Electronic Supplementary Material (ESI) for Molecular Systems Design & Engineering.

Supporting Information – Beyond heuristic analysis: Integration of process and working fluid design for organic Rankine cycles

David H. Bowskill[‡a], Uku Erik Tropp[‡a], Smitha Gopinath[a], George Jackson[a], Amparo Galindo[a] and Claire S. Adjiman[*a]

May 20, 2019

1 ORC process model equations

A summary of the core equations used in each case study are given below:

$$T_{cond,out} = T_1 \tag{1}$$

$$T_{evap,out} = T_3 \tag{2}$$

$$P_4 = P_{4s} = P_1 = P_{cond} \tag{3}$$

$$T_1 = T^{sat}(P_1) \tag{4}$$

$$\dot{Q}_{out} = \dot{m}(h^L(T_1,P_1) - (f_{vap}h^V(T_4,P_4) + (1-f_{vap})h^L(T_4,P_4))) \tag{5}$$

$$P_2 = P_3 = P_{evap} \tag{6}$$

$$T_3 = T^{sat}(P_3) + \Delta T_{extent} \tag{7}$$

$$\dot{Q}_{in} = \dot{m}(h^V(T_3,P_3) - h^L(T_2,P_2)) \tag{8}$$

IF $s^V(T_3,P_3) < s^V(T^{sat}(P_{4s}),P_{4s})$;

$$s^V(T_3,P_3) = f_{vap,s}s^V(T^{sat}(P_{4s}),P_{4s}) + (1-f_{vap,s})s^L(T^{sat}(P_{4s}),P_{4s}) \tag{9}$$

$$h^{V/L}_{out,s} = f_{vap,s}h^V(T^{sat}(P_{4s}),P_{4s}) + (1-f_{vap,s})h^L(T^{sat}(P_{4s}),P_{4s}) \tag{10}$$

ELSE;

$$f_{vap,s} = 1 \tag{11}$$

$$s^V(T_3,P_3) = s^V(T_{4s},P_{4s}) \tag{12}$$

$$h^{V/L}_{out,s} = h^V(T_{4s},P_{4s}) \tag{13}$$

END

$$\frac{h^{V/L}_{out} - h^V(T_3,P_3)}{h^{V/L}_{out,s} - h^V(T_3,P_3)} = \eta_t \tag{14}$$

[0a] *Department of Chemical Engineering, Centre for Process Systems Engineering,Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom*

[0*] Corresponding author

[0‡] These authors contributed equally to this work

IF $h_{out}^{V/L} < h^V(T^{sat}(P_{4s}), P_{4s})$;

$$h_{out}^{V/L} = f_{vap} h^V(T^{sat}(P_4), P_4) + (1 - f_{vap}) h^L(T^{sat}(P_4), P_4) \tag{15}$$

ELSE;

$$f_{vap} = 1 \tag{16}$$

$$h_{out}^{V/L} = h^V(T_4, P_4) \tag{17}$$

END

$$\dot{W}_{out} = \dot{m}(h^V(T_3, P_3) - (f_{vap} h^V(T_4, P_4) + (1 - f_{vap}) h^L(T_4, P_4)) \tag{18}$$

Additional equations for case studies 1 & 2

$$s^L(T_{2s}, P_{2s}) = s^L(T_1, P_1) \tag{19}$$

$$\frac{h^L(T_{2s}, P_{2s}) - h^L(T_1, P_1)}{h^L(T_2, P_2) - h^L(T_1, P_1)} = \eta_p \tag{20}$$

$$\dot{W}_{in} = \dot{m}(h^L(T_2, P_2) - h^L(P_1, T_1)) \tag{21}$$

$$P_{2s} = P_{2a} = P_{evap} \tag{22}$$

$$T_{2a} = T^{sat}(P_{2a}) \tag{23}$$

$$\dot{M}cp(T_{hs,in} - T_{hs,a}) = \dot{m}(h^V(T_3, P_3) - h^L(T_{2a}, P_{2a})) \tag{24}$$

$$\dot{M}cp(T_{hs,in} - T_{hs,out}) = \dot{m}(h^V(T_3, P_3) - h^L(T_2, P_2)) \tag{25}$$

Additional equations for case study 3

$$T_1 = T_2 \tag{26}$$

$$\dot{W}_{in} = \dot{m}\left(\frac{P_2 - P_1}{\rho(T_1, P_1)}\right) \tag{27}$$

$$\eta = \frac{|\dot{W}_{out} + \dot{W}_{in}|}{|\dot{Q}_{in}|} \tag{28}$$

1.1 ORC process constraints

$$T_{evap,out} \leqslant T^U \tag{29}$$

$$T_{cond,out} \geqslant T^L \tag{30}$$

$$T_{cond,out} \geqslant T_m + \Delta T_m \tag{31}$$

$$P_{evap} \geqslant P_{cond} \tag{32}$$

$$\frac{P_{evap}}{P_c} \leqslant P_r^U \tag{33}$$

$$P_{evap} \leqslant P^{U0} \tag{34}$$

$$\frac{P_{cond}}{P_c} \geqslant P_r^L \tag{35}$$

$$P_{cond} \geqslant P^{L0} \tag{36}$$

$$f_{vap} \geqslant f^L \tag{37}$$

Additional constraints for case studies 1 & 2

$$T_{hs,in} \geqslant T_{evap,out} + \Delta T_{min} \tag{38}$$

$$T_{hs,a} \geqslant T_{2a} + \Delta T_{min} \tag{39}$$

$$T_{hs,out} \geqslant T_2 + \Delta T_{min} \tag{40}$$

2 Chemical feasibility constraints

The following equations describe the chemical feasibility constraints:

$$0 \leqslant n_i; \qquad \forall i \tag{41}$$

$$\sum_i n_i - N^{ub} \leqslant 0 \tag{42}$$

$$n_{CH_3} + n_{COOH} + n_{CH_2OH} + n_{CH_2=} = 2 \tag{43}$$

$$n_{cO} \leqslant n_{CH_2} - \frac{n_{CH_2}}{N_{CH_2}^{ub}} \tag{44}$$

$$n_{eO} \leq n_{CH_3} \tag{45}$$

$$n_{eO} \leq 2n_{CH_2} \tag{46}$$

$$n_{eO} \leq 1 + n_{CH_2} - n_{cO} - \frac{n_{COO} + n_{CH=}}{N_a^{ub}} \tag{47}$$

$$n_{CH_2=} \leq 2 + n_{CH=} - \frac{n_{CH_2} + n_{COO} + n_{cO} + n_{eO}}{N_b^{ub}} \tag{48}$$

$$2n_{db} = n_{CH=} + n_{CH_2=} \tag{49}$$

Equation (42) sets the constraint on the maximum number of groups in a molecule. Since only linear molecules are considered, Equation (43) sets the bounds on the number of end groups with a valency of 1. Equations (44)-(47) define the the limits on the ether groups while Equations (48)-(49) ensure functional groups containing double bonds form a feasible molecule.

3 Critical pressure evaluation

Thermophysical property evaluation is carried out using the gSAFT functionality in gPROMS. This routine does not contain an implicitly built method for evaluating the critical properties of a fluid. To approximate the critical pressure and temperature of a fluid a dynamic model was introduced in gPROMS. Here the pressure is incremented over time as in Equation (50). Phase equilibrium between gaseous and liquid phases is ensured by Equation (51). Finally, convergence criteria is set when the ratio of gaseous and liquid volumes is equal to unity with some tolerance ε as in Equation (52). From numerical testing ε was set to 0.21. Once the convergence criteria is met the final pressure is reported as the critical pressure.

$$\frac{\partial P}{\partial t} = 1 \tag{50}$$

$$\mu^{gas}(T,P) = \mu^{liquid}(T,P) \tag{51}$$

$$\frac{v^{gas}(T,P)}{v^{liquid}(T,P)} \leqslant 1 + \varepsilon \tag{52}$$

Due to the difficulty of numerical calculations close to the critical point this routine can be numerically unstable and may fail in some cases. Fortunately due to recent advances in the DAEBDF numerical solver in gPROMS 5.1.1 the above method has been applied successfully for all molecules containing up to 12 functional groups (i.e., 3,175 chemical combinations).

4 Feasibility test gradients

4.1 Test 0

Tests 0 & 1 of the algorithm are used to detect infeasible solutions to the algorithm and pass linearised constraints to the master problem. In Test 0 the two constraints are formulated as follows:

$$\begin{aligned} T_{\boldsymbol{n}}^{L} - T^{U} &\leqslant 0 \\ P_{\boldsymbol{n}}^{L0} - P_{\boldsymbol{n}}^{U0} &\leqslant 0 \end{aligned} \tag{T0}$$

Where:

$$P_{\boldsymbol{n}}^{L0} = \max\{P^{L0}, P_r^L P_c(\boldsymbol{n})\}, \tag{53}$$

$$P_{\boldsymbol{n}}^{U0} = \min\{P^{U0}, P_r^U P_c(\boldsymbol{n})\}. \tag{54}$$

$$T_{\boldsymbol{n}}^{L0} = \max\{T^{L}, T_m(\boldsymbol{n}) + \Delta T_m\} \tag{55}$$

For Test 0, it is highly unlikely (and in fact does not occur in our study) that the constraint on pressure will be violated in Equations (T0) assuming that pressure bounds are chosen sensibly. The temperature bound may however be violated if the working “fluid" is solid over the whole operating region. For this it is desirable to add additional constraints to the master problem to prevent generation of these solutions.

If we formulate the constraint on melting temperature from Equations (T0) and (55) as:

$$(T_m + \Delta T_m) - T^{U} \leqslant 0 \tag{56}$$

This constraint is equivalent to:

$$\exp\left(\frac{T_m}{T_{m,0}}\right) - \exp\left(\frac{T^U - \Delta T_m}{T_{m,0}}\right) \leqslant 0 \tag{57}$$

Finally using the correlation by Hukkerikar *et al.* [1]:

$$\sum_{i=1}^{q}(a_i n_i) - \exp\left(\frac{T^U - \Delta T_m}{T_{m,0}}\right) \leqslant 0 \tag{58}$$

Equation (58) is linear in n_i and can be included directly into the master problem. Therefore, the tests will only eliminate molecules that are solid over the whole feasible region in the first iteration, while Equation (58) will prevent these working fluids from being generated. As a result, we do not calculate gradients of any constraints from Test 0 during the algorithm.

4.2 Test 1

In Test 1, the tests on the saturation behaviour of the fluid are enforced by Equations (T1):

$$\begin{aligned} T^{sat}(P_{\boldsymbol{n}}^{L0}, \boldsymbol{n}) - T^{U} &\leqslant 0 \\ T_{\boldsymbol{n}}^{L0} - T^{sat}(P_{\boldsymbol{n}}^{U0}, \boldsymbol{n}) &\leqslant 0 \end{aligned} \tag{T1}$$

With the upper and lower bounds on pressure set as:

$$P_{\boldsymbol{n}}^{L0} = \max\{P^{L0}, P_r^L P_c(\boldsymbol{n})\}, \tag{59}$$

$$P_{\boldsymbol{n}}^{U0} = \min\{P^{U0}, P_r^U P_c(\boldsymbol{n})\}. \tag{60}$$

Test 1 can then be partitioned into four equations shown below:

$$T^{sat}(P^{L}, \boldsymbol{n}) - T^{U} \leqslant 0 \tag{61}$$

$$T^{sat}(P_r^L P_c(\boldsymbol{n}), \boldsymbol{n}) - T^{U} \leqslant 0 \tag{62}$$

$$T_{\boldsymbol{n}}^{L0} - T^{sat}(P^{U}, \boldsymbol{n}) \leqslant 0 \tag{63}$$

$$T_{\boldsymbol{n}}^{L0} - T^{sat}(P_r^U P_c(\boldsymbol{n}), \boldsymbol{n}) \leqslant 0 \tag{64}$$

All these constraints must be satisfied for a fluid to be feasible. By far the most common form of infeasibility in our design space is due to Equation (61) in that most fluids are not volatile enough to produce a vapour phase within process operating bounds. Should Equation (61) be violated, then gradients of the equation are calculated using forward finite differences with respect to each element in the vector $\boldsymbol{n}$. These finite differences do not take integer values, instead a relaxation is enforced on $\boldsymbol{n}$ and small perturbations are taken. Care is particularly taken for finite differences with respected both associating groups n_{COOH} and n_{CH2OH} such that the perturbations are appropriately small to mitigate numerical issues.

Equations (62) and (64) both have an additional indirect dependence on $\boldsymbol{n}$ through $P_c(\boldsymbol{n})$. Although gradients could be taken with respect to these equations computational times would likely increase. Further, as these constraints are rarely violated gradients are not taken with respect to those equations.

Finally, Equation (63) is not defined for cases when $P^U \geqslant P_c(\boldsymbol{n})$ and can cause numerical issues. For this reason gradients with respect to Equation (63) are also not calculated.

Therefore only linearised constraints of Equation (61) are added to the master problem should a molecule violate this constraint. Regardless a vast proportion of the failures are due to this constraint and little effectiveness is lost by the algorithm through omission of the other constraints. If a molecule should violate any of Equations (62), (63) or (64) and not (61) then it will be removed from the search using an integer cut[2].

5 Primal Problem formulation

The algorithm is initiated from a starting point represented by discrete variables $\boldsymbol{n}^k$, with the iteration counter set to $k = 0$. After obtaining the pressure bounds from Test 2, the primal problem is optimised to find the best possible performance of the working fluid $\boldsymbol{n}^k$. The continuous variable vector $\boldsymbol{x}$ is partitioned such that $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{u},\boldsymbol{x'})$, where $\boldsymbol{u}$ is a vector of independent continuous variables. The remaining variables can be calculated using equality constraints $\boldsymbol{h}(\boldsymbol{x},\boldsymbol{n}^k) = \boldsymbol{0}$ such that $\boldsymbol{x'} = \boldsymbol{x'}(\boldsymbol{u},\boldsymbol{n}^k)$. Therefore, the primal problem is optimised with respect to continuous variables $\boldsymbol{u}$.

$$\begin{aligned} f^k = \min_{\boldsymbol{u}} \; & f(\boldsymbol{u},\boldsymbol{n}^k) \\ \text{s.t.} \quad & \boldsymbol{g}(\boldsymbol{u},\, \boldsymbol{x'}(\boldsymbol{u},\, \boldsymbol{n}^k),\, \boldsymbol{n}^k) \leqslant 0 \\ & \boldsymbol{x'}^L \leqslant \boldsymbol{x'}(\boldsymbol{u},\, \boldsymbol{n}^k) \leqslant \boldsymbol{x'}^U \\ & \boldsymbol{u}^L \leqslant \boldsymbol{u} \leqslant \boldsymbol{u}^U \end{aligned} \tag{P}$$

A value of the objective function f^k is found at each iteration k. The upper bound on the objective function for the problem (P1) is then updated such that $f^{UB} = \min\{f^{UB}, f^k\}$.

5.1 Primal problem initialisation

To successfully implement the primal problem, proper initialisation is needed for each new working fluid to ensure convergence. This is because poor initialisation of process variables far from any primal problem solution makes the optimiser prone to failure even when the molecule itself is feasible. The calculation of equilibrium thermodynamic properties such as entropy and enthalpy using the SAFT-γ Mie equation of state can be particularly problematic when temperature and pressure are initialised far from a feasible solution.

To improve the robustness of state calculations using temperature and pressure, P_{cond} and P_{evap} are initialised close to the lower bound on pressure and all associated temperatures are initialised at at their corresponding saturation points. To be precise, $P_{cond} = P^{L0}_{\boldsymbol{n}}$, $P_{evap} = 1.1 \times P^{L0}_{\boldsymbol{n}}$, $T_4 = T_{4s} = T_1 = T^{sat}(P_{cond},\boldsymbol{n})$ and $T_2 = T_{2s} = T_3 = T^{sat}(P_{evap},\boldsymbol{n})$. All other variables are set arbitrary values. This initialises all four states of the cycle sufficiently close to a solution of the equations $\boldsymbol{h}(\boldsymbol{x},\boldsymbol{n}) = \boldsymbol{0}$. Through rigorous testing, it was found that initialisation of all four states in this way leads to solver convergence for all feasible molecules. Further, even for infeasible molecules, in all cases tested, the primal was successfully initialised even though a solution is impossible.

Numerically, this type of initialisation is enforced by an additional model in gPROMS ModelBuilder 5.1.1.. Saved variable sets from this model are then transferred to the primal problem to ensure its solution.

5.2 Primal problem gradients

After solving the primal problem, the objective function and constraints are linearised with respect to continuous and relaxed discrete variables at the optimal point. The gradients are added to the master problem where a new molecule is generated. If the primal problem is initiated but no solution can be found, then the objective function and linearised constraints are taken at the point of initialisation.

6 Master Problem formulation

The master problem is used to generate a new molecule for evaluation in the feasibility tests and primal problem. The master problem follows the original formulation used by Duran and Grossmann[3] as well as Fletcher and Leyffer[4] with modifications applied specifically to aid in ORC working fluid design. In this proposed algorithm, the master problem is modified to implement an augmented penalty framework with slack variables e_1 and e_2[5] as well as heuristics to guide working fluid selection in the case of degeneracy.

The full formulation of the master problem is summarised by problem (M). Since each iteration may give different outcomes (such as passing or failing the feasibility tests), various sets are defined to keep track of linearised constraints that have come up across all iterations. The set $F^{(k)}$ is used to save the iteration index l when the primal problem is feasible. For the indices $l \in F^{(k)}$, the set $A^{(k)}$ saves the indices (l,m), where m denotes the index of an active constraint $\boldsymbol{g}_2$ in the solution of the primal problem. If the a molecule fails Test 1, the information about all the active and violated constraints are saved in the set $A1^{(k)}$. $A1^{(k)}$ contains the pairs of indices (l,j), where l is the iteration number and j is the index of the active or violated constraint $\boldsymbol{g}_1$ in Test 1. Finally, the indices of all molecules removed by an integer cut[2] are stored in the set $IC^{(k)}$ when a molecule has been generated for a second time by the algorithm, or $PIC^{(k)}$ when a permanent integer is applied at the end of an outer iteration.

$$
\begin{aligned}
\min_{\boldsymbol{u},\boldsymbol{n},\eta}\ & \eta + 0.001(c_p/205.86 - \Delta h/44.47) + 1000(e_1+e_2) \\
\text{s.t. } & f(\boldsymbol{u}^{(l)},\boldsymbol{n}^{(l)}) + \nabla_{\boldsymbol{n}}^T f(\boldsymbol{u}^{(l)},\boldsymbol{n}^{(l)})[\boldsymbol{n}-\boldsymbol{n}^{(l)}] + \nabla_{\boldsymbol{u}}^T f(\boldsymbol{u}^{(l)},\boldsymbol{n}^{(l)})[\boldsymbol{u}-\boldsymbol{u}^{(l)}] \leqslant \eta + e_1,\ \ \forall\ l \in F^{(k)} \\
& g_{2,m}(\boldsymbol{u}^{(l)},\boldsymbol{n}^{(l)}) + \nabla_{\boldsymbol{n}}^T g_{2,m}(\boldsymbol{u}^{(l)},\boldsymbol{n}^{(l)})[\boldsymbol{n}-\boldsymbol{n}^{(l)}] + \nabla_{\boldsymbol{u}}^T g_{2,m}(\boldsymbol{u}^{(l)},\boldsymbol{n}^{(l)})[\boldsymbol{u}-\boldsymbol{u}^{(l)}] \leqslant e_2,\ \ \forall\ (l,m) \in A^{(k)} \\
& g_{1,j}(\boldsymbol{u}^{(l)},\boldsymbol{n}^{(l)}) + \nabla_{\boldsymbol{n}}^T g_{1,j}(\boldsymbol{u}^{(l)},\boldsymbol{n}^{(l)})[\boldsymbol{n}-\boldsymbol{n}^{(l)}] \leqslant e_2,\ \ \forall\ (l,j) \in A1^{(k)} \\
& e_1, e_2 \geqslant 0 \\
& \boldsymbol{u}^L \leqslant \boldsymbol{u} \leqslant \boldsymbol{u}^U \\
& \boldsymbol{Cn} \leqslant \boldsymbol{d} \\
& \eta \leqslant f^{UB} \\
& M^L(1-y_l) + \varepsilon_c \leqslant \sum_i \left(\ln(p_i)(n_i - n_i^{(l)}) \right) \leqslant M^U y_l - \varepsilon_c,\ \ \forall\ l \in IC^{(k)} \\
& y_l \in \{0,1\},\ \ \forall\ l \in IC^{(k)} \\
& M^L(1-y_l) + \varepsilon_c \leqslant \sum_i \left(\ln(p_i)(n_i - n_i^{(l)}) \right) \leqslant M^U y_l - \varepsilon_c,\ \ \forall\ l \in PIC^{(k)} \\
& y_l \in \{0,1\},\ \ \forall\ l \in PIC^{(k)} \\
& \sum_{i=1}^{q} (a_i n_i) - \exp\left(\frac{T^U - \Delta T_m}{T_{m,0}} \right) \leqslant 0
\end{aligned}
\tag{M}
$$

As can be seen the modified constraint on melting point given in Section 4.1 of this supporting information is included in the master problem due to its linearity with $\boldsymbol{n}$. This ensures that any new working fluids generated by the algorithm won't be solid over the entire operating domain.

When no primal problems have been solved, the value of η in the master problem takes the same value (an arbitrary lower bound) for all master-problem feasible values of $\boldsymbol{n}$. Given the vast search space, the master problem is very highly degenerate. The MILP master problem solver usually yields the first-found feasible solution, which in turn depends on the heuristics of the MILP solver. When the master problem is degenerate, the solution yielded by the MILP solver may or may not be a good solution to the primal problem. We modified the objective to provide more information to the master problem when degenerate[6]. To the conventional master problem objective, a new term is added which in effect is a novel primal-heuristic[7]. Some sources disagree on appropriate physico-chemical properties and their correlations with working fluid performance, however here we arbitrarily choose the minimisation of ideal gas heat capacity c_p (which typically correlates strongly with liquid heat capacity) and the maximising of latent heat of vapourisation Δh of the working fluid. The linear group contribution parameters proposed by Joback and Reid[8] are used for the calculation of these properties and both are scaled appropriately. The heuristics appear in the objective function of the master problem with a weight of 0.001 such that the heuristics only have an appreciable affect on the algorithm when no objective function linearisations have been formulated.

Without the use of slack variables it was initially discovered that the master problem would become infeasible (i.e., $\eta \leqslant f^{UB}$ would be violated) and hence the algorithm would stop after approximately 2 primal evaluations. The use of slack variables in an augmented penalty framework facilitates the additional exploration of the search space after which a traditional OA will have converged. The slack variables are included in the objective function of the master problem with a coefficient of 1000 to strongly favour having $e_1, e_2 = 0$ until a traditional OA set-up would become infeasible as the $\eta \leqslant f^{UB}$ would no longer apply. This is necessary due to the high degree of non-convexity in the problem, and otherwise the algorithm may converge without an in depth exploration of the search space.

6.1 Stopping Criteria

The master problem and the primal problem in conjunction with the feasibility tests are solved successively until either of the stopping criteria are met. The arguments of the primal problem, $\boldsymbol{u}$ and $\boldsymbol{n}$, coinciding with the best objective function f^{UB} from all iterations are reported as the solution to the optimisation.

The traditional OA algorithm would terminate once the constraint $\eta \leqslant f^{UB}$ can no longer be satisfied. With the aid of the augmented penalty framework, the algorithm is terminated after three attempts to solve the primal problem are made once either of the slack variables become non-zero. This ensures that a sufficient amount of the search space is explored without a great deal of computational expense. To prevent divergent behaviour in the algorithm, a second stopping criterion is imposed in tandem such that the algorithm is stopped when 20 unique molecules have been analysed either in the feasibility tests or primal problem.

This stopping criteria is applied separately in each outer iteration to generate a new candidate molecule.

7 Physical Properties

Physical property analysis of all top performing working fluids in comparison to a reference set containing all feasible molecules in each case study are shown on Figures 1, 2 and 3. The whiskers show the maximum and minimum values of the reference set for each property, while the box shows the first and third quartile ranges as well as the median. The top ten working fluids for each case study are marked explicitly on the Figures with diamonds. In addition the top performing working fluid for each study id marked with a filled diamond; these correspond to propane, n-butane and pent-1,4-diene for case studies 1, 2 and 3, respectively.

As discussed in the main paper, trends can be seen indicate desirable properties for working fluids. These trends can be seen to be more significant particularly for case study 2. By considering the trend shown by the molecular weight indicator for case study 2 we believe these trends are somewhat artificial due to the inclusion of large molecular weight working fluids completely unsuitable for practical ORC design.

8 Optimal working fluid operating conditions

Optimal working fluid operating conditions for all feasible molecules in each case study can be found in the excel sheet associated with this document – Supporting_information_Data.xlsx

9 Starting Points

Table 1 outlines the 20 starting molecules used to with the algorithm. Each of the molecules is represented with a molecular code where each number signifies the occurrences of a functional group. The groups are given in the following order: -CH_3, -CH_2-, -eO-, -cO-, -COO, -CH=, -COOH, -CH_2OH and =CH_2. For example, a molecule with a code 2-1-0-0-0-0-0-0-0 corresponds to propane as it contains two CH_3-groups and one -CH_2-group.

The test results of each of the starting points are denoted as follows: 1 - tests passed, 2 - tests failed due to saturation constraints, 3 - tests failed due to the melting point constraint.

The results of each starting point can be found in the associated excel document - Supporting_information_Data.xlsx

Table 1 Molecular starting points and test results

#	Molecular code	CS1	CS2	CS3
1	0-0-0-0-0-4-0-0-2	2	1	2
2	2-0-0-0-0-0-0-0-0	1	1	2
3	2-1-0-0-0-0-0-0-0	1	1	1
4	2-2-0-0-0-0-0-0-0	1	1	1
5	2-3-0-0-0-0-0-0-0	1	1	1
6	1-0-0-0-0-1-0-0-1	1	1	1
7	2-2-0-1-0-0-0-0-0	1	1	1
8	0-3-0-0-0-2-0-0-2	1	1	2
9	1-1-0-0-0-3-0-0-1	1	1	2
10	0-0-0-0-5-3-0-1-1	2	2	2
11	0-2-0-1-1-2-0-2-0	2	2	2
12	0-3-0-0-0-3-1-0-1	2	2	2
13	0-3-0-0-4-1-1-0-1	2	2	3
14	1-1-1-0-6-0-1-0-0	3	2	3
15	0-0-0-0-4-2-1-1-0	3	2	3
16	1-0-0-0-0-0-1-0-0	1	1	2
17	2-1-2-0-0-0-0-0-0	1	1	1
18	0-2-0-0-0-0-1-1-0	2	2	2
19	2-2-0-1-1-0-0-0-0	2	2	2
20	2-5-2-0-1-0-0-0-0	2	2	2

References

[1] A. S. Hukkerikar, B. Sarup, A. Ten Kate, J. Abildskov, G. Sin and R. Gani, *Fluid Phase Equilibria*, 2012, **321**, 25–43.

[2] A. P. Samudra and N. V. Sahinidis, *AIChE Journal*, 2013, **59**, 3686–3701.

[3] M. A. Duran and I. E. Grossmann, *Mathematical programming*, 1986, **36**, 307–339.

[4] R. Fletcher and S. Leyffer, *Mathematical programming*, 1994, **66**, 327–349.

[5] J. Viswanathan and I. E. Grossmann, *Computers & Chemical Engineering*, 1990, **14**, 769–782.

[6] S. Gopinath, G. Jackson, A. Galindo and C. S. Adjiman, *AIChE Journal*, 2016, **62**, 3484–3504.

[7] T. Berthold, *Journal of Global Optimization*, 2018, **70**, 189–206.

[8] K. G. Joback and R. C. Reid, *Chemical Engineering Communications*, 1987, **57**, 233–243.

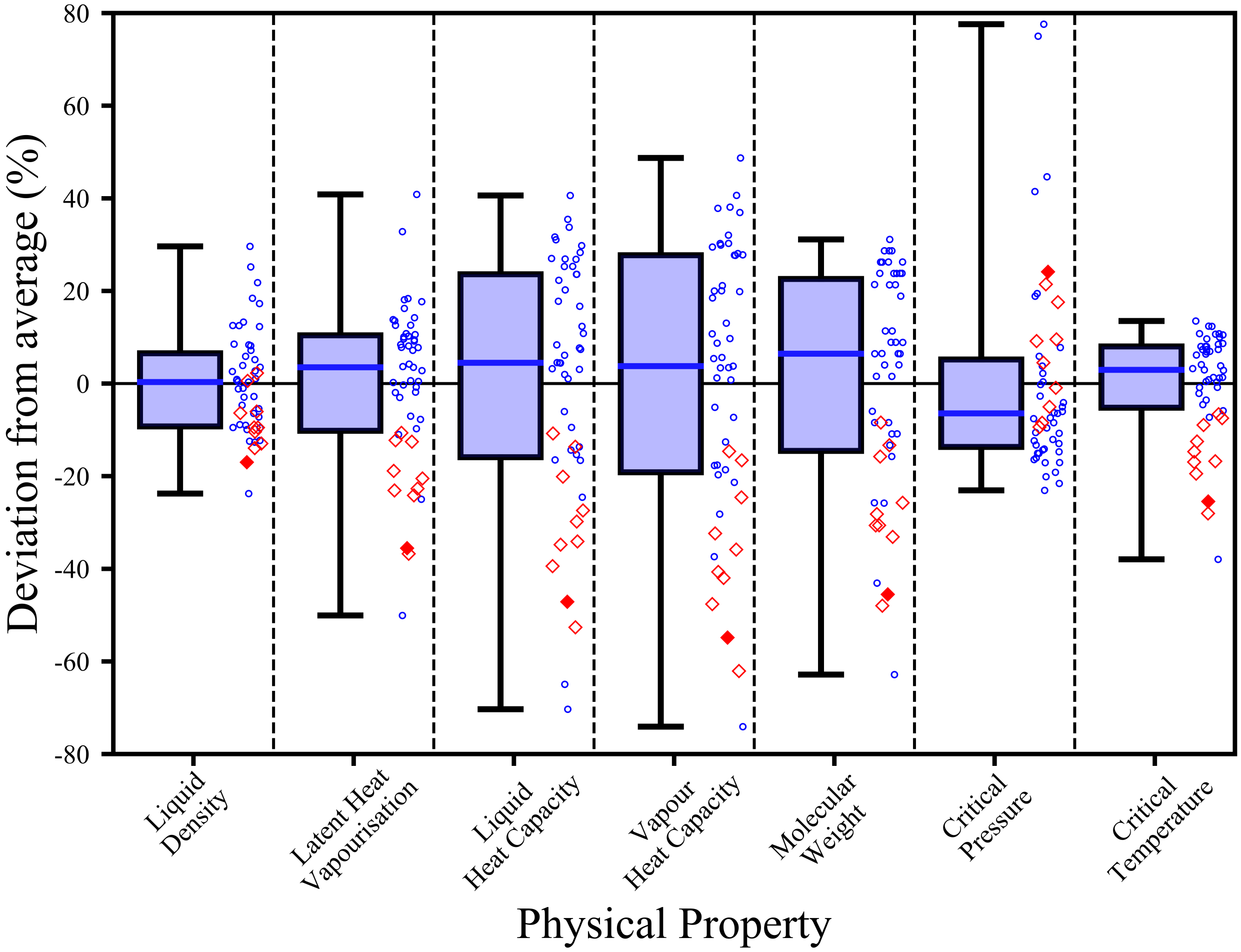

Figure 1 A box plot showing the maximum, minimum, first and third quartiles and median percentage deviations of working fluid properties from the mean of the reference set (cf. text) for case study 1. The data used to construct each box plot is shown on the right of each plot: the optimal working fluid (propane) is marked with a filled diamond, the remaining 9 best working fluids are marked with hollow diamonds and all other working fluids are marked with hollow circles.

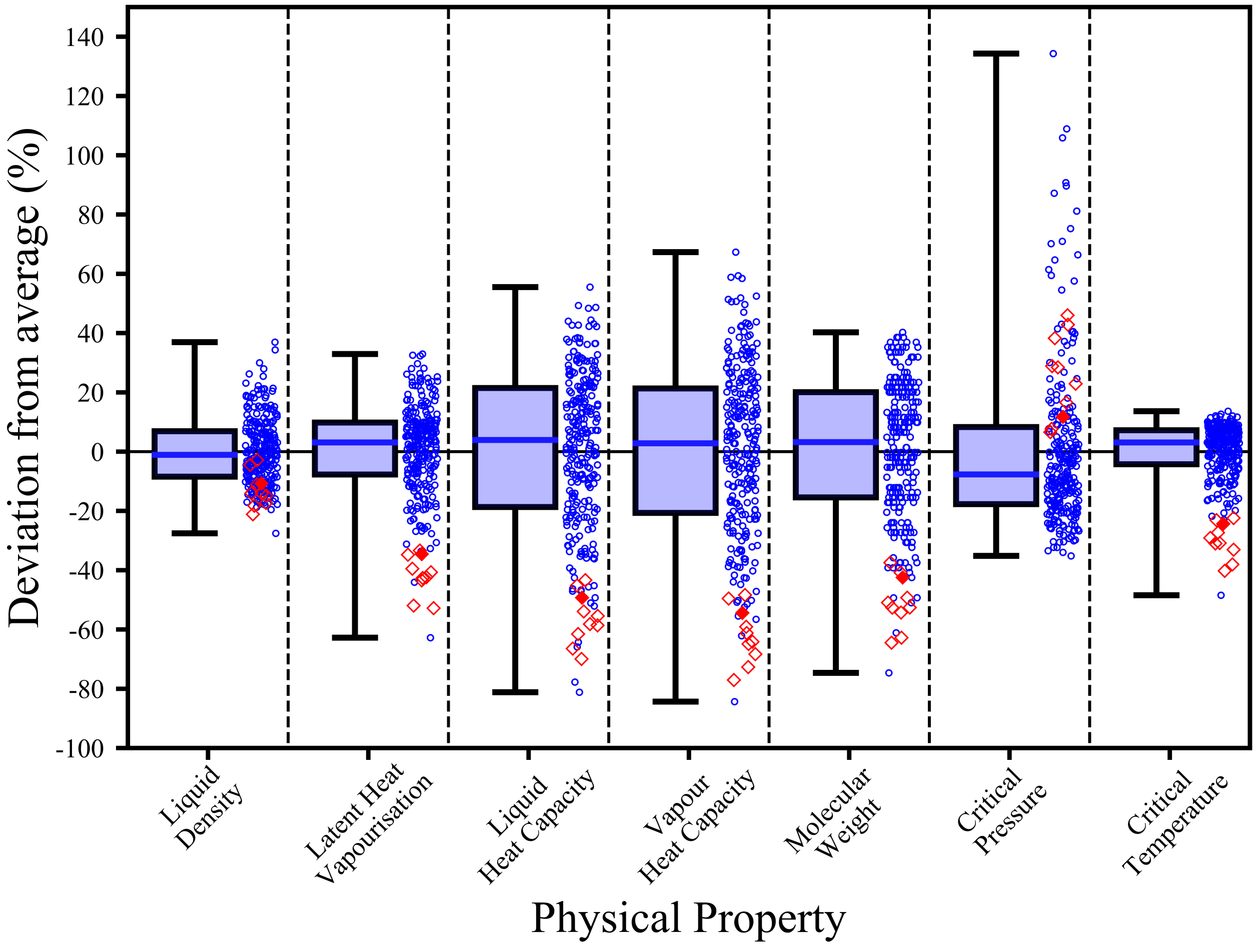

Figure 2 A box plot showing the maximum, minimum, first and third quartiles and median percentage deviations of working fluid properties from the mean of the reference set (cf. text) for case study 2. The data used to construct each box plot is shown on the right of each plot: the optimal working fluid (n-butane) is marked with a filled diamond, the remaining 9 best working fluids are marked with hollow diamonds and all other working fluids are marked with hollow circles.

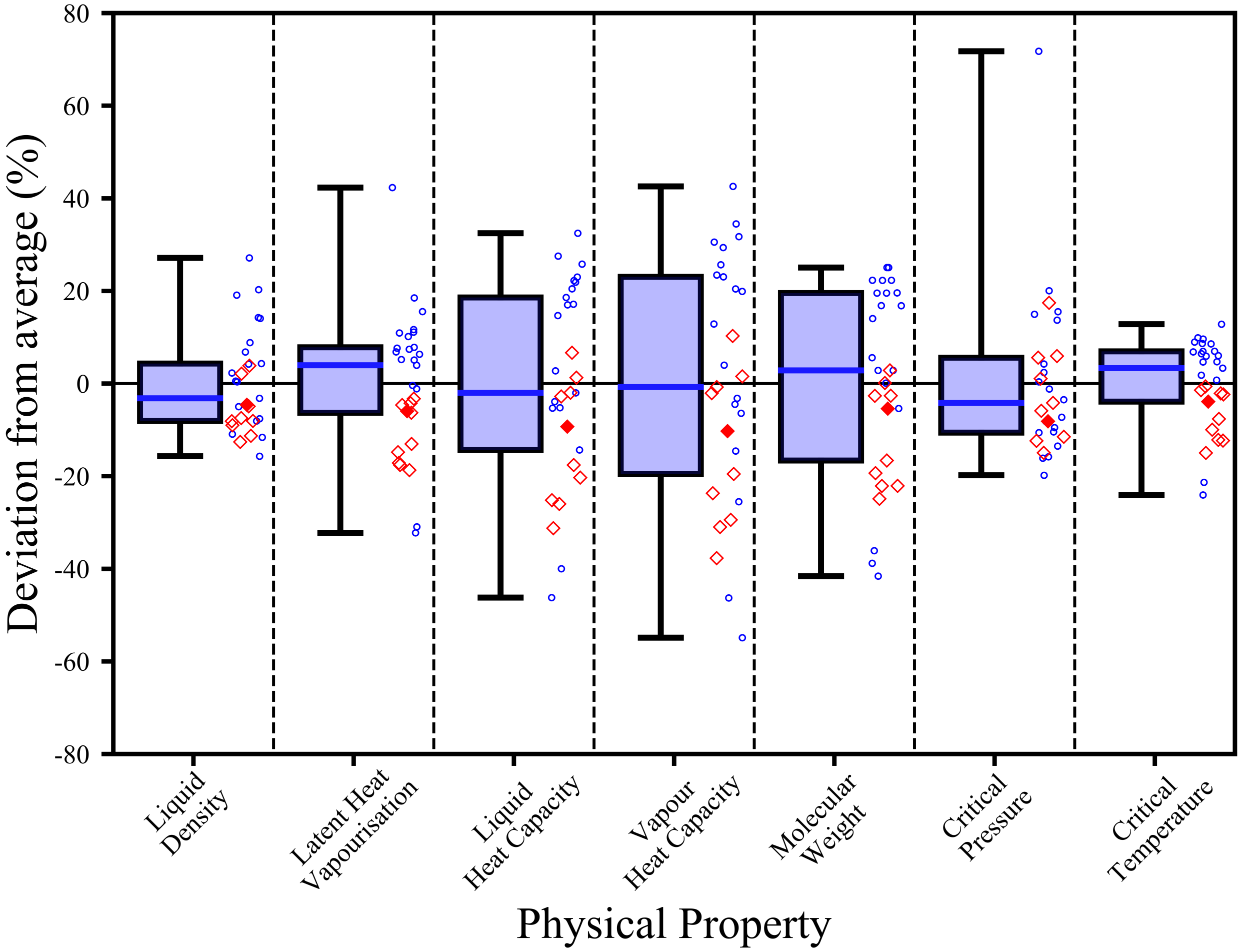

Figure 3 A box plot showing the maximum, minimum, first and third quartiles and median percentage deviations of working fluid properties from the mean of the reference set (cf. text) for case study 3. The data used to construct each box plot is shown on the right of each plot: the optimal working fluid (pent-1,4-diene) is marked with a filled diamond, the remaining 9 best working fluids are marked with hollow diamonds and all other working fluids are marked with hollow circles.