Content

May 10, 2019

PD2M Newsletter

A quarterly update for Pharmaceutical Discovery, Development and Manufacturing Forum Members

Note from the Editor

Carla Luciani

Pharmaceutical Discovery, Development and Manufacturing Forum, Newsletter Chair

Lilly Research Laboratories

Happy 2019 to all PD2M Members!

In this first issue of the PD2M Newsletter, we featured three interesting articles.

- Sal Garcia, Sr. engineer advisor at Eli Lilly, discussed Multivariate Data Analysis (MVDA) using Latent Variable Modeling. In this article, Sal explained why the true potential of these tools in the applications to product development and hybrid modeling is still to be seen and highlighted the work of several academic groups around the globe.
- Andrew Livingston, Professor of Chemical Engineering at Imperial College London, is a world-renowned expert on molecular separation with nanomembranes. In this article, Andrew explains a technique to produce monodisperse (perfect) polyethylene glycol (PEG).
- Shujauddin Changi, Sr. Consultant Engineer at Eli Lilly & PD2M Programming Chair 2019, provided an update of the PD2M session allocation.

Enjoy the first issue of the PD2M Newsletter 2019!

Photo by Franki Chamaki

Abstract

Multivariate latent variable methods like Principal Components Analysis and Partial Least Squares are, in spite of their age, contemporary solutions to the data analysis needs of the pharmaceutical industry. Their true potential in the applications to product development and hybrid modeling is still to be seen and hopefully near in the horizon thanks to the work of several academic groups around the globe.

PD2M Newsletter

A quarterly update for Pharmaceutical Discovery, Development and Manufacturing Forum Members

Multivariate Data Analysis using Latent Variable Modeling: A timeless tool for the engineer

Sal Garcia

Lilly Research Laboratories

"*Nothing new here*" is a necessary declaration since in fact, nothing (or almost nothing) of what is written about here is new. Hopefully such display of honesty will entice you to continue reading this article in spite of its brutal lack of contemporary buzz words like "Big Data", "Machine Learning" or "Process Analytics". The truth is that the practical realization of these catchy phrases requires technology already developed, mature and well established, since the 1900's in some cases.

The application of latent variable models (or projection methods) like Principal Components Analysis [PCA] and Partial Least Squares [PLS] (or Projection to Latent Structures) for industrial process analysis dates to the early 1990's [Kresta et. al. Can. J. Chem. Eng. 69,1991]. The mathematics behind these applications however is in fact much older: PCA was developed by Pearson in 1901 [Philos. Mag.2, 559-572,1901], and PLS much after, by Horst in 1961 [Psychrometrica, 26, 129-149,1961]. And in spite of these methods being out in the literature for more than 118(!) and 67 years respectively, their application in industrial settings is still considered "novel". And the enthusiastic practitioners who want to introduce them still face an uphill battle against corporate culture and the established approach of analyzing variables one at a time.

The multidimensionality of the problems addressed in the development of a commercial process to manufacture a new medicine is well recognized and nowadays not an issue of controversy. PCA and PLS are methods ideally suited to the analysis of such multivariate problems. Particularly where the many uncertain variables involved are not independent but correlated with each other.

How do these methods work? The simplest way (IMHO) to illustrate Principal Components Analysis, is using the analogy of a weighted average. Assume a process outfitted with a collection of *T* thermocouples and *P* pressure sensors; and assume that variation in space is not important. The clever engineer in charge of monitoring such operation might choose to calculate a weighted average of the *T* thermocouples and a weighted average of the *P* pressure sensors to monitor the process. In doing so, our unsung hero has effectively reduced the problem of having to monitor *T*+*P* signals, to a simpler problem of monitoring two (the two weighted averages). The key to the success in our story was the prior knowledge that all thermocouples (and pressure sensors correspondingly) were correlated with each other. The key question in a real scenario is "which variables are correlated with each other, and how correlated are they?" PCA provides these answers.

Principal Components Analysis is a method that will analyze a set of data of *m* variables and *k* observations and help the user determine: *a)* The number of weighted averages (A) needed to properly represent the information contained in the dataset and *b)* the weights for each variable in each of those weighted averages. Variables that contribute significantly for the same given weighted average are inherently correlated. The arithmetic calculations in PCA are exactly those of a weighted average. Each variable has a weight associated per variable, per average to be calculated; all that has to be done is multiply and add:

$$
\begin{aligned}
WAv_k^1 &= x_{1,k}p_1^1 + x_{2,k}p_2^1 + \cdots + x_{m,k}p_m^1 \\
WAv_k^2 &= x_{1,k}p_1^2 + x_{2,k}p_2^1 + \cdots + x_{m,k}p_m^2 \\
&\vdots \\
WAv_k^A &= x_{1,k}p_1^A + x_{2,k}p_2^A + \cdots + x_{m,k}p_m^A
\end{aligned}
$$

The number of "weighted averages" that are independent to each and extractable from the data, is called the *number of principal components*. The collection of the calculated weighted averages per observation are called *scores*, the collection of *weights* for the variables are called *loadings*. Any of the available software tools available today will enable the user to import a set of data, push a "calculate" button and voila! PCA happens.

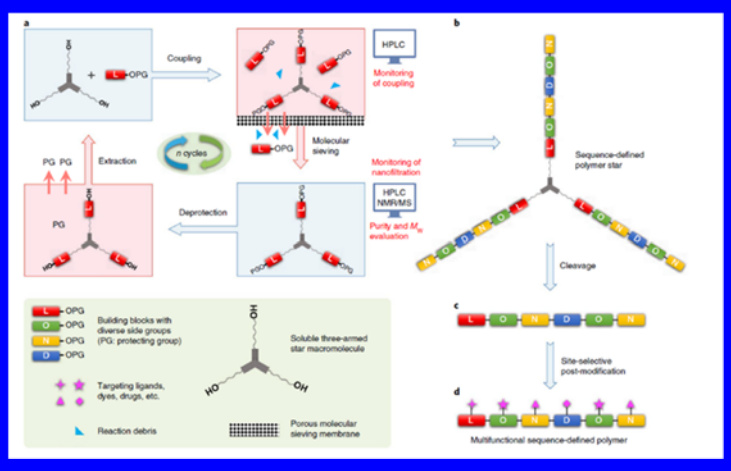
The attractiveness of this method for the engineer is that the number of principal components observable in a data set is analogous to the number of driving forces that are influencing the variations observed in the data. These driving forces (e.g. Heat transfer) can be inferred by analyzing the loadings for the variables for a particular principal component. And the scores (the magnitude of the components at each observations) are indicative of the influence of each driving force on the process in that very moment of operation. A published example of this type of analysis can be found in *Computers & Chemical Engineering,* 33, no. 12 (2009): 2106-2110.

Researchers have continued to advance the application and features of latent variable methods, a comprehensive review of all the additions, alterations, modifications and advances in the field is a gargantuan feat for a later time. For now I would like to comment on three research groups, whose research has caught my attention for their potential to impact industry and their novelty.

Starting with the application side of things I would like to point to the output from the Laboratory of Pharmaceutical Process Analytical Technology in Ghent University, led by Prof. Thomas DeBeer. Prof. DeBeer's group is an outstanding example of education innovation; preparing his students for their future job, not for the jobs of 20 years ago. Students from this group work tirelessly in the lab learning modern techniques and processes (and producing large amounts of good data); and later tirelessly working in the computer applying advanced mathematical methods to analyze the data and produce insightful conclusions as to how materials and processes relate in determining the quality of a product. Their creation of a consolidated data set of properties of a significant number of pharmaceutical powders [Int. J. Pharm. 549 (2018) 415–435] and the later analysis to determine which of the tests are relevant for pharmaceutical applications is one of kind. In a recent conference in Antwerp they presented their ongoing research towards producing a dataset that would enable the identification of surrogate materials for product development.

On the methodological side, I am personally eager to test the recent work on Dynamic Latent Variable Modeling (DiPCA) by Prof. Joe S. Qin and his students from the University of Southern California. His work [Comp. Chem. Eng. 114 (2018) 69–80] has focused on developing a modification to the PCA algorithm that not only identifies the most cross-correlated set of variables, but also the most auto-correlated one. His work has shown that by applying their method, frequencies of periodic operation can be effectively extracted from data additional to the cross-correlational information. Such a method not only makes a more effective monitoring tool, but also offers a data-based approach that yields a model with direct equivalences to a state-space representation of the process dynamics. It's application to a hybrid approach that combines DiPCA and a deterministically derived state space model is imminent [wink wink]. His work is in my view a solid proposal to embed dynamics in a PCA model by more than just lagging the data (the previous approach). Prof. Qin's work is innovative and disruptive and worthy of consideration by industry.

Lastly, and due to the recent attention given to "Big Data" without noting that "Big" and "Informative" are not the same thing; I'd like to highlight the output of the Robust@Leuven group led by Profs. Mia Hubert and Peter Rousseeuw at KU Leuven. Their world-wide recognized research focuses on the development of robust methods for statistical inference. Their seminal work on RobustPCA provides a method that is capable of producing a PCA model, in spite of the presence of outlying samples in the data. The stunning number of high quality papers from this group is noteworthy as well as the software produced. Robust methods like these are the lifesaver that keeps the scientist from drowning in a big sea of, nearly useless, data. Activities that entail the assessment of massive amounts of samples, for example those produced in a manufacturing facility for a large volume product, could benefit from the use of robust methods like those developed at the KU Leuven. Hopefully we will see some applications to pharma in the near future.



Abstract

This article highlights recent work at Imperial College to address the challenges of fabricating “perfect” PEGs using liquid phase etherification coupled to Nanostar Sieving.

References

Dong et al. “Sequence-defined multifunctional polyethers via liquid-phase synthesis with molecular sieving” Nature Chemistry (03 Dec 2018) https://doi.org/10.1038/s41557-018-0169-6

PD2M Newsletter

A quarterly update for Pharmaceutical Discovery, Development and Manufacturing Forum Members

“Perfect” PEGs through Nanostar Sieving

Andrew Livingston
Professor of Chemical Engineering
Imperial College London

Poly(ethylene glycol) (PEG) is widely used in pharmaceutical formulations and enjoys FDA “Generally Recognized As Safe” status. PEGs are used across myriad applications including for chemical modification (PEGylation) of both large and small molecules. PEGylation using PEGs with molecular weight above 2 kDa can offer a variety of beneficial effects, including prolonging circulation time of biologics, protecting against in vivo biological inactivation through proteolysis, improving solubility, and reducing immunogenicity. More than 12 PEGylated products are already approved, and many more are in trails and development.

There are several challenges which arise in PEGylation technology. One of these is molecular accuracy of PEGs over 2 kDa which are typically polydisperse; even narrowly polydisperse materials create an analytical challenge in determining molar mass, and may lead to a range of bioactivity. Another issue is the presence of diol impurities in the methoxy-PEGs typically used for PEGylation. Diols can result in unwanted crosslinking between two proteins. Finally, PEGs are usually only functionalizable at their end groups, leading to low payload capacity when they are considered as pro-drugs and carriers.

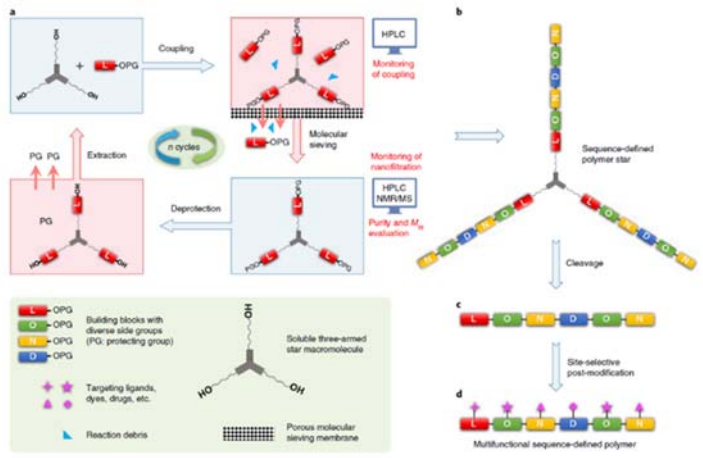


Figure 1. Unimolecular polymer fabrication by iterative synthesis with nanostar sieving (click here to enlarge image).

Recent work at Imperial College has addressed these challenges through fabricating PEGs using liquid phase etherification coupled to Nanostar Sieving (Figure 1). A central hub molecule with 2 or more reactive “arms” is used to grow PEGs outwards using unimolecular oligomers, typically between Eg8 and Eg12. When these oligomers are linear, the result is a high molecular weight, unimolecular PEG. Before cleaving the resulting PEG from the hub, the free end is functionalized, for example by methylation, to provide an mPEG. Subsequently, after cleavage from the hub the other end may be functionalized with any of the usually-desired end groups used in PEGylation. Remarkably, the resulting mPEGs are unimolecular – Figure 2 shows the purity of mPEG 5000 g mol^{-1}, which is over 96% oligomer purity.

The PEG oligomers can also be prepared with side chains, and these in turn may be reactive. This creates the opportunity for these to be added in a defined and discernible sequence of monomers, giving exquisite control over the molecular structure and, if orthogonally reactive side chains are used, the powerful ability to attach a set of payload molecules (for example, an API, and imaging agent, and a cell affinity or penetrating ligand) in a defined sequence.

This technology employs organic solvent nanofiltration membranes, and is scalable. A start-up from Imperial College, Exactmer, has developed a commercial process for linear PEGs approximately 5,000 g mol^{-1} and these are available with a range of functionalities. Exactmer also produces specific custom PEGs on request, and is developing the Nanostar Sieving technology for further synthetic and biological (oligonucelotides, peptides) polymers.

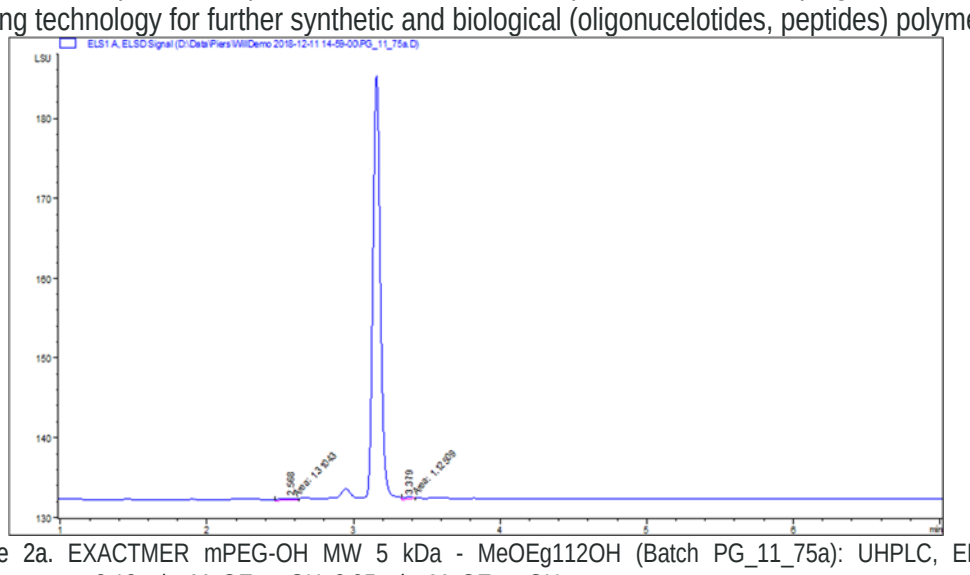


Figure 2a. EXACTMER mPEG-OH MW 5 kDa - MeOEg112OH (Batch PG_11_75a): UHPLC, ELSD chromatogram; 3.18 min, $MeOEg_{112}OH$; 2.95 min, $MeOEg_{100}OH$.

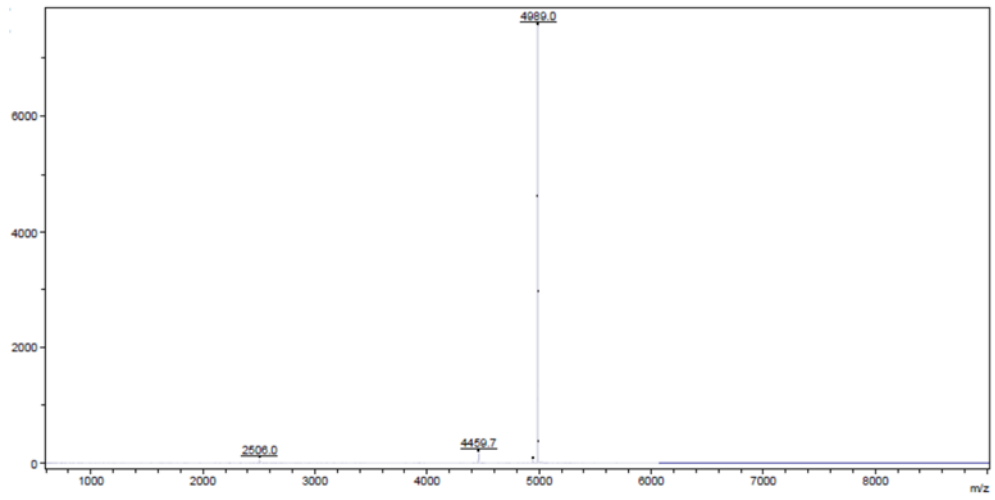


Figure 2b. EXACTMER mPEG-OH MW 5 kDa - $MeOEg_{112}OH$ (Batch PG_11_75a): MALDI+ spectrum. $[MeOEg_{112}OH+Na]^+$=4989.0 (97%), $[MeOEg_{111}OH+Na]^+$ = 4945.0 (0.8%), $[MeOEg_{100}OH+Na]^+$ = 4459.7 (2.8%), $[MeOEg_{112}OH+2.Na]^{2+}$ = 2506.0 (1%). NB the relative height of the Eg100 peak is consistent with its peak area in the resolved ELSD chromatogram.

Abstract

Are you attending the 2019 AIChE Annual Meeting in Orlando? This article includes an update of the PD2M session allocation.

PD2M Newsletter

A quarterly update for Pharmaceutical Discovery, Development and Manufacturing Forum Members

PD2M Session Allocation Update

Shujauddin Changi
Lilly Research Laboratories
PD2M Programming Chair 2019

In going with the "Track" based approach, 11 tracks were open to the audience for submission, with the call for abstract ending on April 12. Similar to previous years, the number of submissions grew exponentially in the last few days. An overwhelming response has been received for this year – partly due to the growing popularity of the PD2M area and otherwise due to the "Orlando" location effect.

In all 216 abstracts have been received. An approximate estimation of the number of sessions (if all abstracts were to be accepted) is reflected in the table below for various Tracks. The most popular session this year appear to be Continuous Processing followed by Integrated Product and Process Design (Experimentation and Modeling). Data Science & Analytics operation and Predictive Scale-Up/Scale-Down sessions have also generated quite an interest with each of the Tracks reaching nearly twice the number of abstracts as last year. Two new tracks that were created this year – a) Advances in drug discovery processes and b) Advances in large molecule processes, which have received enough abstracts to become stand-alone sessions – indicating the success of the Track based approach in drawing researchers to contribute to the "discovery" arm of PD2M. Innovative/Emerging technologies and Enabling advancing in formulations Tracks received somewhat of a lesser participation compared to last year with each receiving nearly half the number of abstracts as last year.

Number	Proposed Tracks	Track chair	Sponsored	Co-sponsored	# Sessions if all talks accepted
1	Continuous Processing in Drug Substance and Drug Product	Jon McMullan (Merck)	15B/26		7
2	Innovative/Emerging Technologies Relevant to Drug Substance and Drug Product	Tim Lee (GSK)	15B/26	12A	3
3	Integrated (Experimentation & Mathematical Models) Product and Process Design with Pharmaceutical Applications	Jacob Albrecht (BMS)	15B/26		5
4	Process Control Strategies in Pharmaceutical Development and Manufacturing	Kevin Seibert (Lilly)	15B/26	12G	3
5	Advancements in Particle Engineering and Material Sciences	Blair Brettman (GA Tech)	15B/26	2B/3	3
6	Data Science & Analytics for Operations Support and Predictions in Pharmaceutical Processes & Products	Shekhar Viswanath (Lilly)	15B/26	10E	3
7	Enabling and Advancing Formulations in Drug Product	Brendon Ricart (Abbvie)	15B/26		1
8	Predictive Scale-Up/Scale-Down for Production of Pharmaceuticals and Biopharmaceuticals	Kushal Sinha (Abbvie)	15B/26		3
9	Computational Solid State Pharmaceutics	Yuriy Abramov (Pfizer)	15B/26	21	1
10	Advances in Drug Discovery Processes	Samir Kulkarni (Pfizer)	15B/26		1
11	Advances in Large Molecule Processes	Michael Hoffman (Abbvie)	15B/26		1
12	Panel: Invited Talks	Tom Jean (BMS)	15B/26		1
13	Plenary Forum: Invited Talks	N/A	15B/26		1
14	Poster Session*	N/A	15B/26		
15	Pharmaceutical Discovery, Development and Manufacturing Forum Awards Ceremony*	N/A	15B/26		1
16	Pharmaceutical Discovery, Development and Manufacturing Forum Planning Session*	N/A	15B/26		1
	Total				35

In light of the above analysis, a good dilemma that presents itself, is that PD2M area would end up with 35 sessions (if all abstracts were to be accepted as is), exceeding the allocation of 26. Track chairs have been working diligently with their group of chairs/co-chairs in getting the abstracts organized into various themes, focusing on the quality and fit of abstracts for their respective Tracks. In parallel, a request has been made to the AICHE committee to increase the number of sessions this year and they will provide their feedback by May 15th. The next set of activities involve breaking down the tracks into various sub-sessions (as per the allocation allowed) before May 17th.

In addition to the above Tracks, couple of invited sessions this year include: 1) Panel session, themed "Precompetitive Collaboration", having 6 speakers with a panel discussion at the end, and 2) Plenary session, themed "Pharma 4.0", led by Moiz Diwan (Abbvie) and myself, where we are working on inviting leading experts from industry, academia, manufacturing, and FDA backgrounds to talk on this topic.

Overall, we are in for an exciting conference with a good all-round participation from all around the world for the PD2M area in 2019.