About the Project

Much of the value of commercial oligomeric or polymeric products is derived from the way the product interacts with liquids in its environment, both during production and during use. For instance, the life-saving abilities of many medicines depend critically on how the molecules that constitute the active ingredients distribute themselves among water-rich tissue and fatty tissue in the human body. Many of these active ingredients are produced using solvents; the active ingredients must then be purified by carefully separating them from the solvents (often with other, non-solvent liquids), or by carefully growing crystals in a liquid medium. The same need, namely to understand in a quantitative and prediction-friendly way, how a given polymeric substance will interact with liquids, exists across a wide range of industries and applications – from paints and coatings to lightweight aerospace composites to batteries and solar panels.

Despite the importance of this need, in many instances the understanding that people rely on is based on qualitative observation, with little in the way of predictive power. As a result, when problems arise with existing products, or new products are needed to respond to changing market conditions, a laborious trial and error process is used to improve our understanding enough to solve the problems. Although good quantitative approaches exist, and in many cases these approaches can solve the same problems at 10x lower cost and 100x faster than trial and error, the use of the best approaches remains limited. Often, if a quantitative approach is used, it tends to be a simple one such as the Hildebrand solubility parameter, which hasn’t changed much since the 1940s!

There are two key challenges that must be overcome in order to facilitate the use of better quantitative approaches, such as the Hansen Solubility Parameter (HSP) approach, to solve these problems. The first challenge is centered on data. The use of HSP requires a set of solubility data on known chemical structures. In the realm of polymers, this data is difficult to obtain. Even more challenging, often the data is available for structures with compositions that are not known precisely, or for business reasons are kept confidential. Although data for other classes of matter (such as pharmaceutical compounds) is more widely available, to date no one has figured how to apply this data to polymer materials. The most well-known “rules of thumb” that exist for predicting HSP for polymers date from the 1970s (although the topic is beginning to get a lot more attention).

The other challenge involves access and communication. For the average polymer chemist or process engineer, predicting HSP either involves coming to grips with a fair bit of unfamiliar math, or purchasing software that takes time to buy, install, and learn how to use. Because the problems that HSP are intended to help solve involve practical, day-to-day decision making, getting an answer quickly is important.

The aim of this project is explore both new, machine-friendly quantitative methods for predicting the Hansen Solubility Parameters of polymers, as well as new web-based interfaces that let users quickly input the information they know, and then provide a useful suggestion about what to do next.

Previously Published Work

The study of Hansen Solubility Parameters dates to the pioneering work of Charles Hansen in the 1960s. Hansen’s three-parameter approach to predicting solubility offered a substantial improvement over the earlier one-parameter Hildebrand approach, while still retaining a significant basis in thermodynamics (although in practice much empirical “fine tuning” was incorporated). In the 1970s, the Hansen method gained notoriety because it successfully predicted that two poor solvents could be mixed to create a good solvent for a difficult-to-dissolve paint product. Although this type of prediction was also a feature of Hildebrand theory, in many cases the mixture predicted to be good solvents by Hildebrand theory were in fact poor solvents. Hansen’s theory proved a much better guide to selecting poor solvents that could be mixed to create a good solvent.

Since the 1970s there have been a number of proposed refinements to the Hansen method, yet none have become especially popular. The Hansen method appears to strike the right balance between simplicity and predictive power. Refinements typically involve additional quantities (pushing the number of adjustable parameters needed well beyond three), and many of these quantities are much less intuitive and easy to visualize.

The typical application of the Hansen method has been with organic macromolecular solutes and small molecule solvents. Some work has also been performed on small molecule solutes such as pharmaceutical compounds. Until recently, though, many types of molecular structure, particularly those exhibiting structure at length scales larger than those of typical molecular solutes or organic polymer repeat units, were not considered good candidates for the Hansen method.

In 2008, my colleagues Dr. Gregory Yandek and Dr. Joseph Mabry at the Air Force Research Laboratory first proposed to the Air Force Office of Scientific Research that solubility parameter approaches should be investigated more deeply for nanostructured materials. In 2010, shortly after Kevin Lamison and I joined their team, the two of us began performing solubility experiments in pursuit of this goal. Soon after the work began, our laboratory was shut down for several weeks for asbestos abatement. Because we could not perform additional experiments, we spent a few weeks on data analysis.

The data analysis problem for solubility parameter determination involves significant non-linear model fitting. One of the difficulties with such methods is that a good method of uncertainty estimation is difficult to obtain. In addition, the optimization “landscape” for this problem often exhibits many local minima, so traditional gradient descent methods can easily miss the best solutions, as well as exhibiting sensitivity to the choice of starting point. To overcome these issues, we built an entirely new method for solving the data analysis problem, one that uses a systematic search to map out parameter space, find the absolute best model, and then, because the optimum is typically not a point but an extended region having a complex (sometimes discontinuous) shape in parameter space, we created simple methods to describe this region. This method required some simple software, but was computationally both tractable and simple to understand despite being “brutish” – as in a lot of brute force. With these methods and our experimental results, we learned that indeed Hansen Solubility Parameters were reliable for nanostructured materials.

I described the first results from this work at the inaugural presentation for the American Chemical Society’s first Silicon-Containing Polymers and Composites Workshop in December 2010. In 2011, Lisa Lubin, then a teacher of high school mathematics, joined the effort as part of the Air Force Research Laboratory’s STEM Education Outreach efforts. Later in 2011, we published the first conference paper on the subject. In 2012, we published a full paper in the peer-reviewed scientific journal *Macromolecules*. In that same year, the research group led by David Schiraldi at Case Western Reserve University also published their results, showing that, just as we had seen, the Hansen Solubility Parameter-based approaches generated useful predictions for nanostructured chemicals.

The study of solubility parameters proved to be a good topic for engaging STEM educators, so in the years after the initial publication we involved more current and aspiring teachers in the research, including Josephus Dossen and Shawn Kirby. Mr. Kirby used these experiences to create teaching tools and educational laboratory experiments related to solubility. Additional results were presented at the American Chemical Society Fluoropolymer Workshop in 2012. In 2014, we extended these methods to the study of extraction of dyes from petroleum, reporting the results at the American Chemical Society spring meeting. In 2015, Anish Tuteja’s group at the University of Michigan used the Hansen Solubility Parameter approach to understand the behavior of the nanostructured chemicals they investigated, publishing the results in 2017. In 2018, these methods were again used in published by Drs. Levi Moore and Timothy Haddad at the Air Force Research Laboratory to characterize some of the most thermally stable organic/inorganic materials ever created.

Due to the continuing interest in this area, in 2019 I decided to rebuild the original code for the analysis (written in Visual BASIC and lost when the Air Force disabled VBA on its applications in the early 2010s), update it to a more modern language (Python) and integrate some concepts (support vectors) from modern machine learning. The process for predicting Hansen Solubility Parameters for a given structure has not changed much since the 1970s. It involves simple algebra and a pencil and paper. As recently pointed out, “pencil and paper” actually do a laudable job of predicting Hansen parameters, at least for small molecules. Nonetheless, I am currently investigating whether the advent of modern data science tools can do better. In addition to presenting results at the American Chemical Society Industrial Polymer Chemistry Symposium coming up in August 2019, the codes and the results will be available on GitHub (<https://github.com/andrewguenthner/datascience>) and on the web at <https://andrewguenthner.com/datascience>.