



Models driven conception of a Computer Aided Mixture Design tool

Juliette Heintz, Vincent Gerbaud, Jean-Pierre Belaud

► To cite this version:

Juliette Heintz, Vincent Gerbaud, Jean-Pierre Belaud. Models driven conception of a Computer Aided Mixture Design tool. ESCAPE 21, 21th European Symposium on Computer-Aided Process Engineering (ESCAPE 21), May 2011, Chalcidique, Greece. pp.562-566. <hal-04087644>

HAL Id: hal-04087644

<https://hal.science/hal-04087644v1>

Submitted on 3 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <http://oatao.univ-toulouse.fr/>
Eprints ID: 4889

To cite this document:

Heintz, Juliette and Gerbaud, Vincent and Belaud, Jean Pierre *Models driven conception of a Computer Aided Mixture Design tool*. (2011) In: 21th European Symposium on Computer-Aided Process Engineering (ESCAPE 21), 29 may - 01 June 2011, Chalkidiki, Grece.

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@inp-toulouse.fr

Models driven conception of a Computer Aided Mixture Design tool

Juliette Heintz ^{a,b}, Vincent Gerbaud ^{a,b}, Jean-Pierre Belaud ^{a,b}

^a Université de Toulouse, INP, UPS, LGC (Laboratoire de Génie Chimique), 4 allée Emile Monso, F-31030 Toulouse Cedex 04 – France. Vincent.gerbaud@ensiacet.fr

^b CNRS, LGC (Laboratoire de Génie Chimique), F-31030 Toulouse Cedex 04– France

Abstract

Within the “sustainable development for chemical industry” french national program, a computer aided mixture design software prototype is being developed. This paper presents the first version of our application, highlighting the different system views (functional, structural and behavioural). The parameters have been chosen so that the software offers a large panel of solutions for mixture design problems. The molecular representation is done by molecular graphs and an adapted genetic algorithm is used coupled with a multi level resolution. The performance of the software is then validated.

Keywords: Mixture design, CAMD, model driven development

1. Introduction

Within the REACH context, finding new molecules that are both environmentally and health friendlier has become a major issue for the chemical industry. To cope with the challenging size of the problem, “trial and error” usual methods must be replaced by more efficient ones, like Computer Aided Molecular Design (CAMD) that can handle simultaneously either functional, economical, health, safety or life cycle constraints.

Our work, performed within the framework of the InBioSynSolv project (French National Research Agency project ANR-09-CP2D-08), aims at developing a software prototype based on the principles of inverse formulation [1]. The main objective is the CAMD of chemical product mixtures that satisfy predefined properties [2] and come from renewable resources. Foreseeing future changes due to more accurate property models, such as toxicity assessment models, a model driven conception is used.

2. Software conception method

The software conception is based on a methodological approach that assures the alignment, as defined by Henderson [3], of the different models of the system. Our design is compliant with a model driven software engineering approach. It is based on different analysis models (UML and BPMN [4]) and object-oriented implementation's models dedicated to Microsoft.NET framework (C# and VB.NET). It formalizes the views of our system: the functional view (use case diagrams), the structural view (class diagrams), the behavioural view (user data sequencing, research sequencing) and the architectural view (component diagrams). This methodology makes the software easy to complement in the future: new properties model can be made accessible, new problems can be devised and stored as default one, new algorithms can be plugged...

Most of the model's alignment is taken care of by the VisualStudio 2010 Architecture and Modeling tool, the remaining is done by hand.

3. Functional and structural view

The software gives the user many problem design capacities. Indeed, problems, properties and models can be created and each problem's parameter can be edited, as shown on the use case diagram (fig. 1), where the relationships between use cases are «include».

3.1. Problem Parameters

Problem parameters are structured in several classes (fig. 2). Three main groups can be observed: “the algorithm parameters”, “the objective function parameters” and “the mixture candidates’ optimization parameters”:

The algorithm parameters for the genetic algorithm, like elitism, are defined for the problem. Following a multilevel approach [5,6], the objective function is a function of the calculable properties for the relevant level. Real properties, like “soluble”, are readily understood but not directly calculable. Thus, each real property is associated to at least one calculable property, like “solubility fraction”, and each calculable property is associated to a model of calculation and to the target value or range defined by the user.

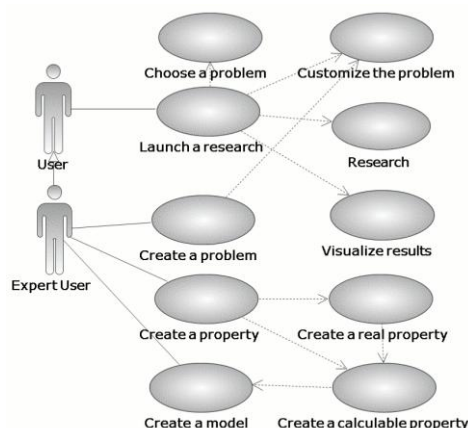


Fig.1. Software's functionalities (use case diagram)

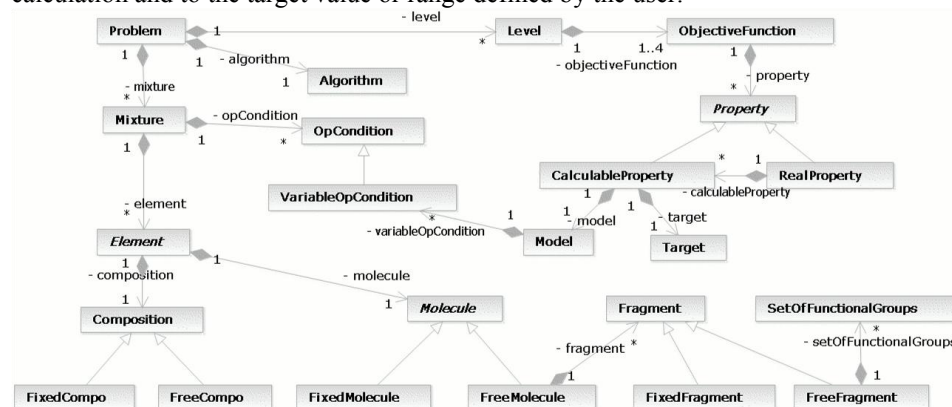


Fig. 2. Problem's data structure (class diagram)

As shown by figure 2, a mixture is defined by its molecules, its composition vector and property calculation conditions. Each can be either fixed or variable. For the molecules, the user can either fix a molecule or build it fragment by fragment (free molecule). Those fragments can then be either fixed or built from functional groups, as it is currently done in usual CAMD tools, or from renewable resources chemical blocks.

3.2. Molecular Representation

The molecular representation is based on the method proposed by Korichi and al. [5]: Molecular graphs are used. Each diagonal elements contains a functional group coded as an integer identifier with indications on its valence, on its integration in a ring or not and on its number of hydrogen. Non diagonal integers represent the bonding type (0: no bond, 1: a single bond, 2: a double bond, 3: a triple bond).

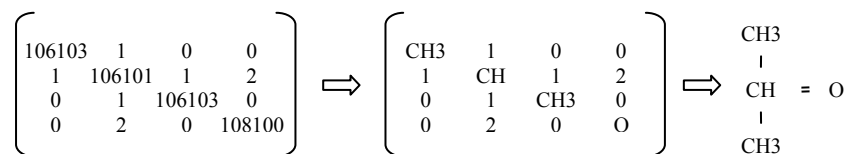


Fig. 3. Molecular graph interpretation

All molecules and fragments in the software are encoded this way. The molecular graph of a given free molecule is the aggregation of its fragment graphs. Unlike from vector-like representation of molecules, either using chemical groups like those used in group contribution methods [7] or atomic signature [8] where molecules are generated after the search of suitable candidates for the CAMD problem, it describes explicitly the bonds, in a similar way than other representations based on adjacency matrices [1].

4. Behavioural view

4.1. Algorithm

In order to handle the combinatorial explosion of the CAMD, the chosen methodology is based on a stochastic method coupled to a multi level resolution [6] (figure 4). The population of candidates consists of mixtures that are evaluated according to an objective function, selected and then modified with the common operators of the genetic algorithm method. The multi level resolution helps to reduce the calculation time. Starting with a large population of candidates and a simple objective function, it reduces the population size at each level shift while complexifying the objective function.

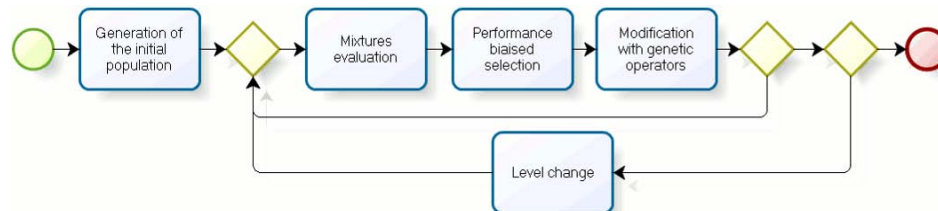


Fig. 4. Multi-level genetic algorithm (sequencing diagram)

4.2. Genetic operators for mixture generations

The genetic operators are applied only on the free fragments of the molecular graph:

Crossover: two graphs are splitted; parts are switched and merged into two new graphs.

Mutation: one functional group is replaced by another with the same connections.

Insertion: a single functional group is inserted into the fragment. If the new functional group has supplementary connections, branches are created by adding other groups.

Deletion: a functional group is deleted which may cause the deletion of whole branches.

4.3. Mixture's performance evaluation

The mixtures are evaluated by the objective function. It is computed as the user-weighted average of the mixture performance for each property calculation model. Performance is obtained by comparing the model value and the target fixed by the user.

$$\text{Objective Function} = \frac{\sum w_i \text{Performance}(\text{modelValue}, \text{target})}{\sum w_i}$$

Each property calculation model is evaluated by a plugged-in component ClassMolPropEstim. Within this component, the properties can be calculated with different methods based on functional groups decomposition procedure [9,10].

Figure 5 displays the interface between the application and the component ClassMolPropEstim. The software passes information of the mixture to evaluate to the component ClassMolPropEstim: molecular graphs (GmList), composition and operating

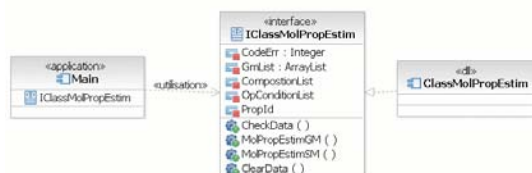


Fig. 5. CAMD tool interoperability (component diagram)

conditions. Then, for each property, the PropId model identifier is set according to the user's choice and level, and the method MolPropEstimGM returns a property value. Finally the mixture information is cleared in order to start over with the next mixture.

5. First results and analysis

The first prototype of the software is able to deal with a mixture where free molecules are built from a single free fragment. The objective function contains only calculable properties with no operating conditions required. The software has been tested with a very simple problem in order to compare the result to the DIPPR data base.

5.1. Test with a $T_b=358.15\text{K}$ target

The mixture consists of one element. The molecule is free and built with one free fragment. It is acyclic and it can be constructed with up to 10 functional groups: CH_3 -, NH_2 -, OH -, Cl -, F -, Br -, I -, $-\text{CH}_2$ -, $-\text{CH}$ -, $-\text{C}$ -, $\text{CH}_2=$ -, $-\text{CH}=$ -, $-\text{C}=$ -, $\text{CH}\equiv$ -, $-\text{C}\equiv$ -. The objective function is composed only of the boiling point temperature calculated with Marrero and Gani's first group model (MG). The target value is 358.15K. The performance is calculated with a Gaussian function where $\alpha=0.001$ and the tolerance is 2. The algorithm consists of 30 iterations on one level, the population size is 100 and the elitism is 20. The genetic operator choice probabilities are 50% for cross-over, 20% for mutation, 10% for insertion and 20% for deletion. The ten best molecules found are:

Table 1. Ten best acyclic molecules matching a 358.15K boiling point target

	molecule (SMILES)	T_b (MG)	T_b exp	Performance
1	<chem>C=CC#CC#C</chem>	358.08 K	NA	0.99999973
2	<chem>CC#CCC#C</chem>	358.24 K	353.7 K	0.99999944
3	<chem>FC#CCC#C</chem>	358.30 K	NA	0.99999854
4	<chem>NCCCCF</chem>	358.84 K	NA	0.99997061
5	<chem>FC=CO</chem>	357.15 K	NA	0.99993720
6	<chem>CC=CO</chem>	357.09 K	NA	0.99993029
7	<chem>OCC#C</chem>	360.07 K	NA	0.99976953
8	<chem>ClC#CC#C</chem>	356.08 K	NA	0.99973128
9	<chem>ClCC=CC</chem>	360.57 K	356.7 K	0.99963449
10	<chem>ClCC#CC#C</chem>	355.51 K	NA	0.99914651

Considering that the first test is very simple, it is not surprising to find many candidate molecules with values close to the 358.15K target. Many molecules have no associated experimental data. When the experimental value exists, the deviation is within the margin of error of the MG model [9].

5.2. Test with a $T_b=358.15\text{K}$ and $T_m=100\text{K}$ target

In this second test, the objective function is composed of two properties: the boiling point temperature with the same target value and the melting point with a target value of 100K. The other parameters are kept alike. The ten best molecules for the first test are:

Table 2. Ten best molecules matching a 358.15K boiling point and 100K melting point target

	molecule (SMILES)	T_b (MG)	T_m (MG)	Performance
1	<chem>FC(Cl)C(C=C)=C</chem>	357.09 K	97.64 K	0.99979073
2	<chem>IC(F)Cl</chem>	368.26 K	98.05 K	0.99669527
3	<chem>ClC(F)C(Cl)=C</chem>	355.08 K	123.55 K	0.98266449
4	<chem>ClC(F)C(C=C)=CF</chem>	389.71 K	139.57 K	0.92321512
5	<chem>C=CI</chem>	346.16 K	167.64 K	0.87117724
6	<chem>BrCC=C</chem>	344.25 K	170.97 K	0.85896164
7	<chem>C=C(C=C)C=C</chem>	333.79 K	167.39 K	0.85826189
8	<chem>C=CCCl</chem>	319.16 K	166.73 K	0.83318863
9	<chem>BrC=C</chem>	307.55 K	158.86 K	0.82871597
10	<chem>BrCCC=C</chem>	375.75 K	182.17 K	0.81830087

The first two molecules match fairly well the constraints. Then, as expected from a more constrained problem than the first test, the candidate molecule has more widespread objective function values.

6. Conclusion and future works

The first prototype of the CAMD software is able to solve simple problems. The next step is to progressively complement the prototype in order to deal with complex and mixture problems: addition of operating conditions, multi property objective functions, multi fragment molecules, multi level resolution... The software object architecture, established thanks to our conception method based on models alignment, enables to insert those changes easily and progressively.

References

1. L.E.K. Achenie, R. Gani, V. Venkatasubramanian, 2003. Computer Aided Molecular Design: Theory and Practice, Computer-Aided Chemical Engineering, vol. 12, 2003, Elsevier.
2. R. Gani, A. Fredenslund, Computer aided molecular and mixture design with specified property constraints, Fluid Phase Equilibria, 82, 39-46, 1993.
3. J.C. Henderson, N. Venkatraman, Strategic Alignment: A Model for Organizational Transformation through Information Technology, in T. Kochan and M. Unseem, eds, Transforming Organisations, Oxford University Press, NY, 1992.
4. www.omg.org
5. M. Korichi, V. Gerbaud, P. Floquet, A.H. Meniai, S. Nacef, X. Joulia, Computer Aided Aroma Design. I-Molecular Knowledge Framework, Chemical Engineering and Processes, 2008.
6. P.M. Harper, and R. Gani, A multi-step and multi-level approach for computer molecular design, Computer and Chemical Engineering, 24, 677-683, 2000.
7. J. Song, H. Song, Computer-Aided Molecular Design of Environmentally Friendly Solvents for Separation Processes, Chemical engineering & technology, 31, 177-187, 2008.
8. D.C. Weis, D.P. Visco, Computer-aided molecular design using the Signature molecular descriptor: Application to solvent, Computer and Chemical Engineering, 34, 1018-1029, 2010.
9. J. Marrero, R. Gani, Group-contribution based estimation of pure component properties, Fluid Phase Equilibria, Volumes 183-184, 2001, Pages 183-208.
10. K.G. Joback and R.C. Reid. 1987. Estimation of pure-component properties from group-contributions. Chemical Engineering Communications. 57(1):233-243.