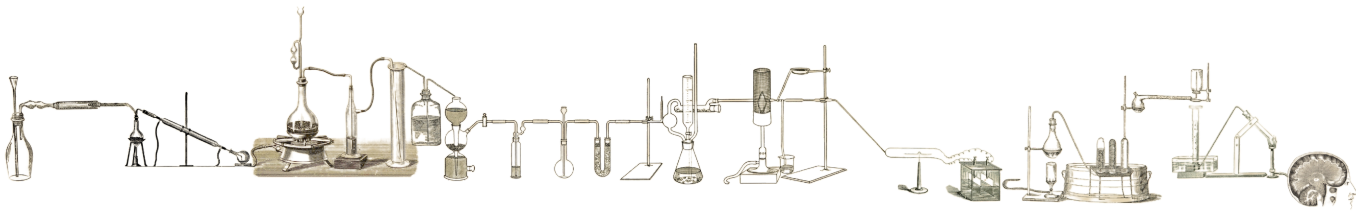
intellegens

WHITE PAPER

Machine Learning for Adaptive Experimental Design

Executive Summary

Identifying the optimal composition and processing parameters to achieve commercial performance goals as quickly as possible is the key objective of formulation design projects. Traditional approaches to formulation design suffer from key disadvantages: expert-driven design is labour-intensive and time-consuming; single-factor analysis misses the effects of correlations between factors; and conventional design of experiments is exhaustive but focused on covering the design space rather than rapidly achieving performance goals. Machine learning identifies improved formulations up to 10 times quicker than traditional approaches, by focusing experimental effort directly on formulations that will lead to successful products in as few experimental cycles as possible, adapting to the results of earlier measurements to ensure maximum value is extracted from the experimental outlay.



Traditional approaches to experimental design

The goal of R&D is to identify new formulations to meet commercial needs as quickly and efficiently as possible. Experimental campaigns to optimise formulations and design new products can consume vast quantities of time and resources. The development of new methodologies that accelerate the discovery and design of new formulations is therefore crucial for achieving time efficiency and cost reductions.

Historically, formulation design has been driven by the knowledge of domain experts, who leverage years of experience to intuit the next experiments to carry out. This enables experts to direct experimental campaigns, but has several limitations for commercial development: the bottleneck of single experts' availability, variability across an organisation as different experts make inconsistent decisions, and the potential for company expertise to be lost when valuable members of staff move on.

The road to more reproducible and methodical experimental design begins with systematically optimising single factors in the so-called COST (Change One Separate variable at a Time) framework. This approach requires the identification of the most important factor for a given formulation, with this factor then optimised, all other factors held constant. This procedure is then repeated for the next most important factor, with all other factors held constant, and so on. Although COST is more systematic than expert-driven design, it is unsuitable for complex systems where there are interactions between the factors or nonlinear responses.

Conventional Design of Experiments methodologies were developed to attempt to overcome the shortcomings of the COST approach. Design of Experiments aims to provide an efficient coverage of the formulation space to build understanding of the way responses change with each factor. The conceptually simplest method is a 'full factorial' experimental design, where N levels for each of F factors are considered, and all possible combinations of each level of each factor are measured: this is shown in Figure 1 for N=4 and F=2. A full factorial design requires F^N experiments, which with even a moderate number of factors quickly becomes prohibitively expensive.

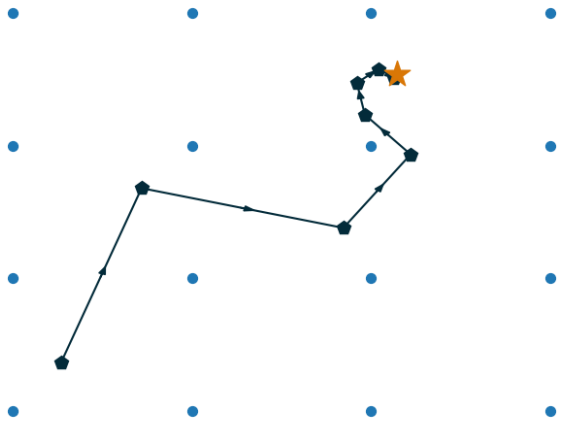


Figure 1. *Traditional Design of Experiments aims to cover the formulation space (blue points), here requiring 16 experiments, whereas machine learning-driven adaptive experimental design (black line) finds the quickest route to the optimal formulation (orange star) in as few experimental cycles as possible, here only requiring 8 experiments to achieve an optimal result*

More advanced Design of Experiments techniques cover the formulation space whilst requiring fewer experiments: one popular approaches is Latin Hypercube sampling, where instead of every combination of factors being used, only one measurement is proposed per level for each factor, ensuring that this is achieved simultaneously for all factors, as shown in Figure 2.

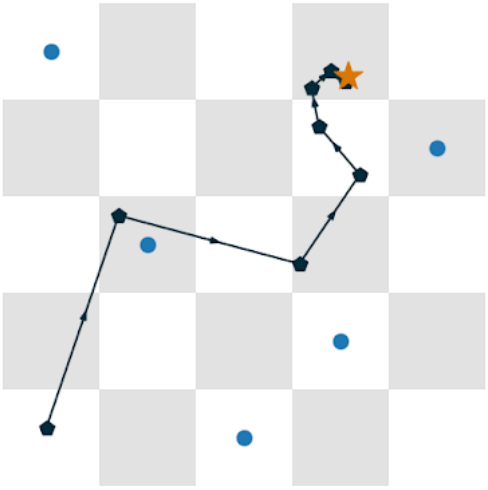


Figure 2. *Latin hypercube sampling of two factors, showing that for each level of each factor only one measurement is proposed. This can provide a more efficient coverage of the formulation space than a full factorial design, although is not as effective at identifying global optima as machine learning driven adaptive experimental design.*

Other popular methods include Box-Behnken designs, Plackett-Burman designs, central composite designs, Taguchi arrays, and definitive screening designs, each of which aims for efficient coverage of the formulation space in different ways. All of these Design of Experiments approaches are fundamentally attempting to answer the same question, namely how most efficiently to sample the formulation space to build understanding of the way each factor influences each response. However, this is not the most commercially relevant question for experimental design: instead of covering the available options, the true aim of a design project is to find the most effective formulation in as few experiments as possible.

Adaptive experimental design

Using machine learning we can shift the frame of experimental design from attempting to cover the formulation space to directly attempting to find the optimal formulation to achieve a given project's goals. By exploiting the predictive power of a machine learning approach we can select which experiments to carry out by which measurement will both be most likely to succeed against the project's goals and also will help improve the machine learning model itself, resulting in a virtuous cycle of a rapidly improving machine learning model suggesting increasing performant new formulations.

Figure 3 shows the performance of a variety of standard Design of Experiments approaches at finding the optimal point of a simple 5-factor analytic function. Almost all of the standard approaches achieve comparable results, finding similarly good optimal values in similar numbers of experiments. Alchemite™ machine learning driven adaptive experimental design finds better results quicker, requiring ten times fewer experimental measurements to find formulations much closer to the optimal result than those identified by the standard Design of Experiments approaches. In discovery projects this translates directly to many-fold savings in the time and resources required to achieve project goals, improving efficiency and productivity of the R&D process.

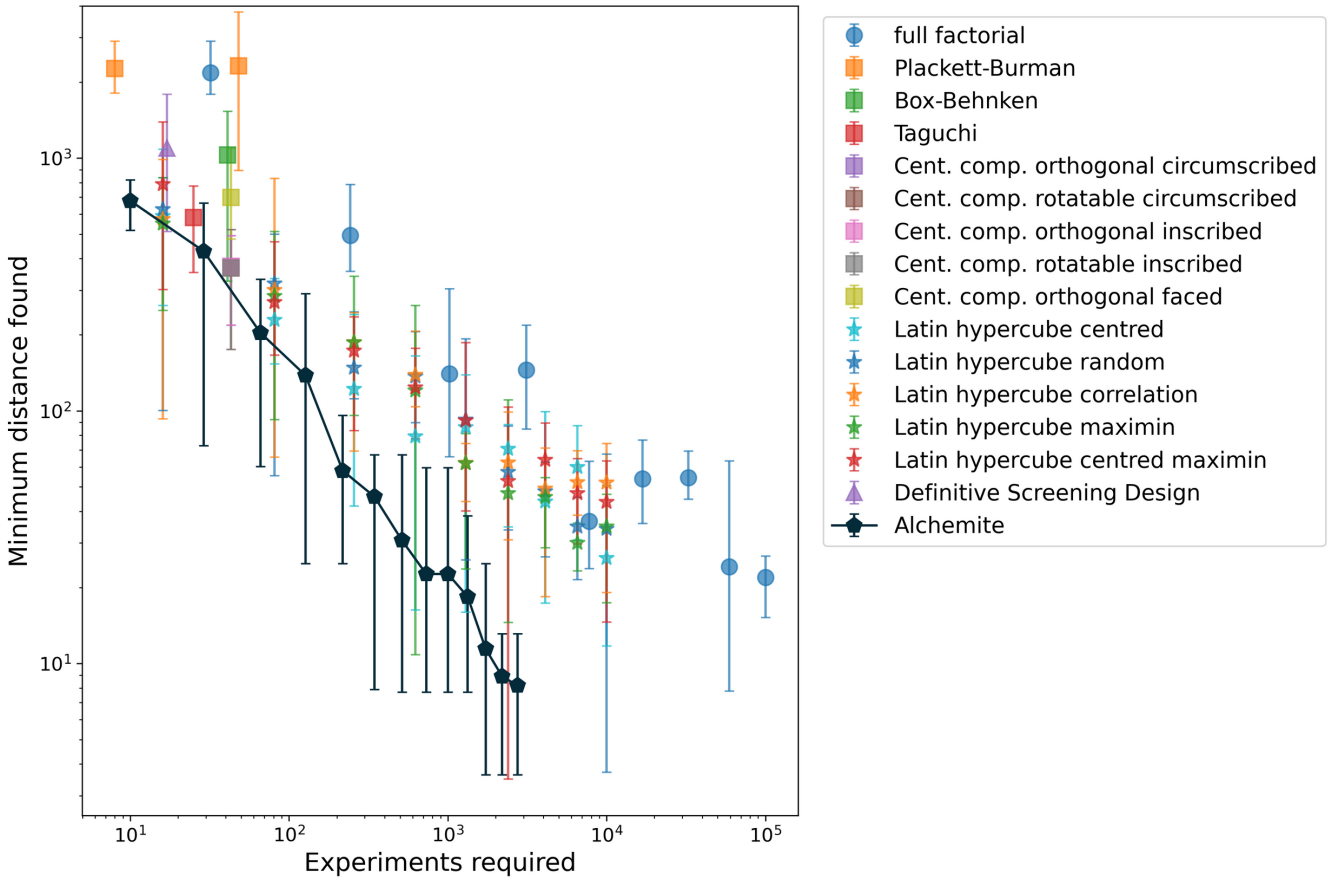


Figure 3. *The performance of a variety of standard Design of Experiments approaches at finding the optimal point of a 5-factor analytic function. The vertical axis shows how far the best result is from the optimal value, with the horizontal axis showing how many experiments are required to achieve this result. Alchemite™ machine learning, in black, achieves comparable accuracy to the best methods but needing ten times fewer experiments.*

The Alchemite™ approach not only offers improved formulations at greater speed, but is also capable of dealing with sparse and noisy data. By predicting and mapping the formulations landscape with associated confidence levels, the approach enables scientists and engineers to effectively identify the next best experiments to run to most rapidly succeed in discovery projects.

Guide your experimental design using the Alchemite™ Analytics platform

With the Alchemite™ Analytics platform, you transform R&D with machine learning by easily experimenting, modelling and visualising sparse and noisy real-world data. Choose the next best experiment to run next by quickly assessing the accuracy and confidence levels of your results.

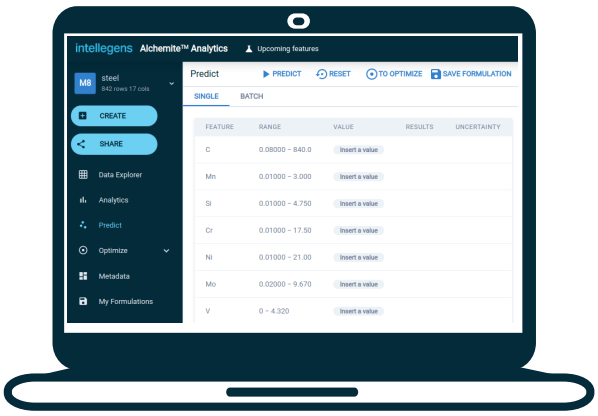
Why is Alchemite™ a better approach?

- Suggests the most important experiments needed (significantly fewer than standard Design of Experiments approaches)
- Improves understanding of specific properties
- Accurately maps the landscape of formulation space
- Provides a model for the direction to follow
- Many variables with many levels can be used
- Data can be sparse, noisy, and unstructured

About Intellegens

Intellegens has developed a unique artificial intelligence engine, Alchemite™, for training neural networks from sparse and noisy data, typical of real-world data. Alchemite™ is the deep learning platform for material and process optimisation to:

- **Validate data**
- **Guide experiments**
- **Optimise formulations**



Alchemite™ can be licensed as a SaaS product for scientists, engineers and technicians or stand-alone use by customer data analytics teams.

For more information, visit our website.

Want to learn more about how our AI technology can be applied to your specific needs? Contact us to learn more at info@intellegens.ai

Click on the link below to subscribe to our latest news and upcoming events
https://intellegens.ai/subscribe

intellegens

intellegens.ai
info@intellegens.ai
@intellegensai