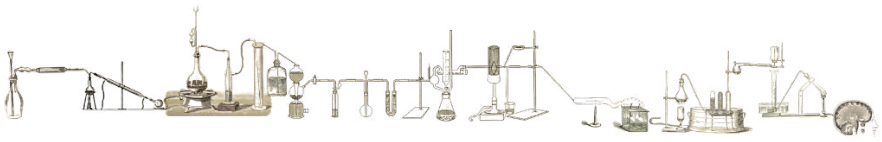
WHITE PAPER

Machine learning for adaptive experimental design

Reducing experimental time and costs by 50-80%



intellegens.com | info@intellegens.com
Intellegens, The Studio, Chesterton Mill, Cambridge, CB4 3NP, UK

intellegens

Applied machine learning

Executive Summary

Identifying the optimal composition or chemistry and the ideal processing parameters to achieve commercial performance goals as quickly as possible is the key objective in the design of **formulations, chemicals, materials**, and **biopharmaceuticals**. Traditional approaches 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 products and processes much faster than traditional approaches, by focusing experimental effort directly on those routes most likely to be successful. With the experimental costs associated with a typical R&D project in industry running to hundreds of thousands of dollars, a **50-80% reduction** in the number of experiments required can deliver a very significant return on investment.

Approaches to experimental design

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

Experimental campaigns to optimise formulations, chemistry, or materials consume vast amounts of time and resources

Historically, such 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, 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 system, with this factor then optimised and all other factors held constant. This procedure is 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 (DOE)** methodologies attempt to overcome the shortcomings of the COST approach. DOE aims to provide an efficient coverage of the design 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. This provides an exhaustive coverage of the experimental possibilities over those factors considered, and often enables powerful insights into the relationships between each factor (and their combinations) and the properties of interest. However, a full factorial design requires FN experiments, which with even a moderate number of factors becomes prohibitively expensive. **Machine learning**, as we shall see, instead aims to identify experiments that find the quickest route to the optimal experiment, achieving project goals with substantially less testing (Figure 1).

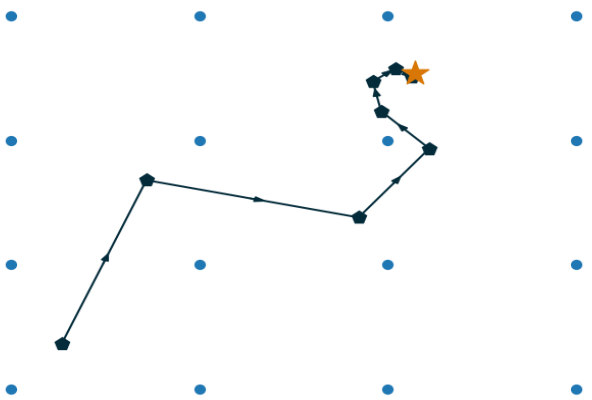


Figure 1. Traditional Design of Experiments aims to cover the formulation space (blue points), here requiring 16 experiments, whereas machine learning-driven adaptive experimental design (black line) finds the quickest route to the optimal formulation (orange star) in as few experimental cycles as possible,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

Of course, there are **more advanced traditional DOE techniques** that cover the design space while requiring fewer experiments. One popular approach 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. These may, however, risk sacrificing accuracy in the search for the optimal solution (Figure 2).

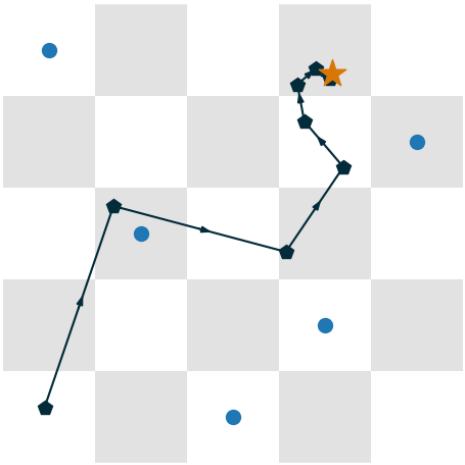


Figure 2. Latin hypercube sampling of two factors, showing that for each level of each factor, only one measurement is proposed. This can provide a more efficient coverage of the formulation space than a full factorial design, although is not as effective at identifying global optima as machine learning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

Other popular methods include Box-Behnken, Plackett-Burman and central composite designs, Taguchi arrays, and definitive screening designs.

There is another factor to consider when comparing traditional DOE with machine learning. DOE works well for understanding linear effects in a system, and some DOE methods are good for quadratic (second order effects). But DOE is not good for studying **higher-order, non-linear effects**. Yet most real experimental problems are characterised by such interactions. Exploring such spaces is another strength of machine learning, as the ML algorithm is able to understand and dynamically leverage relationships between the inputs, in contrast to the more static approach of traditional DOE.

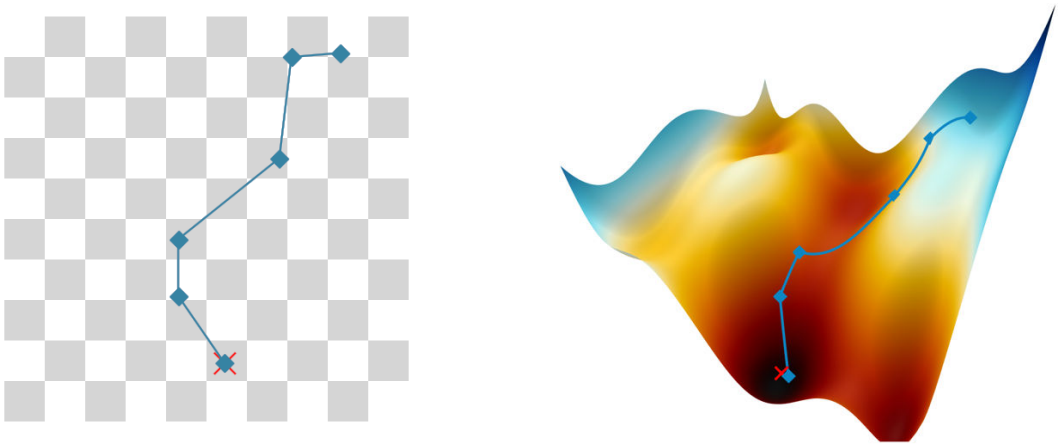


Figure 3. Machine learning allows for more accurate exploration of a design space with non-linear effects on the properties of interest. The figure illustrates how machine learning can navigate a complex design space to guide an experimental programme to its goal.

The key point is that all of the traditional DOE approaches fundamentally attempt to answer the same question: how to sample the design space most efficiently to understand the way that each factor influences each response. But this is not the most commercially relevant question for experimental design: instead of covering all available options, the true aim of a design project is to find the most effective product 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 design space to directly attempting to find the optimal formulation, chemical, material, or process 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 determining which measurement will both be most likely to succeed against the project's goals and also will help improve the machine learning

> Machine learning can determine which experiments are most likely to succeed

model itself, resulting in a virtuous cycle of a rapidly improving machine learning model, suggesting increasingly performant new products.

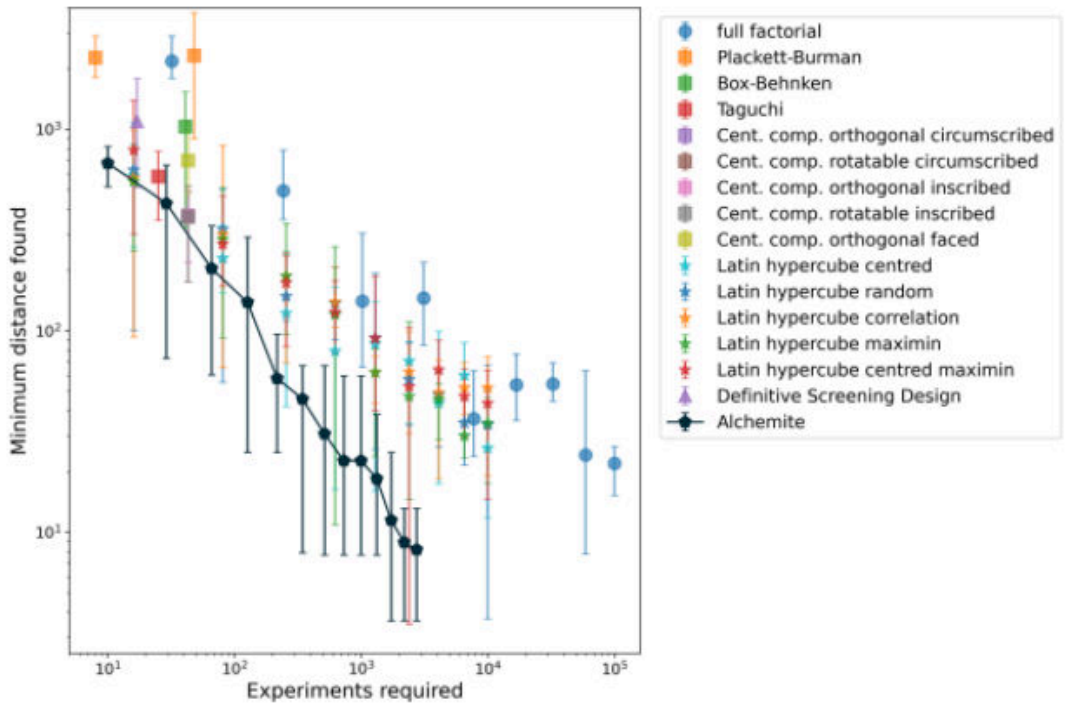


Figure 4. The performance of a variety of standard Design of Experiments approaches at finding the optimal point of a 5-factor analytic function. The vertical axis shows how far the best result is from the optimal value, with the horizontal axis showing how many experiments are required to achieve this result. Alchemite™ machine learning (dark line) achieves comparable accuracy to the best methods but needing ten times fewer experiments.

Figure 4 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, in theory, 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 or development 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.

The Alchemite™ approach not only offers improved solutions at greater speed, but is also capable of dealing with the sparse and noisy data that is typical of experimental datasets. By predicting and mapping the design 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.

Case study examples

Figure 4 showed a theoretical comparison of adaptive and traditional DOE, proposing a 90% advantage through the machine learning approach. What is achievable in practice?

Johnson Matthey, in their studies of catalysis formulations using Alchemite™, have reported halving of experimental workload in one project and identification of a new experimental route in another project that would lead to a five-fold reduction in experimental workload.

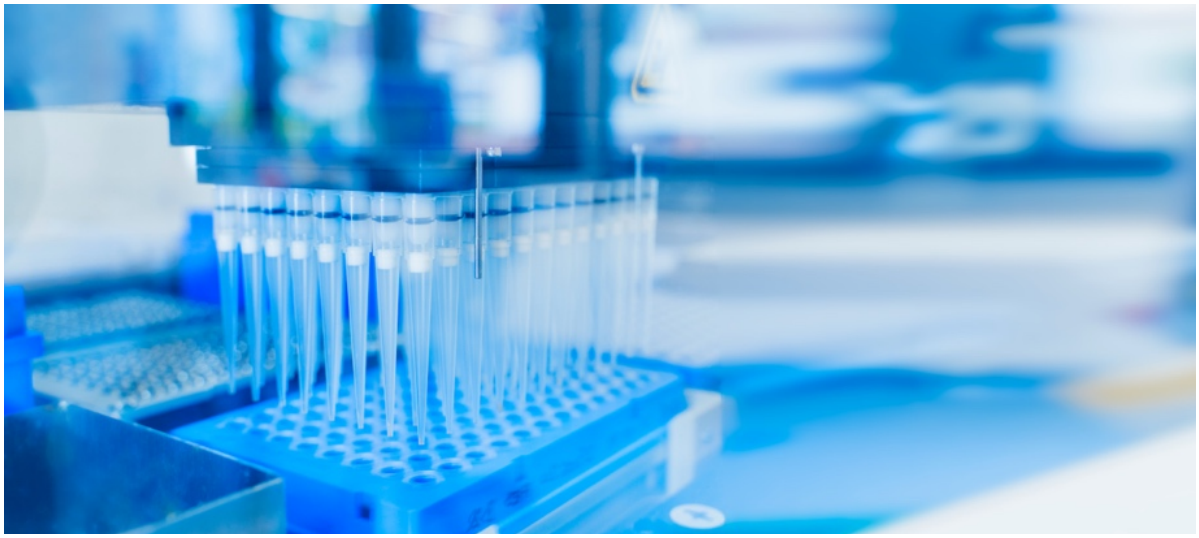
The **Advanced Manufacturing Research Center (AMRC)** found that use of Alchemite™ to plan a test programme for manufacturing research in composite tooling enabled them to achieve project objectives with 80% fewer tests.

Domino Printing Sciences also saw a dramatic reduction in the amount of experimentation needed in an ink reformulation project. They were able to find new formulations after two batches of experiment on 12 formulations proposed by Alchemite™. Traditional DOE would have required 1,800 formulations.

More on these and other case studies can be found at **intellegens.com/casestudies**.

Typical savings in experimental time and cost are in the range 50-80%

The exact savings in time and cost in an experimental programme taking an adaptive DOE approach will vary based on the details of the system being studied and project objectives. But, based on their experience of many such projects, including those above, the Intellegens team expects savings in the range 50-80%.



Guide your experimental design using Alchemite™

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

Some of the advantages of Alchemite™ machine learning versus traditional DOE are shown in Table 1. Alchemite™ also has several advantages compared to other machine learning methods, notably:

- It can build an ML model based on **sparse, noisy data** (as is often found in experimental datasets). Many ML methods fail in this situation.
- It has highly accurate methods for quantifying **uncertainty** in its predictions. This is important for Adaptive DOE in identifying experiments most likely to succeed.

Table 1. Alchemite™ vs traditional DOE

Limitation of traditional DOE	Alchemite™ approach
They still result in a high experimental burden	Suggests the most important experiments, resulting in 50-80% reduction in the number required
It's hard to address cross-correlations; they often model one parameter at a time	ML model captures complex, high dimensional, non-linear relationships, accurately mapping design space for materials, formulations, chemicals, or processes.
Can require statistical expertise	Method learns from the data provided to build a model – the user does not need statistical knowledge to set it up
Maps out a set of experiments, but the analysis and understanding of results is a separate task	Creates a machine learning model as part of the process and this can be used to understand what drives specific properties, and as a predictive tool

Finally, although adaptive DOE is a powerful machine learning-based evolution of traditional DOE methods, the latter remain widely-used and valuable in many scenarios. The approaches can be regarded as complementary, as indicated in Table 2.

Table 2. Complementary methods

Traditional DOE	Alchemite™
Use it to explore all options	Use it to efficiently achieve a goal
Use it to gather first data on a new problem	Use it to gain value from existing data, including merging data from previous projects
All experiments usually run at once	Typically iterate through experiments
Useful if you require a guarantee on the number (not the quality) of experiments	Aim for the highest quality experiments
Uses / requires advanced statistical knowledge	Pose question, get an answer in the domain language (no statistical / ML knowledge needed)

About Intellegens

Our mission is to be the leading machine learning solution for real-world, sparse and noisy data problems in industrial R&D and manufacturing processes. Our focus is on making it easy to apply machine learning to accelerate innovation. Alchemite™ originated at the University of Cambridge and development is on-going at Intellegens, in close collaboration with our growing community of Alchemite™ customer organisations. These represent sectors including alloys, additive manufacturing, aerospace, batteries, ceramics, chemical processes, composites, consumer products, cosmetics, drug discovery, energy, food and beverage, formulated products, paints, plastics, and printing technology.

www.intellegens.com | info@intellegens.com | @intellegensai