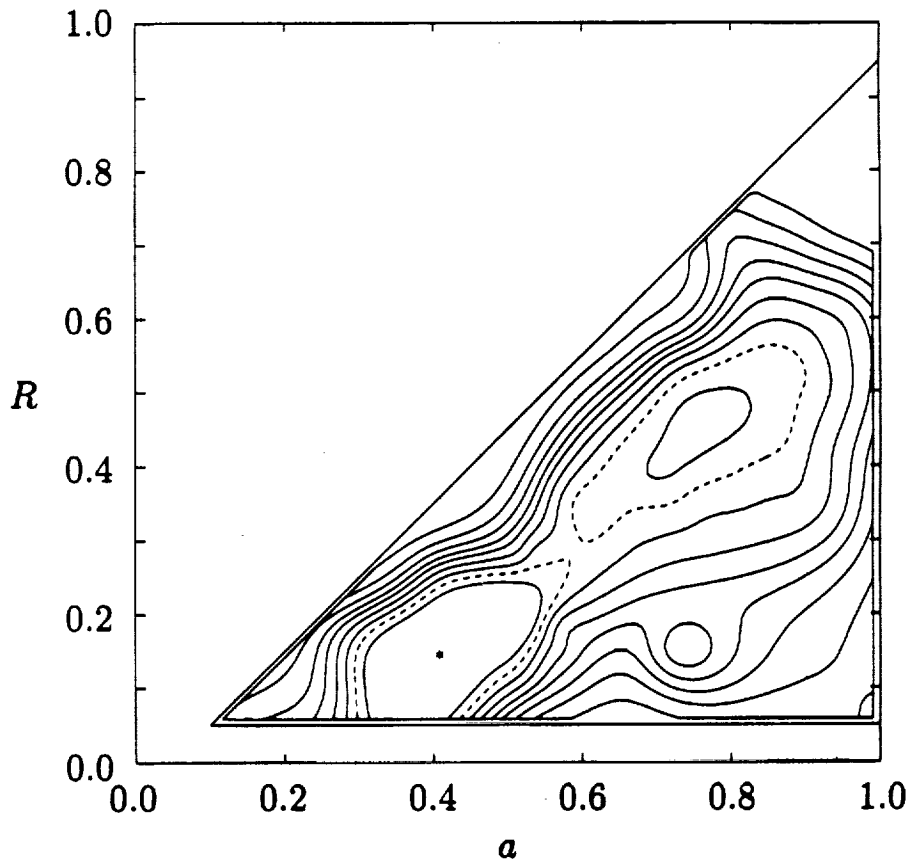


Figure 8: Contour plot for $\tilde{\Phi}^{\text{EP}}(\mathbf{p}^{\text{EP}}; \Delta^{\text{EP},1})$. The surrogate minimum over Ω^{EP} is $\tilde{\Phi}_{\min}^{\text{EP}}(\Delta^{\text{EP},1}) = .223$; the surrogate maximum over Ω^{EP} is .550; contours delineate level sets of relative weighted volume .1, .2, ..., .9. The dashed contour encloses \mathcal{RR}_s for $\tau = 2$.

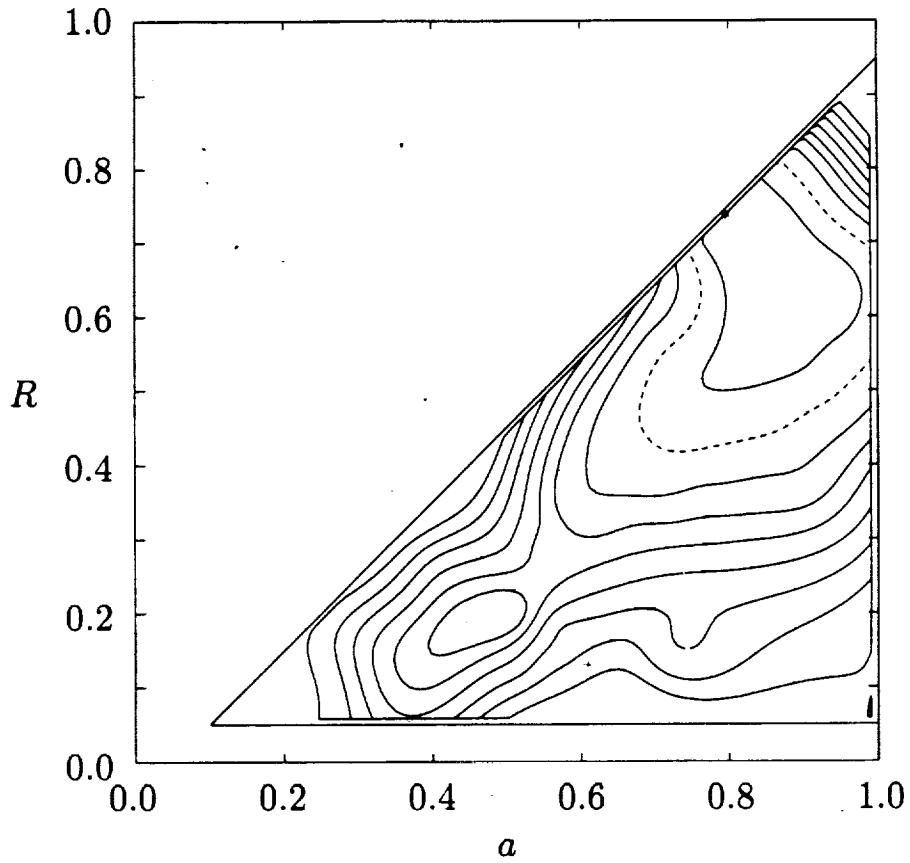


Figure 9: Contour plot for $\tilde{\Phi}^{\text{EP}}(\underline{p}^{\text{EP}}; \underline{\lambda}^{\text{EP},2})$. The surrogate minimum over Ω^{EP} is $\tilde{\Phi}_{\min}^{\text{EP}}(\underline{\lambda}^{\text{EP},2}) = .190$; the surrogate maximum over Ω^{EP} is .385; contours delineate level sets of relative weighted volume .1, .2, ..., .9. The dashed contour encloses \mathfrak{RR}_δ for $r = 2$.

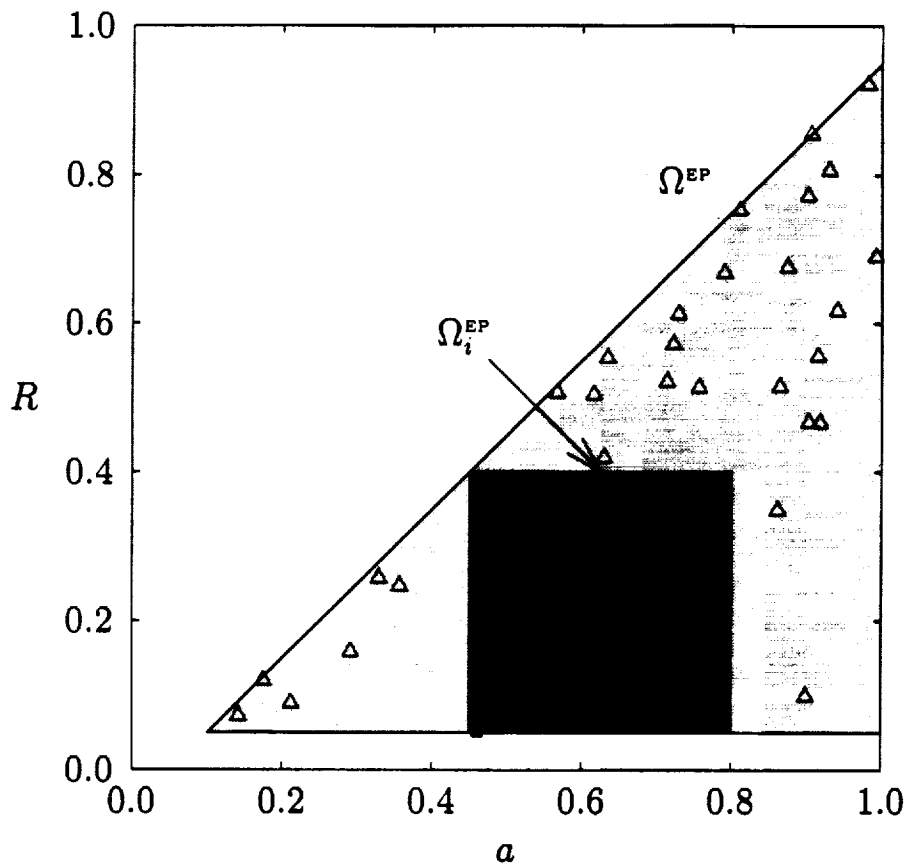
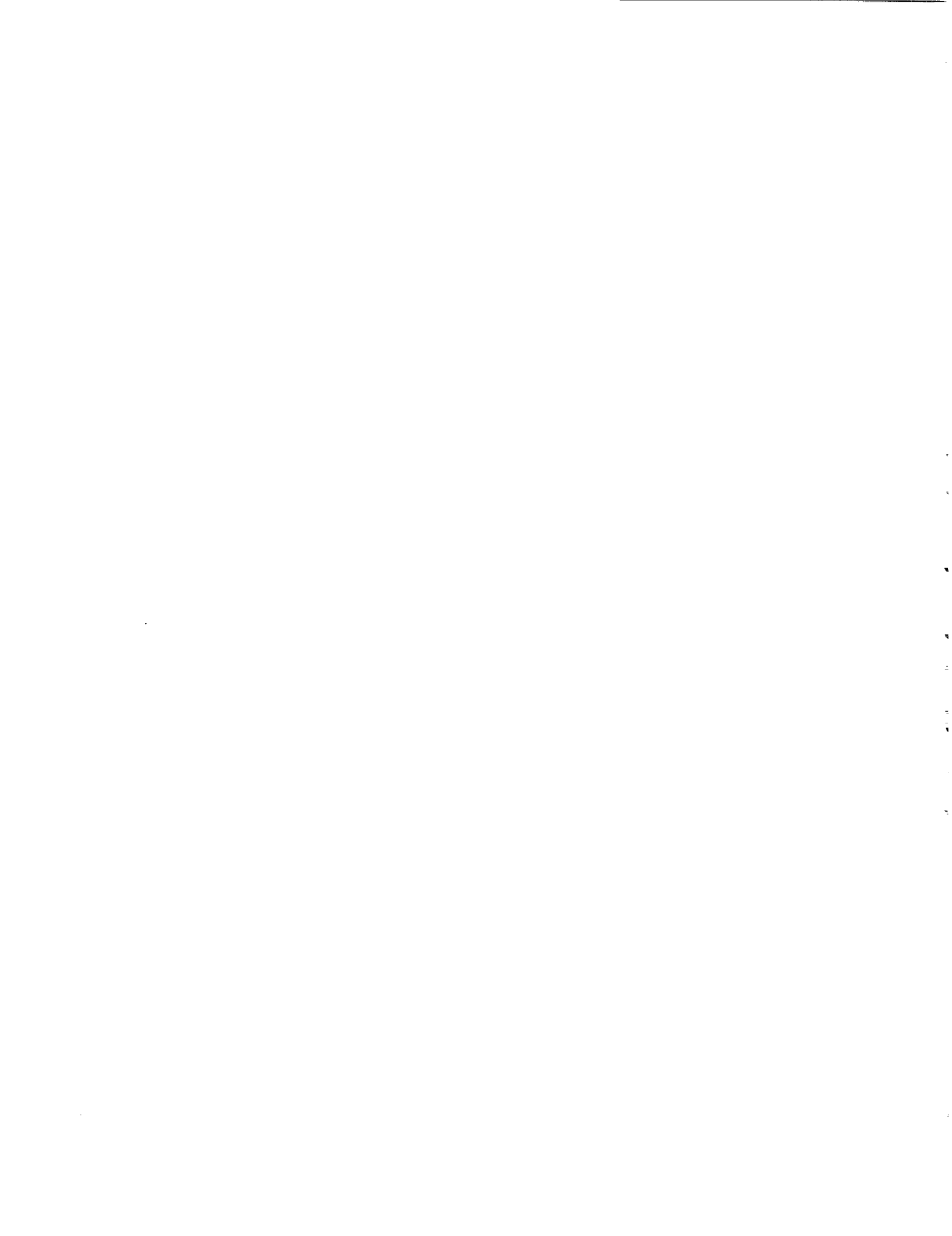


Figure 10: Hypothetical initial, Ω_i^{EP} , and subsequent, Ω^{EP} , design spaces for databoard example. Input points for the subsequent study comprise new points (Δ) and points recycled from the initial study (\square).







REPORT DOCUMENTATION PAGE

Form Approved
GME No. 0704-0185

This reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204 Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Project Room (1224), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE August 1993	3. REPORT TYPE AND DATES COVERED Contractor Report
4. TITLE AND SUBTITLE SURROGATES FOR NUMERICAL SIMULATIONS; OPTIMIZATION OF EDDY-PROMOTER HEAT EXCHANGERS		5. FUNDING NUMBERS C NAS1-19480	
6. AUTHOR(S) Serhat Yesilyurt Anthony T. Patera		WU 505-90-52-01	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Computer Applications in Science and Engineering Mail Stop 132C, NASA Langley Research Center Hampton, VA 23681-0001		8. PERFORMING ORGANIZATION REPORT NUMBER ICASE Report No. 93-50	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Langley Research Center Hampton, VA 23681-0001		10. SPONSORING/MONITORING AGENCY REPORT NUMBER NASA CR-191510 ICASE Report No. 93-50	
11. SUPPLEMENTARY NOTES Langley Technical Monitor: Michael F. Card Final Report		Submitted to Computer Methods in Applied Mechanics and Engineering (CMAME)	
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified - Unlimited Subject Category 34		12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) Although the advent of fast and inexpensive parallel computers has rendered numerous previously intractable calculations feasible, many numerical simulations remain too resource-intensive to be directly inserted in engineering optimization efforts. An attractive alternative to direct insertion considers models for computational systems: the expensive simulation is evoked only to construct and validate a simplified input-output model; this simplified input-output model then serves as a simulation surrogate in subsequent engineering optimization studies. We present here a simple "Bayesian-validated" statistical framework for the construction, validation, and purposive application of static computer simulation surrogates. As an example, we consider dissipation-transport optimization of laminar-flow eddy-promoter heat exchangers: parallel spectral element Navier-Stokes calculations serve to construct and validate surrogates for the flowrate and Nusselt number; these surrogates then represent the originating Navier-Stokes equations in the ensuing design process.			
14. SUBJECT TERMS numerical analysis; fluid dynamics; optimization; surrogates; Monte Carlo Method; modelling		15. NUMBER OF PAGES 56	
		16. PRICE CODE A04	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT