



## Computer-aided molecular product-process design under property uncertainties – A Monte Carlo based optimization strategy

**Frutiger, Jérôme; Cignitti, Stefano; Abildskov, Jens; Woodley, John M.; Sin, Gürkan**

*Published in:*  
Computers & Chemical Engineering

*Link to article, DOI:*  
[10.1016/j.compchemeng.2018.08.021](https://doi.org/10.1016/j.compchemeng.2018.08.021)

*Publication date:*  
2019

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Frutiger, J., Cignitti, S., Abildskov, J., Woodley, J. M., & Sin, G. (2019). Computer-aided molecular product-process design under property uncertainties – A Monte Carlo based optimization strategy. *Computers & Chemical Engineering*, 122, 247-257. <https://doi.org/10.1016/j.compchemeng.2018.08.021>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## Accepted Manuscript

Computer-aided molecular product-process design under property uncertainties – A Monte Carlo based optimization strategy

Jérôme Frutiger , Stefano Cignitti , Jens Abildskov ,  
John M. Woodley , Gürkan Sin

PII: S0098-1354(18)30866-4  
DOI: <https://doi.org/10.1016/j.compchemeng.2018.08.021>  
Reference: CACE 6193



To appear in: *Computers and Chemical Engineering*

Received date: 25 January 2018  
Revised date: 16 July 2018  
Accepted date: 14 August 2018

Please cite this article as: Jérôme Frutiger , Stefano Cignitti , Jens Abildskov , John M. Woodley , Gürkan Sin , Computer-aided molecular product-process design under property uncertainties – A Monte Carlo based optimization strategy, *Computers and Chemical Engineering* (2018), doi: <https://doi.org/10.1016/j.compchemeng.2018.08.021>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Highlights

- Property uncertainties incorporated into computer-aided molecular design problems
- Novel Monte Carlo based optimization strategy for product-process design
- Quantification and sampling of group contribution factor uncertainties
- Robust-reliable as well as optimistic-explorative analysis approach
- HFO working fluids for waste heat recovery of marine diesel engine

ACCEPTED MANUSCRIPT

# Computer-aided molecular product-process design under property uncertainties – A Monte Carlo based optimization strategy

*Jérôme Frutiger, Stefano Cignitti, Jens Abildskov, John M. Woodley, Gürkan Sin \**

Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Søtofts Plads, Building 229, DK-2800 Kgs. Lyngby, Denmark

\*corresponding author

## **ABSTRACT:**

A methodology is presented to solve a computer-aided molecular design (CAMD) and process design model problems under consideration of fluid property uncertainty. The uncertainties of the group contribution (GC) property prediction models are quantified for which asymptotic approximation of the covariance of parameter estimation errors is performed following a regression analysis. A Monte Carlo sampling technique generates GC factor samples within the respective uncertainties, which are evaluated separately as constraints to the CAMD optimization problem. The methodology is applied to identify working fluid candidates for an organic Rankine cycle used as waste heat recovery system in a marine diesel engine. CAMD under property uncertainties allows 1) identifying robust and more reliable molecules with respect to property uncertainties (conservative approach) and 2) enhancing the search space in order to find

potentially globally optimal working fluids (optimistic approach). Suitable Hydrofluoroolefins (HFO) have been identified as potential working fluids for waste heat recovery.

Keywords

Chemical product-process design; Property uncertainty, Group contribution; Working fluid; Organic Rankine cycle

## 1. Introduction

In recent years computer-aided molecular design (CAMD) has been combined with classical process design problems in order to simultaneously find the most suitable process chemicals along with an optimal process layout and process conditions [1]. Hence, product-process design [2] as an integrated solution approach has been applied to provide a solution to a variety of problems and challenges. Papadopoulos et al. [3] give a detailed overview over both solution strategies and application domains of product-process design problems. Among others product-process design principles have been used to develop novel refrigerants [4][5] and working fluids [6][7], polymers [8][9], fuels [10] and biofuels [11], as well as solvents [12][13] and ionic liquids [14].

CAMD relies heavily on chemical property values which can either be obtained through prediction, e.g. using group contribution (GC) methods [15]), as well as from databases containing experimental data, such as NIST ThermoDataEngine (TDE) [16][17] or DIPPR AIChE [18].

In this work, we would like to emphasize and tackle a major challenge with respect to property models in product-process design problems namely the uncertainties caused by the property prediction models [19]. In every CAMD algorithm molecules are generated and their respective

properties need to be predicted. These uncertainties can influence the result of the selection or optimization algorithm [20]. Hence, even though the optimization algorithm identified an optimal solution, the solution may be suboptimal when verified with experimental data or further experimentation due to property uncertainties.

A prediction model for a certain chemical property is usually obtained through a regression analysis which estimates the parameters of the model structure describing the relationship between the property value and the structures of the molecules. In case of GC methods, one of the commonly used property prediction models for CAMD, the GC factors are the parameters to be estimated [21]. Uncertainties in these parameters obtained from the regression process are inevitable due to the incomplete knowledge and approximation of the physical property with the model [22]. Frutiger et al. [15] described in detail, how the uncertainty of GC model parameters (i.e. GC factors) and predictions can be analyzed and quantified.

Several authors have addressed the problem of property prediction uncertainty in the context of CAMD [23]. Maranas [20][24] used multi-variate probability density distributions to reflect the likelihood of the realization of property parameters. The probability distributions were imposed on a nominal value for each property that accounts for the difference between the experimental value and the prediction of the property. Molecules were afterwards identified through stochastic property matching or optimization. Kim et al. [25] as well as Xu et al. [26] suggested to use uncertainty factors for the parameters that account for the discrepancy between predictive and experimental property values. Ng et al. [22][23] have introduced a fuzzy optimization algorithm to take into account property uncertainties in connection with molecular design. One of the main difficulties when taking property uncertainties into account is that the parameter uncertainty should be estimated adequately from the regression and not arbitrarily estimated. In linear and

non-linear regression, the parameter uncertainties are reflected through the 95% confidence interval obtained from the parameter covariance matrix [15]. The second major challenge is that taking uncertainties into account usually increases the computational demand in finding a feasible solution to the product-process design optimization [27].

In this contribution, we present a new approach to solve product-process design problems taking into account property uncertainties, namely the use of a Monte Carlo based sampling strategy [28][29]. Monte Carlo sampling principles have been known for 2 decades as a tool for molecular design [30][31][32]. However, only recently Monte Carlo method has been applied in the context of assessing property uncertainties in process systems engineering applications. Monte Carlo based sampling has been used to propagate property parameters to the process model output uncertainty [33] as well as to quantify equations of state parameters and subsequently assess the uncertainty of equations of state on the level of process model output uncertainty [34]. Furthermore, Monte Carlo based global sensitivity analysis process models with respect to property descriptors has been suggested [35]. In process design Monte Carlo based propagation of property uncertainties to process model output uncertainties can be of use, when a compound needs to be selected for a process from a list of candidate molecules. Each candidate compound can be ranked according to the uncertainty range of the process model output subject to the compounds property uncertainties [33]. Most recently Frutiger et al. [36] introduced a novel reverse engineering approach for the fluid selection process based on Monte Carlo sampling: A Monte Carlo sampling algorithm has been used to generate sets of different property parameter combinations (virtual molecules), which are subsequently evaluated in a process model. The distance between the property values of the virtual molecules and the uncertainty bound of the properties of real compounds (collected from a database) are calculated.

The molecules that lie closest to the top-performing virtual molecules are suggested as the optimal compounds for the given process and are further analyzed through evaluation.

To illustrate the relevance of property uncertainties in CAMD problems, the case of working fluids for thermodynamic cycles is considered. For energy systems, such as heat pumps and power cycles, a working fluid needs to be selected. However, the selection of the working fluids affects the cycle parameters to be set. This makes the selection of a working fluid for a given energy system a product-process design problem. An optimization of a product-process design problem for a given cycle can give one optimal working fluid for a given system. However, as it has been shown by Frutiger et al. [37] the models are highly affected by property model uncertainties. Simple screening of working fluids can lead to different optimal working fluids depending on the property uncertainties.

CAMD has been applied by various authors for the development of working fluids as described in the review of Linke et al. [38]. Property uncertainties have not been integrated systematically into CAMD optimizations for working fluid design. Only Papadopoulos et al. [39] showed a CAMD algorithm to account for property uncertainties of working fluids, by applying nonlinear sensitivity analysis. However, the authors analyzed the sensitivity of mixture properties in the CAMD optimization and did not quantify the uncertainty in the pure component properties. Even though the impact on pure component property uncertainties on working fluid selection [33] has been recognized, there is, to our best knowledge, no CAMD optimization algorithm considering working fluids properties. Considering property uncertainties when solving CAMD optimization problems for energy systems is of particular relevance for the development of novel climate-friendly working fluids such as Hydrofluoroolefins (HFOs). Property prediction models for HFOs are often associated with uncertainties, as Mondejar et al. [40] showed. If the property

uncertainties are not considered in the CAMD optimization, it is possible that one rules out a potentially powerful HFO as working fluids, or one focuses the development on a generated HFO whose properties measured properties are in fact very different from the predicted ones.

The method, described in this work for solving CAMD problems considering property uncertainties based on Monte Carlo sampling, address the following new features:

- Usage of Monte Carlo sampling technique to obtain different sets GC property parameters (GC factors) within their respective true uncertainty range obtained from the regression analysis and covariance matrix
- Solution of CAMD problem as a scenario based optimization using each sample set of GC factors as a scenario of constraints
- Interpretation of the product-process model results including its uncertainties with conservative or optimistic approaches depending on the design intent.

We apply the methodology to an industrial case study of energy systems engineering namely the identification of a working fluid for an organic Rankine cycle (ORC) system for waste heat recovery from an industrial size ship diesel engine from MAN Diesel A/S on a Mærsk A/S container ship [41].

The paper is organized as follows: i) the overall methodology for the solution of the CAMD problem under uncertainty is described ii) the CAMD problem is formulated and solved for the identification of suitable working fluids for the ORC system for waste heat recovery iii) the results for the generated working fluids are compared and analyzed with respect to the modelling uncertainties.

ACCEPTED MANUSCRIPT

## 2. Method and Tools

### 2.1. Methodology for Monte Carlo based CAMD under property uncertainty

The methodology consists of the major steps given in Table 1 and illustrated in Figure 1. The methodology, and in particular Step 1 and Step 4, is based on the work of Cignitti et al. [1], who developed an integrated product-process design framework, in. The authors also applied their framework for the development of working fluids for thermodynamic cycles [41] without considering property uncertainties.

Table 1: Overview of the methodology.

Step 1	CAMD optimization problem with process and property constraints and selection of property prediction models
Step 2	Quantification of property parameter uncertainties
Step 3	Monte Carlo based sampling of property parameters
Step 4	Solution of CAMD optimization for each property parameter set
Step 5	Analysis of optimal molecules obtained from CAMD solution subject to corresponding uncertainty

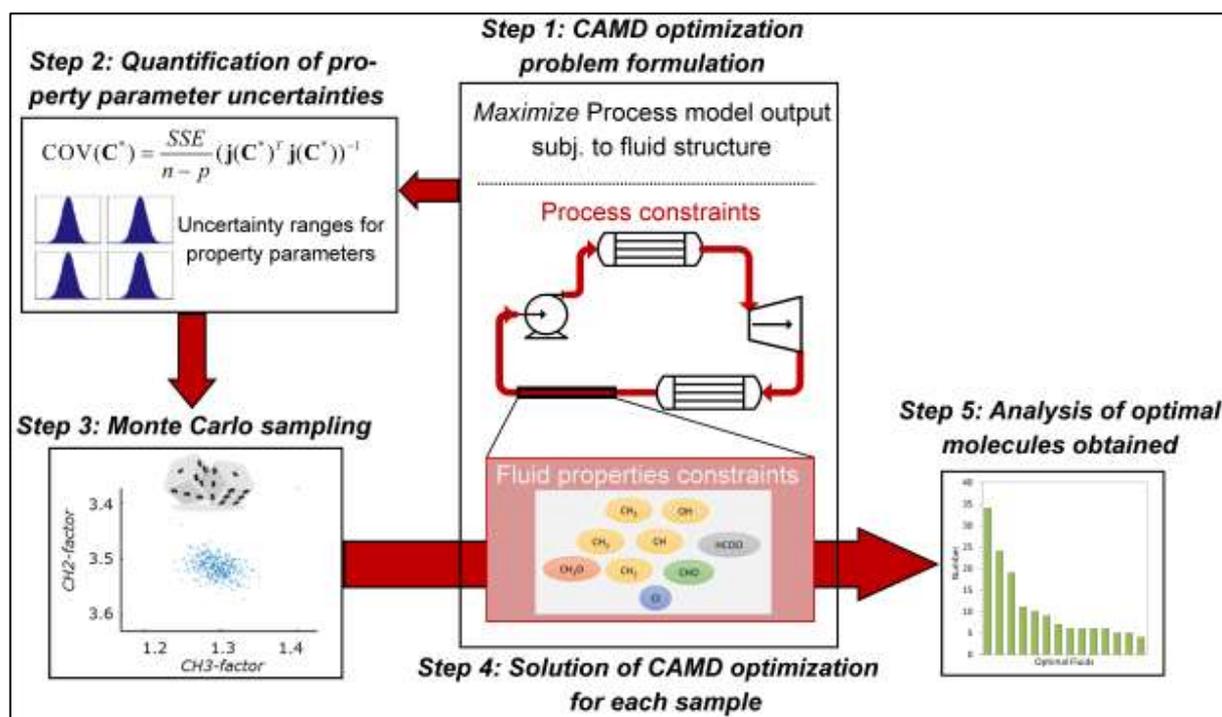


Figure 1. Illustration of the overall methodology.

*Step 1: CAMD optimization problem with process and property constraints and selection of property prediction models*

A CAMD problem can be formulated as a Mixed Integer Nonlinear Programming (MINLP) as described by Zhang et al. [42]. The objective function of CAMD optimization problem can correspond to desired product property (e.g. high solubility or low environmental impact) or to the desired performance of a given process, where the product is used (e.g. low energy demand, low cost). The variables of a product-process optimization problem are the structural compositions of the molecules, integer values for connectivity between functional groups, and process conditions. A CAMD MINLP problem consists of constitutional model equations and constraints related to the property prediction models as well as mass, energy and fugacity

balances to describe the process. The MINLP problem of this work is formulated in general form according to Cignitti et al. [1]:

$$\text{Objective function:} \quad \max J(\mathbf{x}, \boldsymbol{\theta}) \quad (1)$$

$$\text{Process equations:} \quad 0 = f_1(\mathbf{x}, \boldsymbol{\theta}) \quad (2)$$

$$\text{Product equations:} \quad 0 = f_2(\mathbf{x}, \boldsymbol{\theta}, \mathbf{Q}) \quad (3)$$

$$\text{Property models:} \quad 0 = g_1(\mathbf{x}) - \boldsymbol{\theta} \quad (4)$$

$$\text{Process conditions:} \quad 0 \geq g_2(\mathbf{x}, \boldsymbol{\theta}) \quad (5)$$

$J$  is the objective function.  $f_1$  represents process equations such as energy, mass and momentum balances,  $\mathbf{x}$  is a vector of continuous process variables,  $\boldsymbol{\theta}$  is a vector of product properties, and  $\mathbf{Q}$  is a vector containing structural information (e.g. GC groups and occurrences) about the molecule. Property models are represented through  $g_1$  and  $g_2$  represents conditional equations that the process solution must satisfy, such as mole fraction definition. The property models  $g_1$  can be expressed through connectivity equations and prediction equations [43]. In this work the Marrero/Gani GC method [44] is used as property prediction method. By definition this method combines the structural contributions from a specific functional group (1st order parameters), from polyfunctional (2nd order parameters) as well as from structural groups (3rd order parameters) to estimate a certain property of a molecule. However, in order to simplify the prediction approach, only first order GC factors have been considered in this work.

$$\text{Connectivity equation:} \quad \sum_i n_i (2 - v_i) = 2 \quad (6)$$

$$\text{Property prediction:} \quad F(y^{\text{pred}}) = \sum_i n_i C_i \quad (7)$$

Eq. (6) represents the octet rule,  $n_i$  is the number of first-order GC groups and  $v_i$  is the valency of the respective group  $i$ . Eq. (7) is the general formulation of a first-order Marrero/Gani GC model

$F(y^{\text{pred}})$  for a property  $y$  with  $C_i$  being the first order GC-factors.  $F(y^{\text{pred}})$  is individual for each property. The GC factors  $C_i$  are obtained through fitting the property model to experimental data of known molecules.

*Step 2: Quantification of property parameter uncertainties*

The GC factors are considered to be uncertain for the CAMD optimization problem. The uncertainty is quantified by applying the methodology of Frutiger et al. [45], which describes in detail the development, parameter estimation and uncertainty analysis methods for GC based property models. The GC parameter estimation and uncertainty analysis is summarized in the following.

1. Experimental data  $y^{\text{pred}}$  of chemical compounds for the respective properties are collected from DIPPR AIChE database [18], or alternatively from NIST TDE ) [16][17].
2. The Marrero-Gani GC equations are fitted to the experimental data using non-linear robust regression to obtain the values of the GC-factors [46]:

$$\mathbf{C}^* = \arg \min \sum_i w_i \cdot (y_i^{\text{exp}} - y_i^{\text{pred}})^2 \quad (8)$$

$$w_i = \frac{1}{1 + (y_i^{\text{exp}} - y_i^{\text{pred}})^2} \quad (9)$$

Eq. (8) describes the minimization of the residuals to obtain the vector of GC factor estimates  $\mathbf{C}^*$ .  $w_i$  is a weight factor [47] by which each residual is weighted. In this example the Cauchy weight is used, placing high weights on small residuals and small weights on large residuals. The weights are updated recursively. This decreases the influence of data points producing large residuals (not following the model), i.e. potential outliers.

3. The covariance matrix of the parameter estimation errors for the corresponding GC-factors is calculated based on the asymptotic approximation of the covariance matrix estimators [47][48] as follows:

$$\text{COV}(\mathbf{C}^*) = \frac{SSE}{n-p} (\mathbf{j}(\mathbf{C}^*)^T \mathbf{j}(\mathbf{C}^*))^{-1} \quad (10)$$

$$SSE = \sum_i (y_i^{\text{exp}} - y_i^{\text{pred}})^2 \quad (11)$$

$n$  is the number of data points and  $p$  the number of parameters. The Jacobian  $\mathbf{j}$  is the local sensitivity of the property model with respect to the GC factor values  $\mathbf{C}^*$ .  $SSE$  is the sum of squared errors between the experimental and predicted data.

4. The correlation coefficients between the GC factors can be obtained from the covariance-matrix:

$$\text{Corr}(C_i^*, C_j^*) = \frac{\text{COV}(C_i^*, C_j^*)}{\text{Var}(C_i^*) \text{Var}(C_j^*)} \quad (12)$$

In Eq. (12)  $\text{COV}(C_i^*, C_j^*)$  is the respective element of  $C_i^*$  and  $C_j^*$  of the covariance matrix and  $\text{Var}(C_i^*)$  and  $\text{Var}(C_j^*)$  are the variances of the respective GC factors.

5. The square-root of the diagonal elements of the covariance-matrix defines the standard deviation of the GC-factors. Hence, the uncertainty range of the corresponding GC-factors can be reported corresponding to 95% confidence:

$$\mathbf{C}_{1-\alpha}^* = \mathbf{C}^* \pm \sqrt{\text{diag}(\text{COV}(\mathbf{C}^*))} \cdot t(n-p, \alpha_t / 2) \quad (13)$$

In Eq. (13) describes the 95% confidence interval of the GC factors.  $t(n-p, \alpha_i/2)$  is the  $t$ -distribution value corresponding to the two tailed students  $t$ -distribution, here  $\alpha_i=0.05$ , corresponding to 95% confidence.

The obtained GC-factor 95% confidence interval is used as an input uncertainty ranges for probabilistic Monte Carlo sampling.

*Step 3: Monte Carlo based sampling of property parameters*

Monte Carlo procedure is used for the sampling of the GC parameter sets [22]. The range of uncertainty (corresponding to the 95% confidence intervals calculated in Step 2) of the different GC-factors defines the parameter input space. Latin Hypercube Sampling method [49] is utilized to generate random samples from this input space. We assume that the probability of uncertainty follows a normal distribution, which is the default assumption in nonlinear regression theory (where parameter estimation errors are assumed to follow normal distribution around the mean value of the estimates with covariance characterized by  $C$  calculated above). GC-factors can be strongly correlated, which influence the prediction of the respective properties [45]. Hence, the correlation between the GC-parameters should be taken into account by using the rank-based method for correlation control of Iman and Conover [50]. In total a number of 200-250 sample sets of GC-factors should be generated.

*Step 4: Solution of CAMD optimization for each property parameter set*

Each GC sample set forms a scenario for the CAMD optimization [51] [52]. The GC models are part of the constraints of the optimization problem. Hence the CAMD problem is solved for 200-250 alternative property constraints. Each optimization solution provides a potentially alternative

solution of the optimal molecular structure that gives an optimal solution for the objective function. In each scenario optimal molecular structure may be different. Considering all rankings of the respective optimization solutions, those fluids that appear the most often are considered to be optimal subject to the property model uncertainty.

*Step 5: Analysis of optimal molecules obtained from CAMD solution subject to corresponding uncertainty*

There are several ways for the analysis of the results of the corresponding optimizations. In this work we suggest to rank an obtained molecule according to two criteria: 1) the number of times the respective molecule occurred as optimal solution from the different scenarios and 2) the corresponding value of the objective function. The number of times of optimality reflects how uncertain the optimization result is with respect to property uncertainties and how sensitive the result is with respect to the properties of the corresponding molecule. A molecule that occurs many times as the optimal solution is more reliable with respect to its real and *a priori* unknown performance, which can be assessed in further detailed simulations (with more accurate property models) or experimental measurements. On the other hand, ranking according to the objective function value corresponds to a more optimistic approach of the product identification, since it is possible that the highest objective function outcome of all scenarios is assigned with a molecule that only occurred a few times as the optimal solution. This corresponds to a solution with higher uncertainty with respect to the properties and higher risk that the molecule does not show the same performance in detailed simulations or experiments. However, the optimistic approach is desirable for an explorative research, when the goal of the product design is to find new and unknown molecular alternatives for a certain process.

## 2.2. Case study: Working fluid design for waste heat recovery of marine diesel engine

The methodology described in 2.1 is applied step-by-step to a case study for the identification of a working fluid for an organic Rankine cycle (ORC) system for waste heat recovery from the exhaust gas of an industrial size ship diesel engine from MAN Diesel A/S on a Mærsk A/S container ship [28][53][54]. The aim of the case study is to identify suitable working fluids for the given problem under consideration of the uncertainties in the property prediction model. The waste heat recovery system reduces the fuel consumption by decreasing the load on main or auxiliary engines [55].

*Step 1: CAMD optimization problem with process and property constraints and selection of property prediction models*

According to Step 1 of the methodology, the case study objective problem (objective function and constraints) needs to be formulated. It is based on the work of Cignitti et al. [41], which describes the MINLP problem and solution approaches for the optimization in detail including all equations. Here, we summarize the most important parts of the problem formulation.

The product design problem is the generation of optimal chemical compounds for an ORC, the process design problem is the identification of the optimal process variables, the lower and upper pressure level of the cycle. The two problems are solved simultaneously, which makes the MINLP problem a product-process design problem.

Figure 2 gives an overview over the ORC cycle. The main goal of the ORC is to recover heat from the engine exhaust gas and convert it into power, i.e. electricity for the ship. Hence, the objective function of the product-process design is the net power output  $W_{\text{NET}}$  of the cycle.

Case study objective function  $\max W_{\text{NET}}$  (14)

The ORC consists of an evaporator system (including pre-heater and super-heater), a turbine, a condenser and a pump. The heat is taken up from the engine exhaust gas, partly converted into power in the turbine and released to the sea water (see Figure 2). In order to increase the efficiency of the cycle an internal heat exchanger (recuperator) is used. In total there are 7 state points in the cycle, where pressure, temperature, enthalpy and entropy need to be calculated. The specification for the cycle model have been taken from the work of Cignitti et al. [41].

Table 2 summarizes all process parameters, process variable constraints, and target property ranges.

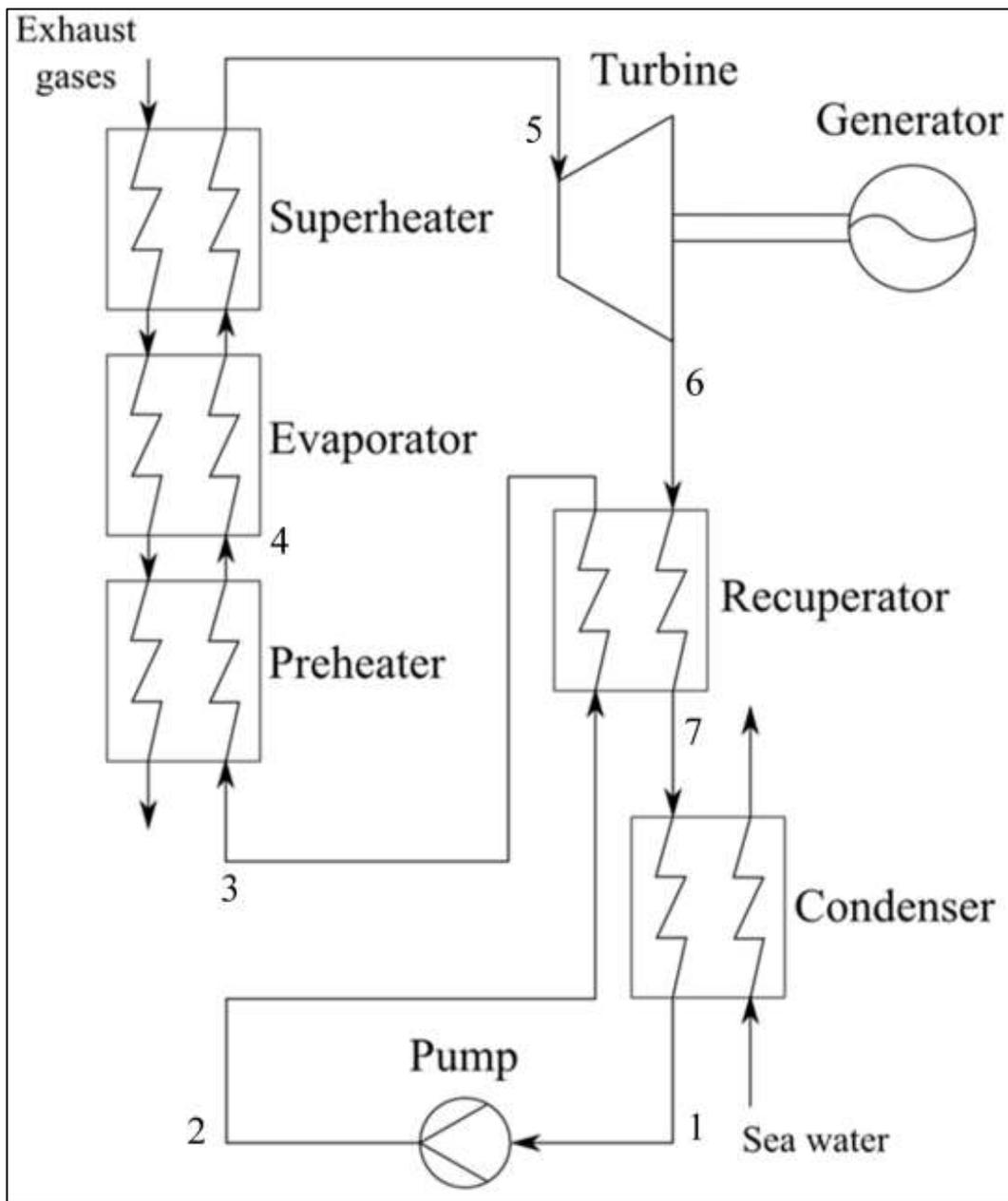


Figure 2: Overview over the ORC for waste heat recovery from the exhaust gas of Marine Diesel engine [28].

Table 2. Process parameters and specifications for the ORC system [41].

<i>Process parameter</i>	<i>Value</i>
Exhaust gas inlet temperature	509K
Exhaust gas mass flow	2161.15mol/s
Exhaust gas heat capacity	31.86J/(molK)
Sea water inlet temperature	299K
Sea water outlet temperature	294K
Pump isentropic efficiency	0.7
Turbine isentropic efficiency	0.8
<i>Process variable constraints</i>	<i>Value</i>
Minimum temperature difference in superheater	10K
Minimum temperature difference in evaporator	20K
Minimum temperature difference in recuperator	10K
Minimum working fluid condenser outlet temperature	300K
Minimum working fluid recuperator outlet temperature (cold side)	408K
Minimum turbine inlet pressure	0.05bar
<i>Target molecules boundaries</i>	<i>Value</i>
Number of groups per molecule	3-9
Critical temperature $T_c$ range	368K-600K
Critical pressure $P_c$ range	18bar-100bar
Acentric factor $\omega$ range	0-1
Heat capacity $c_p$ range	No boundaries

At each state the thermodynamic properties (i.e. specific enthalpies and entropies as well as fugacities) of the desired working fluid are calculated (see the work of Cignitti et al. [1][41] for all detailed equations). Hence, an equation of state (EoS) is required consisting of an ideal contribution (i.e. the ideal gas enthalpy and entropy) and a departure function (difference between ideal and real fluid behavior). The ideal gas heat capacity  $C_p$  formulated in polynomial form according to Joback and Reid [56] is used to estimate the ideal gas enthalpy and entropy. Soave-Redlich-Kwong Equation of State (SRK-EoS) was selected to determine the departure functions of the thermodynamic properties [57]. The advantage of SRK-EoS is its relatively small number of required fundamental parameters as a 3-parametric cubic equation of state: The critical temperature ( $T_c$ ), the critical pressure ( $P_c$ ) and the acentric factor ( $\omega$ ).

MINLP problem solution will find a working fluid with optimal target properties in  $T_c$ ,  $P_c$ ,  $\omega$ ,  $C_p$  and the optimal process conditions (turbine inlet temperature  $T_t$  and pressure  $P_t$ ) that maximize the power output  $W_{NET}$ .

The cycle model equations are included into the MINLP as equality and inequality constraints, which need to be fulfilled. For example on each stage the SRK-EoS for the enthalpy needs to be fulfilled with following departure function [41]:

$$h_{st} - h_{ref} = \int C_p dT_{st} + RT_{st} (Z_{st} - 1) - \frac{\Psi R}{\Omega} \left( \sqrt{\alpha T_{st} T_c} \kappa + \alpha T_c \right) \ln \left( \frac{Z_{st} + \Omega \frac{T_c P_{st}}{P_c T_{st}}}{Z_{st}} \right) \quad (15)$$

$$\alpha = \left( 1 + \kappa \left( 1 - \sqrt{T_{st} / T_c} \right) \right)^2 \quad (16)$$

$$\kappa = 0.37464 + 1.54226\omega - 0.26992\omega^2 \quad (17)$$

where  $h_{st}$  is the stage enthalpy,  $h_{ref}$  the reference enthalpy,  $T_{st}$  the stage temperature,  $P_{st}$  the stage pressure,  $Z_{st}$  the stage compressibility factor,  $\Psi$  and  $\Omega$  the SRK-parameters. The enthalpy is used for calculation of heat flow for energy balances and for the net power output calculation. Likewise, entropy is calculated similarly using the departure function for isentropic states (pump and turbine), and the real pump and turbine stage is then calculated with 70% and 80% efficiency, respectively. At any given equilibrium point, the isofugacity condition needs to be fulfilled through the calculation of fugacity coefficient through the same EoS.

The fluid target properties to be optimized in the MINLP problem are the input parameters to SRK EoS and the ideal gas heat capacity constants  $C_p$  [1]. The GC property prediction methods [19] are based on the method of Marrero and Gani [44]. The group contribution property prediction models for the critical temperature  $T_c$ , the critical pressure  $P_c$ , and the acentric factor  $\omega$  are defined according to Hukkerikar et al. [19]:

$$T_c = T_{c0} \ln \left( \sum_i n_i T_{ci} \right) \quad (18)$$

$$P_c = P_{c0} + \left( \frac{1}{P_{c1} \sum_i n_i P_{ci}} \right)^2 \quad (19)$$

$$\omega = \omega_A \left[ \log \left( \sum_i n_i \omega_i + \omega_C \right) \right]^{\omega_B} \quad (20)$$

where  $T_{ci}$ ,  $P_{ci}$  as well as  $\omega_i$  are the group contribution factors of Marrero-Gani group  $i$ ,  $n_i$  is the number of respective group factors.  $T_{c0}$ ,  $P_{c0}$ ,  $P_{c1}$ ,  $\omega_A$ ,  $\omega_B$ , and  $\omega_C$  are the correlation constants of the Marrero-Gani GC method to experimental data [45]. The heat capacity GC model for is given as follows

$$C_p = \left( \sum_i n_i A_i - A_0 \right) + \left( \sum_i n_i B_i + B_0 \right) T_{st} + \left( \sum_i n_i C_i - C_0 \right) T_{st}^3 + \left( \sum_i n_i D_i - D_0 \right) T_{st}^3 \quad (21)$$

where  $A_i$ ,  $B_i$ ,  $C_i$  and  $D_i$  as well as  $\omega_i$  are the group contribution factors of Joback Reid [56] group  $i$ ,  $n_i$  is the number of respective group factors.  $A_0$ ,  $B_0$ ,  $C_0$ ,  $D_0$  are the correlation constants.  $T_{st}$  is the temperature of the corresponding stage.

A set of first order groups Marrero-Gani groups is selected to build the working fluids: CH<sub>3</sub>, CH<sub>2</sub>, CH, C, CH<sub>2</sub>=CH, CH=CH, CH<sub>2</sub>=C, CH=C, C=C, CF<sub>3</sub>, CF<sub>2</sub>, CF, CH<sub>2</sub>F, CHF, CHF<sub>2</sub>, HCCIF, CCIF<sub>2</sub>. These groups allow the generation of hydrofluoro-olefines (HFO), consisting of fluorine with C=C double-bonds. HFO's are currently investigated as an alternative to current hydrofluorocarbon (HFC) working fluids due to their low global warming potential and zero ozone depletion potential [58]. It is one of the goals of this case study to explore novel HFOs as working fluids for organic Rankine cycles.

The MINLP consists of 79 blocks of equations, 21 blocks of variables and 21 discrete variables, the CPU time for solving the problem one time is 5.04s on a conventional personal computer [41].

*Step 2: Quantification of property parameter uncertainties*

The property uncertainties for each of the GC factors of the property models ( $T_c$ ,  $P_c$ ,  $\omega$ ,  $C_p$ ) are obtained through regression analysis and uncertainty analysis as described in the methodology in section 2.1. Experimental data from DIPPR AIChE database [18] has been taken to fit the GC models.

As an example for this step, the outcomes are shown for  $P_c$ . Figure 3 illustrates the agreement between experimental and the predicted values of the including the respective 95%-confidence interval of the prediction for every substance used in the regression.

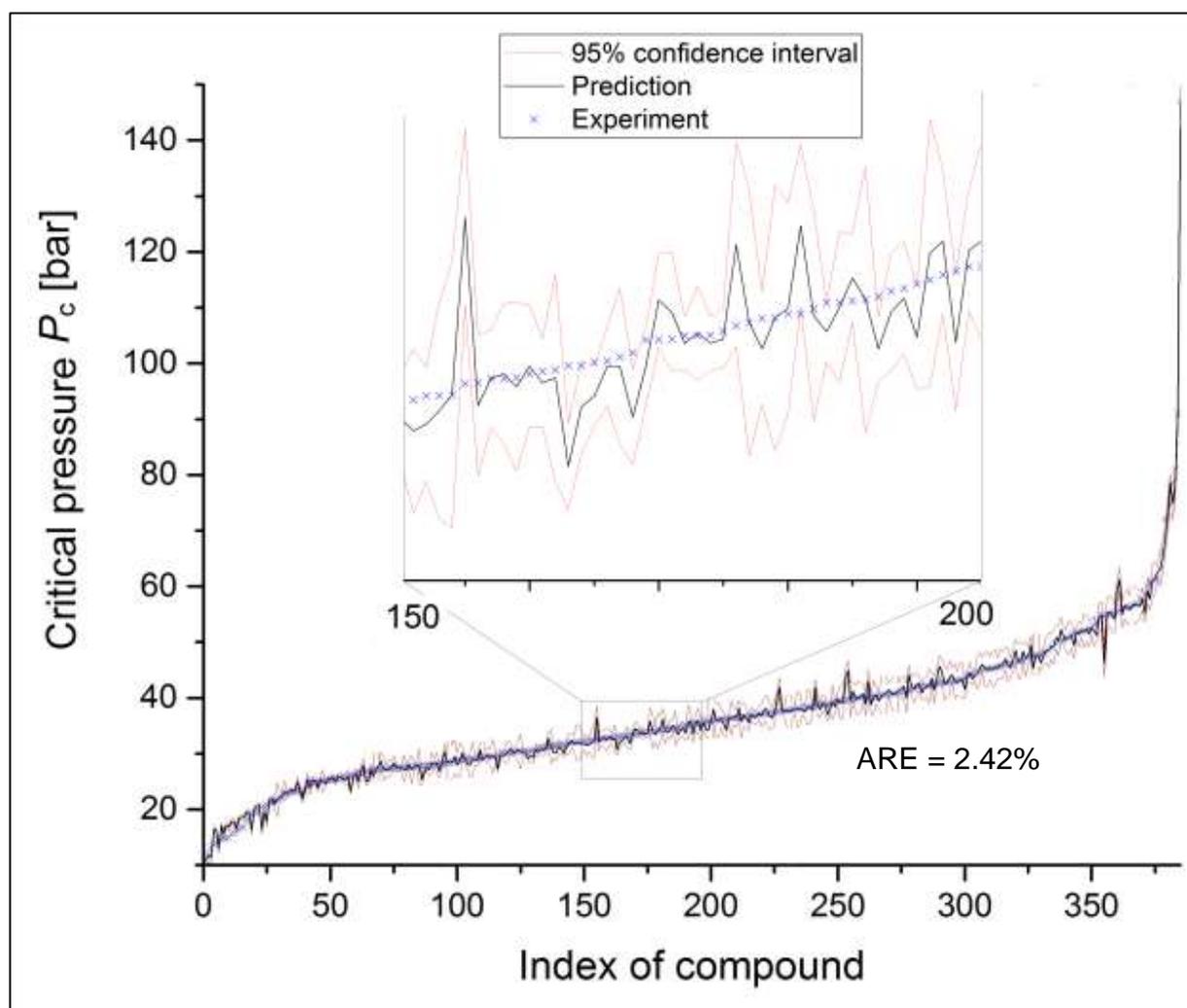


Figure 3: Experimental as well as predicted value with 95%-confidence intervals from regression analysis of the compounds ordered from smallest to highest value. 50 compounds have been enlarged to show the distribution of the experimental values around the prediction. *ARE* is the average relative error between the experimental and predicted value.

The group contribution values including their corresponding uncertainties for the selected groups in the CAMD problem are shown in Table 3. Furthermore, the correlation matrix between the GC factors is presented in Table 4. If the uncertainties of a GC-factor are comparatively large, e.g. the standard deviation is a high fraction (>50%) or even larger than the mean value, samples from this uncertainty ranges may lead in feasible solutions. It is therefore recommended

to use an upper limit for the uncertainty bounds. The user of the methodology is free to choose this limit and it is a simple recommendation, which may not necessary in a specific problem. A sensitivity analysis or a simple evaluation of the samples in the process model without the optimization may support the decision on feasibility. In his study if the uncertainty of a GC factor was higher than 25% of the actual, the corresponding uncertainty bound was set to 25%, in order to obtain a feasible solution to the optimization problem.

Table 3. GC factors for the selected CAMD groups for  $P_c$  property prediction model.

Group	$P_c$ -GC factor	Uncertainty range (95%-confidence interval)	
CH3	$5.16 \cdot 10^{-3}$	$4.61 \cdot 10^{-3}$	$5.72 \cdot 10^{-3}$
CH2	$8.73 \cdot 10^{-3}$	$6.38 \cdot 10^{-3}$	$11.07 \cdot 10^{-3}$
CH	$1.23 \cdot 10^{-2}$	$1.05 \cdot 10^{-2}$	$1.41 \cdot 10^{-2}$
C	$1.50 \cdot 10^{-2}$	$1.06 \cdot 10^{-2}$	$1.94 \cdot 10^{-2}$
CH2=CH	$1.09 \cdot 10^{-2}$	$0.77 \cdot 10^{-2}$	$1.41 \cdot 10^{-2}$
CH=CH	$1.55 \cdot 10^{-2}$	$1.40 \cdot 10^{-2}$	$1.71 \cdot 10^{-2}$
CH2=C	$1.38 \cdot 10^{-2}$	$1.30 \cdot 10^{-2}$	$1.47 \cdot 10^{-2}$
CH=C	$1.97 \cdot 10^{-2}$	$1.86 \cdot 10^{-2}$	$2.09 \cdot 10^{-2}$
C=C	$4.70 \cdot 10^{-2}$	$4.54 \cdot 10^{-2}$	$4.86 \cdot 10^{-2}$
CH2F	$-1.53 \cdot 10^{-3}$	$-1.98 \cdot 10^{-3}$	$-1.07 \cdot 10^{-3}$
CHF	$-3.71 \cdot 10^{-2}$	$-5.37 \cdot 10^{-2}$	$-2.05 \cdot 10^{-2}$
CHF2	$6.47 \cdot 10^{-3}$	$3.30 \cdot 10^{-3}$	$9.65 \cdot 10^{-3}$
CF2	$8.91 \cdot 10^{-3}$	$6.14 \cdot 10^{-3}$	$11.69 \cdot 10^{-3}$
CF3	$3.06 \cdot 10^{-2}$	$2.13 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$

Table 4. Correlation matrix of the selected GC factors for  $P_c$  property prediction model.

	CH3	CH2	CH	C	CH2=CH	CH=CH	CH2=C	CH=C	C=C	CH2F	CHF	CHF2	CF2	CF3
CH3	1.00													
CH2	0.03	1.00												
CH	0.04	0.45	1.00											
C	0.11	-0.19	-0.19	1.00										
CH2=CH	0.04	-0.01	-0.01	0.78	1.00									
CH=CH	0.07	-0.01	-0.01	0.77	0.87	1.00								
CH2=C	0.03	0.06	0.06	0.73	0.91	0.89	1.00							
CH=C	0.01	0.07	0.07	0.72	0.87	0.90	0.93	1.00						
C=C	0.07	-0.24	-0.24	0.24	0.04	0.04	-0.03	-0.05	1.00					
CH2F	-0.04	-0.36	-0.35	0.24	0.03	0.03	-0.05	-0.07	0.36	1.00				
CHF	0.09	-0.11	-0.11	0.09	0.01	0.02	-0.01	-0.02	0.13	0.13	1.00			
CHF2	-0.03	-0.16	-0.15	0.10	0.01	0.01	-0.02	-0.03	0.16	0.35	0.04	1.00		
CF2	0.08	0.06	0.06	0.07	0.07	0.08	0.08	0.07	0.04	0.00	0.01	0.00	1.00	
CF3	0.48	0.21	0.17	-0.05	0.00	0.01	0.03	0.03	-0.13	-0.20	-0.01	-0.10	0.02	1.00

*Step 3: Monte Carlo based sampling of property parameters*

The uncertainty ranges of the GC factors as well as the correlation matrices are the input to step 3 of the methodology. 250 sets of GC parameters for all property prediction models are obtained through Monte Carlo procedure including correlation control. The 95% confidence interval of the Gc factors forms the uncertainty range in which sampling is taking place. Sampling has been performed in MATLAB [59]. Figure 4 illustrates the sampling results for three GC factors for the  $P_c$  property prediction model.

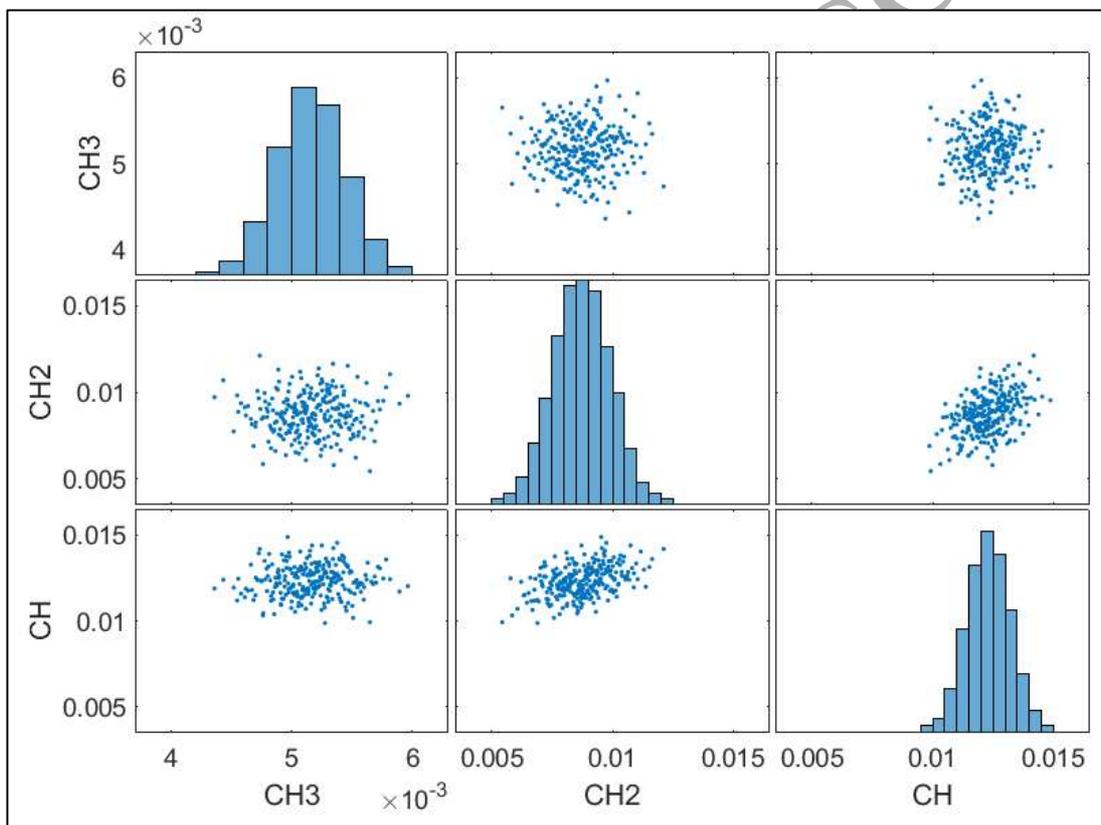


Figure 4. Illustration of the Monte Carlo samples for three GC factors (CH3, CH2 and CH) of the  $P_c$  property prediction model. As it can be seen, the samples are centered around the mean value through normal distributions and the correlation between the GC factors (see Table 3 and Table

4. Correlation matrix of the selected GC factors for  $Pc$  property prediction model.) are reflected in the samples.

The number of samples needs to be sufficiently high such that the results are reproducible. The reproducibility of the results can be checked for a given number of samples. The Monte Carlo integration error  $\sigma_M$  can be calculated as follows

$$\sigma_M = \sigma_{\text{obj}} / \sqrt{N} \quad (1)$$

where  $\sigma_{\text{obj}}$  is the standard deviations of the different outputs of the objective function and  $N$  is the number of samples [60].  $\sigma_{\text{obj}}$  is obtained from the distribution of the obtained optimal netpower outputs of the objective function (i.e.  $W_{\text{NET}}$ ). The number of samples  $N$  is sufficiently high, if  $\sigma_M$  stays constant, when reproducing the calculation.

*Step 4: Solution of CAMD optimization for each property parameter set*

Each of the sampling set is used separately to solve the MINLP problem. Hence, 250 optimization problems, each with a different set of GC factors are solved. A simultaneous approach, where the molecule and process are optimized in an integrated manner, was selected as the solution strategy. The global optimization solver selected was LINDOGlobal [41]. The software in which the MINLP has been implemented was GAMS [61]. The current methodology is independent of the solver. Hence, other state-of-the-art solvers such as BARON [62] and ANTIGONE [63] can be used as well.

*Step 5: Analysis of optimal molecules obtained from CAMD solution subject to corresponding uncertainty*

The solution of each of the 250 optimizations problems provides an optimal working fluid including the corresponding optimal process conditions (turbine inlet conditions) and maximum net power output value. The results are analysed according to the conservative and optimistic approach as described in section 2.1

ACCEPTED MANUSCRIPT

### 3. Results and Discussion

Figure 5 shows the outcomes of the solution of the 250 optimization problems with different. The working fluids are ranked according to the number of times that they occurred as the optimal solution. Furthermore, the corresponding objective function value (i.e. the net power output) is shown in the same plot for comparison.

ACCEPTED MANUSCRIPT

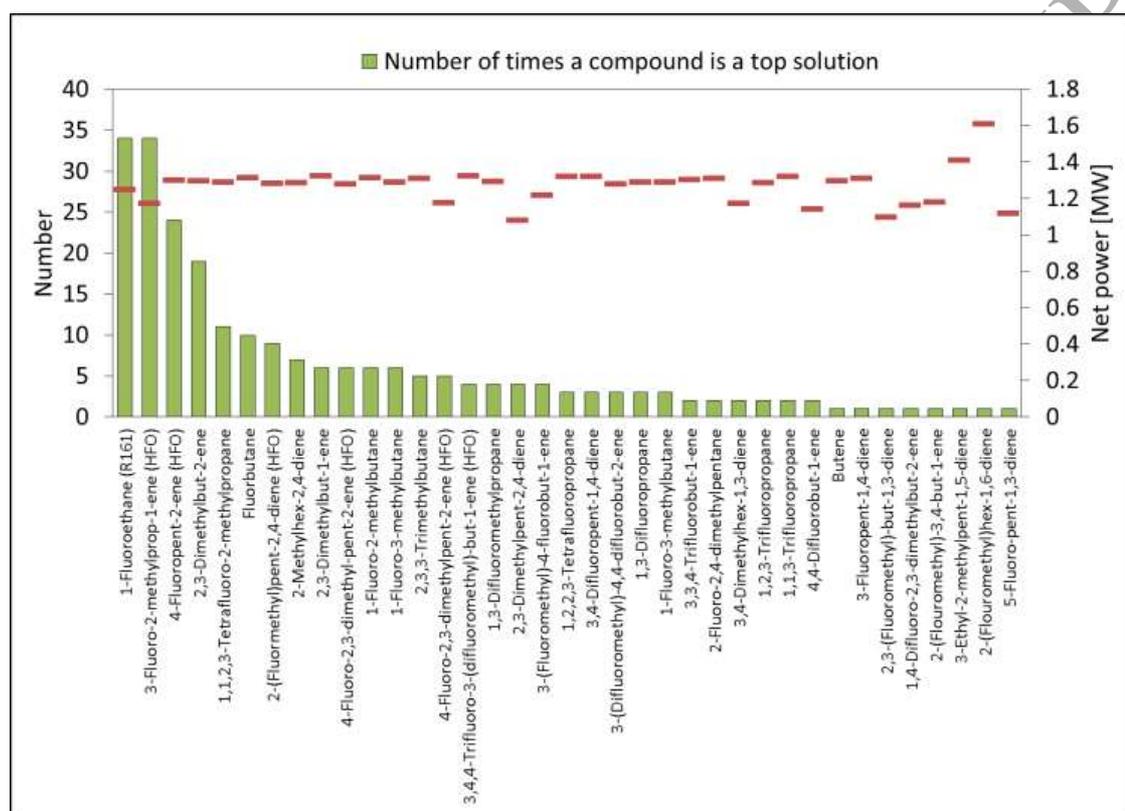
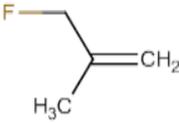
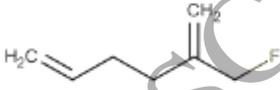
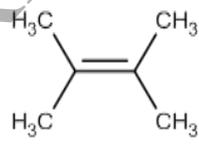


Figure 5: Ranking of working fluids according to occurrences as optimal working fluids (green bar) and corresponding net power output (red lines).

Figure 5 reflects the CAMD output uncertainty (molecular structure of the product and objective function) subject to the property input uncertainty. On the one hand, the number of occurrences as an optimal working fluid solution differs a lot (e.g. 1-fluoroethane (R161) 34 times, butene one time), on the other hand the net power output  $W_{\text{NET}}$  varies only by 5-10% (with two exceptions, that are discussed further down). There are two major causes this variation of the molecular structure: 1) The different uncertainty ranges of the GC factors can lead to high or low property uncertainties 2) The cycle state points may be in a sensitive region in terms of a certain fluid property or GC factor. The results of this study confirm that the solution of a product-process design problem (i.e. the molecular structure) can be very sensitive to property uncertainties.

The results allow the selection of working fluids for further analysis (detailed simulation or experimental investigation) based on different criteria. Hence, we present two different interpretation possibilities for the given results. Table 5 shows the detailed results for two promising HFO fluid candidates representing the more conservative (3-fluoro-2-methylprop-1-ene) as well the optimistic interpretation result (2-(fluoromethyl)hex-1,6-diene) and compares it to the base case optimization result (2,3-dimethylbut-2-ene).

Table 5: Summary of promising molecules as working fluids for waste heat recovery considering a conservative and optimistic approach compared to the base case optimization results.

Solution	Conservative	Optimistic	Base case optimum
Name	3-fluoro-2-methylprop-1-ene	2-(fluoromethyl)hexa-1,5-diene	2,3-dimethylbut-2-ene
Type	HFO	HFO	Alkene
Structure			
Number of time as optimal solution	34	1	19
$W_{NET}$ [kW]	1173	1609	1297
$P_c$ [bar]	44	45	24
$T_c$ [K]	496	579	513
$\omega$ [-]	0.25	0.28	0.16
$C_p$ [J/(molK)]	A	-60	-79.87
	B	0.72	1.13
	C	$-7.90 \cdot 10^{-4}$	$-1.24 \cdot 10^{-3}$
	D	$2.24 \cdot 10^{-6}$	$1.91 \cdot 10^{-6}$
			$3.45 \cdot 10^{-6}$

In a more conservative analysis approach, molecules from the left-hand side of the plot (e.g. 1-Fluoroethane, 3-fluoro-2-methylprop-1-ene, 4-fluoropent-2-ene, 2,3-dimethylbut-2-ene) are selected for further analysis, because they can be considered to be the most robust solutions subject to property uncertainties, (i.e. the variation of the GC factors within their corresponding uncertainty range). One of the aims of this study was also to identify novel HFOs as climate-friendly working fluids to existing ones. The 3-fluoro-2-methylprop-1-ene (on the left-hand side of Figure 5 with an occurrence of 34 times) is a compound shows robust behaviour with respect

to uncertainties and should potentially be explored further through experimental measurements of the properties in the cycle.

2,3-dimethylbut-2-ene is the result using the mean value of GC factors obtained from the regression analysis and not the samples (base case optimization result). However, when the GC factors are varied within their model-based uncertainty (i.e. 95% confidence interval), three other compounds occur more often as an optimal solution. This can be explained that the properties of Dimethylbut-2-ene may be more sensitive regions of the SRK EoS, leading to lower  $W_{\text{NET}}$  upon variation.

In a more optimistic approach, the molecules showing the highest net power output on the right hand side of the plot (in particular) should be considered. Each of these compounds occurs only one time as an optimal solution, but they are potentially the best performing working fluids and their net power output is much higher than the base case optimization result. They represent the global optimum of the overall optimization problem.

Both approaches of interpreting the results can be useful depending on the initial goal of a study. If the user needs a solution that is more robust with respect to uncertainty, a conservative approach can be desired. However, if the user wants to explore potentially unknown molecules, the optimistic approach may support his decision.

When candidate fluids have been selected, a more detailed analysis of the fluid performance is necessary, in particular more the fluid behaviour with respect to more accurate or different property models and thorough process component models needs to be investigated.

## 4. Conclusion

The study presented a methodology to include property prediction uncertainties in CAMD using Monte Carlo based sampling. The methodology has been applied to identify optimal working fluids for a waste heat recovery system in a marine diesel engine.

The following are the main conclusions from CAMD study under property:

- Uncertainties in GC property prediction factors can have a strong influence on the optimal molecular structure obtained from the CAMD and hence the optimal design of the cycle process.
- A more conservative approach to the results allows finding robust (or more reliable) molecules with respect to property uncertainties.
- A more optimistic approach to the results allows identifying globally optimal more uncertain molecules (high risk – high gain approach).
- The study found several optimal HFO as climate-friendly potential working fluids, in particular 3-fluoro-2-methylprop-1-ene and 2-(fluoromethyl)hex-1,6-diene.

CAMD has always served as an explorative technique for the identification of suitable chemical products in chemical engineering application from a given search space. The consideration of property uncertainties adds an additional dimension to this exploration and provides much more information than a single solution of the optimization problem. In this study both more robust compounds and molecules giving higher objective function compared to the base case optimization were identified through solution of CAMD using Monte Carlo based sampling

strategy. These results emphasize strongly that uncertainties in property prediction models should be taken into account in CAMD studies and more general in product-process design problems.

## NOTATION

### Symbols

$J$	objective function
$f_1$	process equations
$f_2$	product equations
$\mathbf{x}$	vector of continuous process variables
$\boldsymbol{\theta}$	vector of product properties
$\mathbf{Q}$	vector containing structural information (e.g. GC groups)
$g_1, g_2$	conditional equations
$n_i$	number of first-order GC groups
$v_i$	valency of the respective group
$i$	group of group contribution model
$F(y^{\text{pred}})$	general formulation of a first-order Marrero/Gani GC model
$y$	generic fluid property
$y^{\text{pred}}$	prediction of generic fluid property
$y^{\text{exp}}$	experimental value of generic fluid property
$C_i$	group contribution factor
$n$	number of data point
$p$	number of parameters

$\mathbf{j}$	Jacobian, the local sensitivity of the property model with respect to the GC factor values
$\mathbf{C}^*$	GC factor values
$SSE$	sum of squared errors
$COV$	covariance matrix
$Corr$	correlation coefficient
$Var(C_i^*)$	variance of the respective GC factors
$t(n-p, \alpha_t / 2)$	t-distribution value corresponding to the two tailed students t-distribution
$w_i$	weight factor
$h_{st}$	stage enthalpy
$h_{ref}$	reference enthalpy
$T_{st}$	stage temperature
$P_{st}$	stage pressure
$Z_{st}$	stage compressibility factor
$\Psi$	SRK-parameter
$\Omega$	SRK-parameter
$\omega$	acentric factor [-]
$T_r$	reduced temperature [-]
$P_c$	critical pressure [bar]
$T_c$	critical temperature [C]
$W_{NET}$	net power output [kW]
A	constant of temperature dependent heat capacity [J/molK]
B	constant of temperature dependent heat capacity [J/molK]
C	constant of temperature dependent heat capacity [J/molK]
D	constant of temperature dependent heat capacity [J/molK]
$C_p(T)$	temperature dependent heat capacity [J/(mol K)]

## ACKNOWLEDGEMENTS

This work was supported by the Danish Council for Strategic Research in Sustainable Energy and Environment under the THERMCYC project (grant number 1377-00037A).

## AUTHOR INFORMATION

### Corresponding Author

\*Tel.: +45 45252806, E-mail address: [gsi@kt.dtu.dk](mailto:gsi@kt.dtu.dk)

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ACCEPTED MANUSCRIPT

## REFERENCES

- [1] Cignitti S, Mansouri SS, Woodley JM, Abildskov J. Systematic optimization-based integrated chemical product–process design framework. *Ind Eng Chem Res* 2018;57:677–88. doi:10.1021/acs.iecr.7b04216.
- [2] Gani R. Chemical product design: challenges and opportunities. *Comput Chem Eng* 2004;28:2441–57.
- [3] Papadopoulos AI, Linke P, Seferlis P. *Integrated Multiobjective Molecular and Process Design: Operational and Computational Frontiers*. vol. 39. Elsevier; 2017.
- [4] Harper PM, Gani R. *Computers & Chemical Engineering A multi-step and multi-level approach for computer aided molecular design* 2000;24:677–83.
- [5] Roskosch D, Atakan B. Reverse engineering of fluid selection for thermodynamic cycles with cubic equations of state, using a compression heat pump as example. *Energy* 2015;81:202–12.
- [6] Papadopoulos AI, Stijepovic M, Linke P. On the systematic design and selection of optimal working fluids for Organic Rankine Cycles. *Appl Therm Eng* 2010;30:760–9.
- [7] Cignitti S, Zhang L, Gani R. Computer-aided Framework for Design of Pure, Mixed and Blended Products. 12th Int Symp Process Syst Eng 25th Eur Symp Comput Aided Process Eng 2015;37:2093–8.
- [8] Venkatasubramanian V, Chan K, Lafayette W. Computer-aided molecular design using genetic algorithms. *Comput Chem Eng* 1994;18:833–44.
- [9] Pavurala N, Achenie LEK. Identifying polymer structures for oral drug delivery - A molecular design approach. *Comput Chem Eng* 2014;71:734–44.
- [10] Yunus NA, Gernaey K, Woodley J, Gani R. An integrated methodology for design of tailor-made blended products. *Proc 22nd Eur Symp Comput Aided Process Eng* 2012.
- [11] Hada S, Solvason CC, Eden MR. Characterization-based molecular design of bio-fuel additives using chemometric and property clustering techniques. *Front Energy Res* 2014;2:1–12. doi:10.3389/fenrg.2014.00020.
- [12] Harper PM, Gani R, Kolar P, Ishikawa T. Computer-aided molecular design with combined molecular modeling and group contribution. *Fluid Phase Equilib* 1999;158–160:337–47.
- [13] Papadopoulos AI, Badr S, Chremos A, Forte E, Zargiannis T, Seferlis P, et al. Computer-aided molecular design and selection of CO<sub>2</sub> capture solvents based on thermodynamics, reactivity and sustainability. *Mol Syst Des Eng* 2016;1:313–34.
- [14] Chong FK, Foo DCY, Eljack FT, Atilhan M, Chemmangattuvalappil NG. Ionic liquid design for enhanced carbon dioxide capture by computer-aided molecular design approach. *Clean Technol Environ Policy* 2015;17:1301–12. doi:10.1007/s10098-015-0938-5.
- [15] Frutiger J, Marcarie C, Abildskov J, Sin G. A comprehensive methodology for

- development, parameter estimation, and uncertainty analysis of group contribution based property models - an application to the heat of combustion. *J Chem Eng Data* 2016;61:602–13.
- [16] Frenkel M, Chirico RD, Diky V, Yan X, Dong Q, Muzny C. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. *J Chem Inf Model* 2005;45:816–38.
- [17] Diky V, Chirico RD, Muzny CD, Kazakov AF, Kroenlein K, Magee JW, et al. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 8. Properties of Material Streams and Solvent Design. *Chem Inf Model* 2013;53:249–66.
- [18] Rowley RL, Wilding W V., Oscarson, L. J, Knotts TA, Giles NF. DIPPR® Data Compilation of Pure Chemical Properties, Design Institute for Physical Properties, AIChE, New York, NY 2014.
- [19] Hukkerikar AS, Sarup B, Ten Kate A, Abildskov J, Sin G, Gani R. Group-contribution+ (GC+) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilib* 2012;321:25–43.
- [20] Maranas C. Optimal molecular design under property prediction uncertainty. *AIChE J* 1997;43:1250–64.
- [21] Kontogeorgis G, Gani R. Chapter 1: Introduction to computer aided property estimation. *Comput Aided Chem Eng* 2004;19:3–26. *Comput Aided Chem Eng* 2004;19:3–26.
- [22] Sin G, Gernaey K V., Eliasson Lantz A. Good Modeling Practice for PAT Applications: Propagation of Input Uncertainty and Sensitivity Analysis. *Biotechnol Prog* 2009;25:1043–53.
- [23] Ng LY, Chong FK, Chemmangattuvalappil NG. Challenges and opportunities in computer-aided molecular design. *Comput Chem Eng* 2014;81:115–29. doi:10.1016/j.compchemeng.2015.03.009.
- [24] Maranas C. Optimization accounting for property prediction uncertainty in polymer design. *Comput Chem Eng* 1997;21:1019–24.
- [25] Kim K-J, Diwekar UM. Efficient Combinatorial Optimization under Uncertainty. 1. Algorithmic Development. *Ind Eng Chem Res* 2002;41:1276–84.
- [26] Diwekar UM, Xu W. Improved genetic algorithms for deterministic optimization and optimization under uncertainty. Part I. Algorithms development. *Ind Eng Chem Res* 2005;44:7132–7.
- [27] Ng LY, Chemmangattuvalappil NG, Dev VA, Eden MR. Chapter 1 – Mathematical Principles of Chemical Product Design and Strategies. *Comput Aided Chem Eng* 2017;39:3–43.
- [28] Frutiger J, Cignitti S, Abildskov J, Woodley JM, Sin G. Computational chemical product design problems under property uncertainties. vol. 40. Elsevier Masson SAS; 2017. doi:10.1016/B978-0-444-63965-3.50164-1.

- [29] Frutiger J, Bell I, O'Connell JP, Kroenlein K, Abildskov J, Sin G. Uncertainty assessment of equations of state with application to an organic Rankine cycle†. *Mol Phys* 2017;115:1225–44. doi:10.1080/00268976.2016.1275856.
- [30] Rothman JH, Kroemer RT. A concerted molecular design method by temperature-guided Monte Carlo search/potential applications for textitab initio ligand design. *J Mol Model* 1997;3:261–86. doi:10.1007/s008940050038.
- [31] Hu X, Beratan DN, Yang W. A gradient-directed Monte Carlo approach to molecular design. *J Chem Phys* 2008;129:064102. doi:10.1063/1.2958255.
- [32] Lu F. Molecular design of oil-based drilling fluid emulsifiers by Monte Carlo simulation. *Drill Fluid Complet Fluid* 2013;30.
- [33] Frutiger J, Andreasen JG, Liu W, Spliethoff H, Haglind F, Abildskov J, et al. Working fluid selection for organic Rankine cycles - impact of uncertainty of fluid properties. *Energy* 2016;109:987–97.
- [34] Frutiger J, Bell I, O'Connell JP, Kroenlein K, Abildskov J, Sin G. Uncertainty assessment of equations of state with application to an organic Rankine cycle. *Mol Phys* 2017;115:1225–44.
- [35] Frutiger J, Abildskov J, Sin G. Global sensitivity analysis of computer-aided molecular design problem for the development of novel working fluids for power cycles. vol. 38. 2016. doi:10.1016/B978-0-444-63428-3.50052-7.
- [36] Frutiger J, Zühlsdorf B, Abildskov J, Elmegaard B, Sin G. Reverse engineering of working fluid selection for industrial heat pump based on Monte Carlo sampling and uncertainty analysis. *Ind Eng Chem Res* 2018;submitted.
- [37] Frutiger J, Andreasen J, Liu W, Spliethoff H, Haglind F, Abildskov J, et al. Working fluid selection for organic Rankine cycles - Impact of uncertainty of fluid properties. *Energy* 2016;109. doi:10.1016/j.energy.2016.05.010.
- [38] Linke P, Papadopoulos A, Seferlis P. Systematic Methods for Working Fluid Selection and the Design, Integration and Control of Organic Rankine Cycles—A Review. *Energies* 2015;8:4755–801.
- [39] Papadopoulos AI, Stijepovic M, Linke P, Seferlis P, Voutetakis S. Toward Optimum Working Fluid Mixtures for Organic Rankine Cycles using Molecular Design and Sensitivity Analysis. *Ind Eng Chem Res* 2013;52:12116–33.
- [40] Pierotti GJ, Deal CH, Derr EL. Activity Coefficients and Molecular Structure. *Ind Eng Chem* 1959;51:95–102. doi:10.1021/ie50589a048.
- [41] Cignitti S, Andreasen JG, Haglind F, Woodley JM, Abildskov J. Integrated working fluid-thermodynamic cycle design of organic Rankine cycle power systems for waste heat recovery. *Appl Energy* 2017;203:442–53. doi:10.1016/j.apenergy.2017.06.031.
- [42] Zhang L, Cignitti S, Gani R. Generic mathematical programming formulation and solution for computer-aided molecular design. *Comput Chem Eng* 2015;78:79–84.
- [43] Odele O, Macchietto S. Computer aided molecular design: a novel method for optimal

- solvent selection. *Fluid Phase Equilib* 1993;82:47–54.
- [44] Marrero J, Gani R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilib* 2001;183–184:183–208.
- [45] Frutiger J, Marcarie C, Abildskov J, Sin G. A comprehensive methodology for development, parameter estimation, and uncertainty analysis of group contribution based property models – an application to heat of combustion. *J Chem Eng Data* 2016;61:602–13.
- [46] Huber PJ. Robust Estimation of a Location Parameter. *Ann Math Stat* 1964;35:73–101.
- [47] Seber G, Wild C. *Nonlinear Regression*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1989.
- [48] Sin G, Gernaey K V., Neumann MB, van Loosdrecht MCM, Gujer W. Global sensitivity analysis in wastewater treatment plant model applications: Prioritizing sources of uncertainty. *Water Res* 2011;45:639–51. doi:10.1016/j.watres.2010.08.025.
- [49] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab Eng Syst Saf* 2003;81:23–69.
- [50] Iman R, Conover W. A distribution-free approach to inducing rank correlation among input variables. *Commun Stat Part B-Simulation Comput* 1982;11:311–34.
- [51] Calafiore G, Campi MC. Uncertain convex programs: Randomized solutions and confidence levels. *Math Program* 2005;102:25–46.
- [52] Ben-Tal a., Nemirovski a. *Robust Convex Optimization*. *Math Oper Res* 1998;23:769–805. doi:10.1287/moor.23.4.769.
- [53] Cignitti S, Andreasen JG, Mansouri SS, Haglind F, Abildskov J, Woodley JM. Optimization-based Integrated Chemical Product-Process Design: Application to an industrial organic Rankine cycle unit for waste heat recovery. *Ind Eng Chem Res* 2017.
- [54] Cignitti S, Frutiger J, Zühlendorf B, Bühler F, Andreasen JG, Müller F, et al. Forbedring af industrielle processers energieffektivitet. *Dansk Kemi* 2016;97.
- [55] Andreasen JG, Larsen U, Knudsen T, Pierobon L, Haglind F. Selection and optimization of pure and mixed working fluids for low grade heat utilization using organic rankine cycles. *Energy* 2014;73:204–13.
- [56] Joback K, Reid R. Estimation of pure-component properties from group-contribution. *Chem Eng Commun* 1987;57:233–43.
- [57] Soave G. Equilibrium constants from a modified Redlich-Kwong equation of state. *Chem Eng Sci* 1972;27:1197–203.
- [58] Mondejar ME, Cignitti S, Abildskov J, Woodley JM, Haglind F. Prediction of properties of new halogenated olefins using two group contribution approaches. *Fluid Phase Equilib* 2017;433:79–96. doi:10.1016/j.fluid.2016.10.020.
- [59] MATLAB 2017b, The MathWorks, Inc., Natick, Massachusetts, United States n.d.
- [60] Sobol IM. On quasi-Monte Carlo integrations. *Math Comput Simul* 1998;47:103–12.

- [61] Bussieck MR, Meeraus A. General algebraic modeling system (GAMS). In: Kallrath J, editor. *Model. Lang. Math. Optim. Appl. Optim.*, Boston, MA.: Springer; 2004.
- [62] Sahinidis N. BARON: A general purpose global optimization software package. *J Glob Optim* 1996;8:201–5.
- [63] Misener R, Floudas CA. ANTIGONE: Algorithms for continuous / integer global optimization of nonlinear equations. *J Glob Optim* 2014;59:503–26. doi:10.1007/s10898-014-0166-2.

ACCEPTED MANUSCRIPT