17th European Symposium on Computer Aided Process Engineering – ESCAPE17
V. Plesu and P.S. Agachi (Editors)

Simultaneous polymer property modeling using Grid technology for structured products

Kavitha Chelakara Satyanarayana[a], Jens Abildskov[a], Rafiqul Gani[a]

[a] *CAPEC, Department of Chemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark, kac@kt.dtu.dk, ja@kt.dtu.dk, rag@kt.dtu.dk.*

Abstract

Polymer property modeling requires the handling of linear and non-linear parameters; solving those needs high computational resources and time. When modeling several properties, time taken for parameter estimation and computational power demands in general, increases proportionately with the number of properties modeled. Grid technology greatly facilitates this work different property models (based on either group contributions, connectivity indices or other methods) can be generated in parallel. Here we present a suite of new polymer property models based on groups and atomic contributions. These models are more suitable for computer-aided polymer design algorithms than their predecessors. A comparison of the times required for generating the models in series and in parallel, using grid technology, is also presented.

Keywords Polymer properties, group contribution, connectivity index, grid technology.

1. Introduction

Increasing consumer demands for new and better chemical products based on specific properties, is driving the chemical, material and pharmaceutical industries to search for new chemicals and investigate whether they meet the specific demands of the end users. New structured chemical products, their i) design (identification of the molecular structure); ii) manufacture (how to produce them reliably as well as efficiently); and iii) performance (how to

verify that the manufactured product will perform significantly better than benchmark products) are needed to improve the quality of human life while maintaining Nature's capacity to renew itself. Because of the size of the molecular structures, the number of possible alternatives for any type of product can run into millions. It is no longer feasible to depend on old methodologies which involve production and experimental testing of many alternative chemicals for the specified properties. Thus, a systematic strategy to limit the search space where the promising products can be found needs to be identified, so that the time consuming and costly resources can be applied more effectively, thereby addressing the issues related to time and cost to market. Rather than 'create-and-test' which is the paradigm of current approaches to chemical product development (that is first create a product and then test if this is suitable), the objective should be to fix a performance target (defined through product needs) and identify alternatives that match the target. In this way, numbers of trials are reduced and resources are spent only on the final selection of the already identified feasible alternatives. Emerging technologies involve the use of computer-aided polymer design techniques. Establishing structure-property relationship for polymers, suitable for polymer design is one objective of this work.

2. Polymer Property Prediction Methods

Group contribution (GC) methods are useful for correlating a material property with the chemical composition and state of matter of a substance. In these methods, functional groups (like CH_3, CH_2, OH, etc.) contribute to molecular properties in a systematic additive fashion. Thus, a GC method treats a property of a given compound as an additive function of parameters characterizing the groups in the molecule. Van Krevelen [1] proposed GC-based methods for predicting polymer properties, the groups defined by this method were relatively large fragments. The method is often quite accurate, but has a lot of less obvious features, like a large number of corrections to nearest neighbour interactions in the basic level of property prediction. For example, to enhance accuracy, the same group has different contributions depending on its position in the polymer repeat unit (side chain, main chain etc.). These drawbacks limit the flexibility of this method and its usefulness in computer-aided polymer design algorithms. As an alternative to GC methods the method of Bicerano [9] is based on topological information related to molecular structure.

The Marrero/Gani [2] GC method is here used for developing a new and simple GC-based model for predicting polymer properties. It has a larger range of groups, classified under first-order, second-order and third-order groups. This GC method is described as:

$$f(X) = \sum_i N_i C_i + w \sum_j M_j D_j + z \sum_k O_k E_k \qquad (1)$$

where C_i is the contribution of the first-order group of type-*i* that occurs N_i times, D_j and E_k are the contributions of the second-order group of type-*j* and the third-order group of type-*k,* that occurs M_j and O_k times, respectively. In the first level of estimation, the constants *w* and *z* are assigned zero values because only first-order groups are employed. In the second level, the constant *w* is unity and *z* is zero, because only first- and second-order groups are involved. In the third level, both *w* and *z* are unity. The left-hand side of Eq. (1) is a simple function of *f*(X) of the property X. If at least one part of the molecular structure of a given polymer is not described by any of the available groups, the method is not applicable. To overcome this problem, we would – ideally – introduce a new group and determine its contribution from regression, if sufficient experimental data is available for doing so. Often this is not the case. This difficulty may be overcome by developing a method in which the properties are expressed in terms of connectivity indices (CI's) combined with atomic contributions. Such a CI-based method is equivalent to the prediction of properties by a GC approach but it has the contributions of atoms and bonds instead of groups. It can not be expected to be as accurate as a GC approach, but for creating the contributions of a 'missing groups' this might not be a problem. Here we will develop GC methods based on Marrero/Gani groups, and atomic CI-based methods of the form:

$$f(Y) = \sum_i (a_i A_i) + b({}^{v}\chi^{0}) + 2c({}^{v}\chi^{1}) + d \qquad (2)$$

where Y is the polymer property to estimate, A_i is the number of atom *i* occurring in the molecular structure a_i times. ${}^{v}\chi^{0}$ and ${}^{v}\chi^{1}$ is the zeroth- (atom) and first-order (bond) connectivity index, respectively, as described by Kier and Hall [4]. *b*, *c* and *d* are adjustable parameters. Generating different polymer property models using GC- and CI-based methods involves the handling of both linear and non linear parameters. Solving them can be time-consuming. As different properties are being modeled, time taken for determining group or atom contributions increases proportionately with the number of properties. Grid technology [5] is used here to develop property models in parallel execution, decreasing the time for model generation by using freely available computer resources in a network.

3. Grid Computing

Grid computing enables higher throughput computing by taking advantage of many networked computers to model a virtual computer architecture capable of distributing process execution across a parallel infrastructure [6]. Here we use 'OfficeGRID' [5] of MESH-technologies. With 'OfficeGRID', the user can submit the jobs from command line. It also supports the working of different platforms simultaneously. Moreover, it provides good flexibility as the machines can join and leave the grid freely and the machines can accept one or

more jobs concurrently. 'OfficeGRID' also includes advanced error handling like detecting node failure, automatic resubmission of jobs sent to the node, and network failures - even failures during large network transfers - are dynamically corrected.

4. Methodology

This work involves inductive learning. That is, experimental data are collected, a property model is derived from the data set and finally the properties are predicted using the developed model. Experimental data on polymer densities (25 °C), glass transition temperatures and solubility parameters (25 °C) are collected. These data are collected from various literature sources [1, 7, 8]. Each polymer repeat unit structure in the data set are represented by a set of Marrero/Gani groups for developing a GC-method, and its set of atoms for developing CI-based model. Contributions of the groups/ atoms and constants are determined from fitting the model to the experimental data, by adjusting its parameters to minimize a sum of squared residuals (differences between experimental and estimated values of the target properties). The Levenberg/Marquardt minimization method is used for solving the least squares problem. The models developed are tabulated in Table 1.

Table 1: GC-and CI-based property model

Properties	GC-based model
Glassy Amorphous density (ρ_g)	$M_w/\rho_g - V_{g0} = \Sigma_i N_i V_{g1i} + w \Sigma_j M_j V_{g2j} + z \Sigma_k O_k V_{g3k}$
Rubbery amorphous density (ρ_r)	$M_w/\rho_r - V_{r0} = \Sigma_i N_i V_{r1i} + w \Sigma_j M_j V_{r2j} + z \Sigma_k O_k V_{r3k}$
Crystalline density (ρ_c)	$M_w/\rho_c - V_{c0} = \Sigma_i N_i V_{c1i} + w \Sigma_j M_j V_{c2j} + z \Sigma_k O_k V_{c3k}$
Glass transition temperature (T_g)	$T_g \cdot M_w - Y_{g0} = \Sigma_i N_i Y_{g1i} + w \Sigma_j M_j Y_{g2j} + z \Sigma_k O_k Y_{g3k}$
Solubility parameter (δ)	$\delta^2 V_a - H_{coh0} = \Sigma_i N_i H_{coh1i} + w\Sigma_j M_j H_{coh2j} + z\Sigma_k O_k H_{coh3k}$
	CI-based model
Glassy Amorphous density (ρ_g)	$M_w/\rho_g = \Sigma_i a_i A_{gi} + 8.54({}^v\chi^0) - 4.1({}^v\chi^1)$
Rubbery amorphous density (ρ_r)	$M_w/\rho_r = \Sigma_i a_i A_{ri} - 12.9({}^v\chi^0) - 17.04({}^v\chi^1)$
Crystalline density (ρ_c)	$M_w/\rho_c = \Sigma_i a_i A_{ci} - 2.4({}^v\chi^0) + 3.71({}^v\chi^1)$
Glass transition temperature (T_g)	$T_g \cdot M_w - Y_{gc0} = \Sigma_i a_i A_{gi} + 9056.87({}^v\chi^0) - 15842.3({}^v\chi^1)$
Solubility parameter (δ)	$\delta^2 V_a - H_{cohc0} = \Sigma_i a_i A_{cohi} - 46226.9({}^v\chi^0) - 59656({}^v\chi^1)$

The symbols in GC-based models, V_g, V_r, V_c, Y_g and H_{coh} represent the contributions (C_i, D_j and E_k) of the first, second and third order groups, respectively, for the corresponding properties. $V_{g0} = 1.4274$, $V_{r0} = -0.9842$, $V_{c0} = -1.149$, $Y_{g0} = 7641.53$ and $H_{coh0} = 3374.697$ are additional parameters of the estimation model. M_w and V_a are the repeat units molecular weight and amorphous volume, respectively. The symbols in CI-based models, A_{gi}, A_{ri}, A_{ci}, A_{gi} and A_{cohi} represent the atomic contributions to the corresponding property. $Y_{gc0} = -5371.355$ and $H_{cohc0} = -2696.465$ are additional parameters of the estimation model.

5. Results and Discussion

The main idea behind the development of new, simple and accurate models is to predict polymer properties using methods, which are convenient for computer-aided polymer design algorithms. Table 2 gives the correlation statistics of models developed using these two approaches. From Table 2, the correlation coefficient values show that the Marrero/Gani based GC-methods fit the experimental results well. There is a significant difference in the accuracy of first-, second- and third-order predictions. For the CI-based method, the average relative errors are greater than the Marrero/Gani GC-based models.

Table 2: Correlation statistics for Marrero-Gani GC-based method and CI-based method

Variable type	Properties				
	Marrero-Gani GC-based method				
Units	ρ_g (g/cm^3)	ρ_r (g/cm^3)	ρ_c (g/cm^3)	T_g (K)	Δ (J^{½}/cm$^{3/2}$)
Data points	58	35	43	105	37
Overall Correlation coefficient R^2	0.9931	0.9965	0.9919	0.942	0.9546
Avg. relative error (ARE)	0.699%	0.861%	1.102%	3.856%	1.660%
First-order ARE	0.747%	0.947%	1.11%	4.55%	1.832%
Second order ARE	0.709%	0.869%	1.102%	3.965%	1.661%
Third order ARE	0.699%	0.861%	1.102%	3.856%	1.660%
	CI-based method				
Correlation coefficient R^2	0.9782	0.9936	0.934	0.6914	0.9265
Avg. relative error, (ARE)	1.435%	1.287%	4.756%	11.237%	16.339%

As the data points available for density is limited, the method of relating the van der Waals volume to the molar volume suggested by Van Krevelen [1] is used. This gives the volume contributions to a large set of Marrero/Gani first-order groups. Glass transition temperatures predicted by Van Krevelen are accurate to within 20 K for 80 % of around 600 polymers. The Marrero/Gani GC-based method predicts glass transition temperatures with the accuracy of 84 % within 20K for a data set of 105 polymers. Regarding solubility parameter, the Van Krevelen method is in the order of 10 % accuracy while the Marrero-Gani GC-based method given here is 1.66% accurate. The program generating the above mentioned models was executed in series and the total time taken for execution of ten codes was 137 seconds. These program was also executed such that the minimizations were made parallely using ‘OfficeGRID’ (with three ‘client’ machines connected to a ‘master’ machine). The total time taken for these ten codes execution was just 53 seconds. This reduced the time for model prediction considerably.

6. Conclusions

Generating properties from the structural description is a starting point of computer aided molecular design (CAMD), which finally gives out the polymer structure for the given set of target properties. So far, the predicted Marrero/Gani GC model looks good considering the average relative error. But, as the data points were limited in number, more data points will be collected and these models will be validated for those data points in the future. As soon as the property models for other properties like refractive index, permeability, dielectric constant are generated, polymer design software can be developed for finding polymer structures with desired properties. 'OfficeGRID' proved very efficient. But, at this point, the model development is not so time consuming. In course of time, these models will be extended for predicting polymer properties for polymer structure with several side chains (repeated at some frequency, etc.), where the computations could be really massive and grid technology will be needed to make the unsolvable problems solvable in a small time period.

Acknowledgements

The project is funded by the NABIIT program.

References

1. Van Krevelen, D.W. Properties of Polymers, Their Estimation and Correlation With Chemical Structure, Elsevier, 1990.
2. Marrero, J.; Gani, R.; Fluid Phase Equilibria, 2001, 183.
3. Gani, R.; Harper, P.M.; Hostrup, M.; Ind. Eng. Chem. Res., 2005, 44, 7262.
4. Kier, L.B.; Hall, H.L.; Molecular connectivity in Structure Activity Analysis; John Wiely & Sons; New York, 1986.
5. http://www.meshtechnologies.com/products.php?id=OfficeGRID
6. http://en.wikipedia.org/wiki/Grid_computing
7. Brandrup, J.; Immergut, E.H.; Grulke, E.A; Editors,Polymer Handbook, Wiley Interscience, 1999.
8. James E. Mark; (editor), Physical properties of Polymers Handbook, American Institute of Physics, 1996.
9. Bicerano, J.; Prediction of Polymer Properties, Marcel Dekker, Inc., 2002.
10. Abildskov, J.; Development of a new group contribution method, M Sc. Thesis, February 1995.