WHITE PAPER

The smart blend: Next-gen mixture formulation

Optimize mixtures-of-mixtures using AI

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

Executive Summary

Machine learning delivers outstanding benefits in the design and analysis of formulations, chemicals, materials, biopharmaceuticals, and beyond. Often the components for these products are themselves mixtures and it is the relative amounts of these underlying ingredients that drives the final product's performance. Yet many experimental design tools do not adequately account for this 'mixture-of-mixtures' scenario. In this white paper, we demonstrate how **Alchemite™** machine learning offers a workflow that allows designers to specify the available feedstocks, then seamlessly optimize feedstock levels, all whilst retaining the scientific knowledge about the underlying ingredients. This enables users to design practical formulations with confidence that they will perform.

Introduction

The **Alchemite™ Suite** software allows scientists to design experimental campaigns using dozens or hundreds of ingredients, leveraging advanced machine learning and providing detailed insight into formulated products, chemicals, materials, and biopharmaceutical systems. However, formulations are often mixtures of complex components; for example, components in the food and beverage industry are themselves usually formulated products containing multiple ingredients. The properties of these underlying ingredients may strongly impact the performance of the final product, but many experimental design tools are unable to incorporate this information and so miss out on scientifically relevant knowledge.

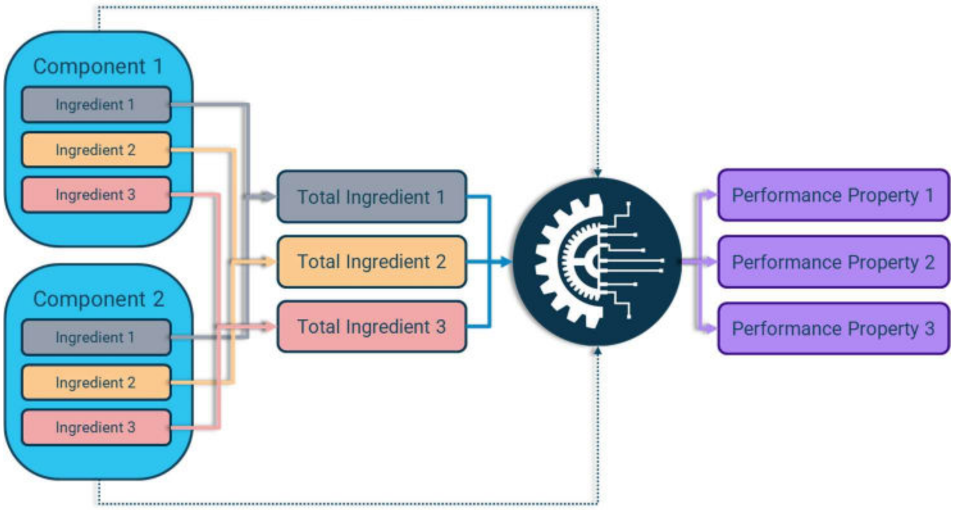


Figure 1. schematic of the Mixtures-of-Mixtures problem, where components available for blending each contain constituent ingredients.

Alchemite™ enables formulation chemists to elegantly and intuitively incorporate information about the ingredients behind their components into their experimental designs and machine learning models, ensuring all relevant information is leveraged. In the case studies below, we show how these 'mixtures-of-mixtures' problems are specified in Alchemite™ using the software's *calculated columns* or *dependent columns* features. We provide examples of how this enables us to design new experiments that more effectively achieve project goals using this increased scientific knowledge.

The added information on underlying ingredients enriches analysis of the usual requirements for a formulation experimental design. These constraints typically include ensuring weight/molar percentages sum to 100%; enforcing practicality by limiting the number of feedstocks used in each formulation, rather than allowing small amounts of dozens of feedstocks; and maintaining ratios of different component or ingredient classes to ensure performance. Machine learning that can explore this complex formulation space is a powerful tool. When it can also tackle mixtures-of-mixtures, it enables next-generation formulation.

Case study 1: Industrial cleaning products

Industrial cleaning formulations often rely on commercial components (e.g., surfactant blends or solvent mixtures) that are themselves composed of multiple active and inert ingredients. The performance of the final product, including cleaning efficiency, foamability, and regulatory compliance, is driven by the underlying ingredients rather than the blends alone.

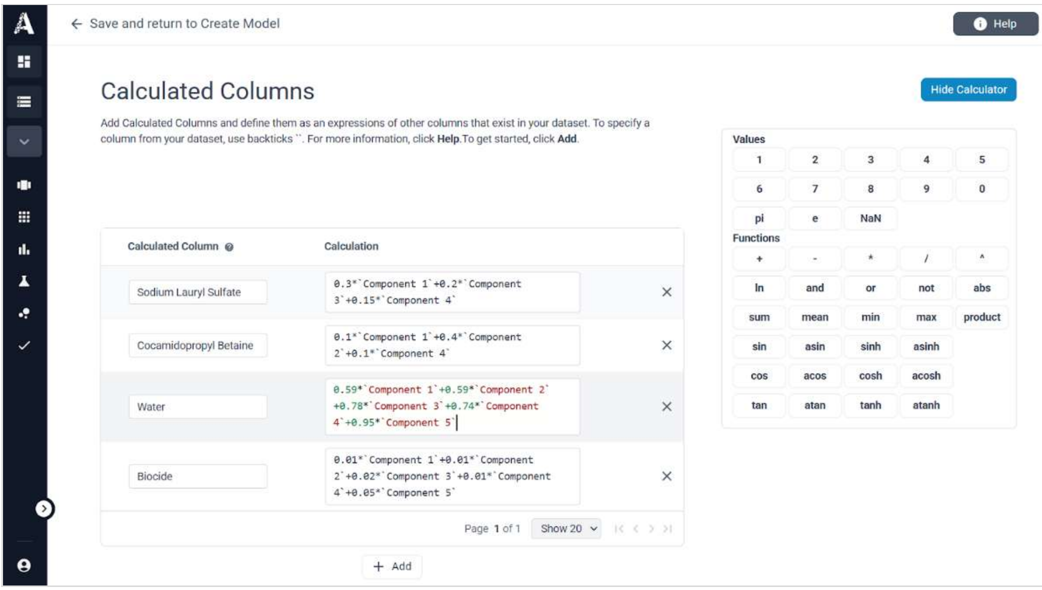


Figure 2: Creating calculated columns to encapsulate the composition of each component.

To capture the composition of each component and the effect on the total formulation, we use the Alchemite™ *calculated columns* feature to input the makeup of each component. In Figure 2, we show how to provide Alchemite™ with the composition of each component: new calculated columns for the total amount of sodium lauryl sulfate and cocamidopropyl betaine (surfactants), water (the solvent), and biocide in each formulation are constructed from the composition of each component. This allows the machine learning model to understand which ingredients truly affect cleaning performance, while keeping the formulation space that chemists need to explore constrained to components available from suppliers.

In Figure 3 we see the *importance table* for the Alchemite™ model trained from this data using the calculated columns. The model has identified that the most important variables for predicting the output properties (the darkest colour cells) are the total ingredient amounts, with the amount of each component being less important. This aligns with the chemical intuition that the total amount of surfactant, for example, is likely more important for foamability and cleaning performance than which component the surfactant came from. The model is then applied to identify candidate formulations that optimize the outputs, enabling the research team to focus-in on the best cleaning formulation.

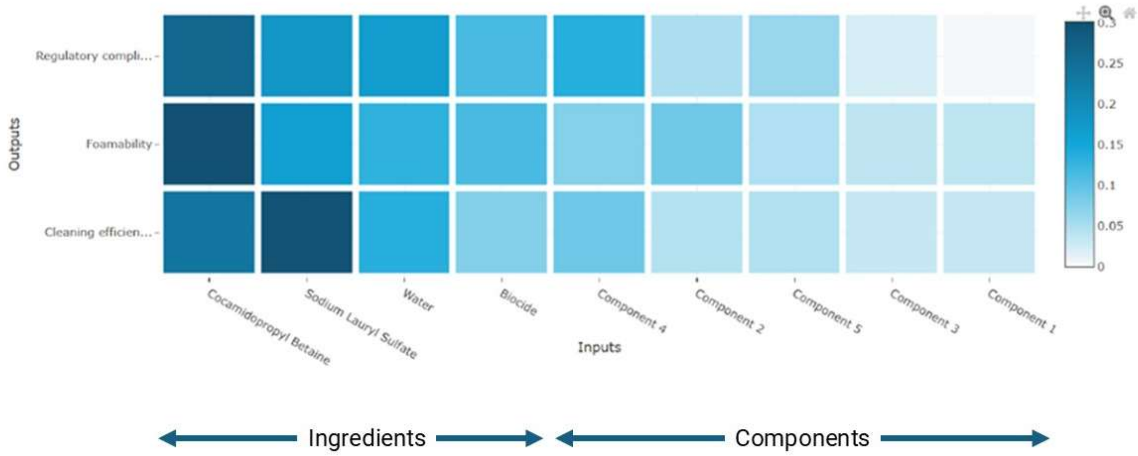


Figure 3: Importance table showing that the total amount of each ingredient is more important than which component it came from.

Case study 2: Food formulation

Many food formulations are made from components that each have their own macronutritional profile. Key properties of the food, including consumer preferences, are strongly influenced by these macronutritional profiles, so it is important to include this information in experimental design. Although these macronutrients are not ingredients in the usual sense, we can apply the mixtures-of-mixtures approach to capture this vital information.

We inform Alchemite™ about the macronutritional profile of each component as the machine learning model is being trained using the *calculated columns* feature, similarly to the previous case study. For example, we can specify three calculated columns representing macronutrients, “Protein”, “Fat”, and “Carbohydrate”, each of which is present in each of the components. The calculated columns represent the total available amount of each macronutrient in the final formulation, which is expected to make a significant impact on the user satisfaction with the final food product. The calculation of the available amount of each macronutrient can be more complex than a simple sum of contributions from each component. For example, a percentage of the protein may be denatured during processing, or components may interact in a way that affects the bioavailability of certain macronutrients. Once the model is trained, the user can immediately use the nutritional profiles of the feedstocks without needing to adjust their workflow.

The Alchemite™ *Formulator* page automatically calculates the total macronutritional profile of the components using the relationships that a food scientist specifies and uses this to inform the model's predictions. The macronutritional profiles are shown (Figure 4) in purple on the right of the Alchemite™ *Predict* page, alongside machine learning predictions for the properties of the final product (consumer scores of overall enjoyment, taste, and sweetness) in green. Thus, information on these critical factors informs experimental design from an early stage.

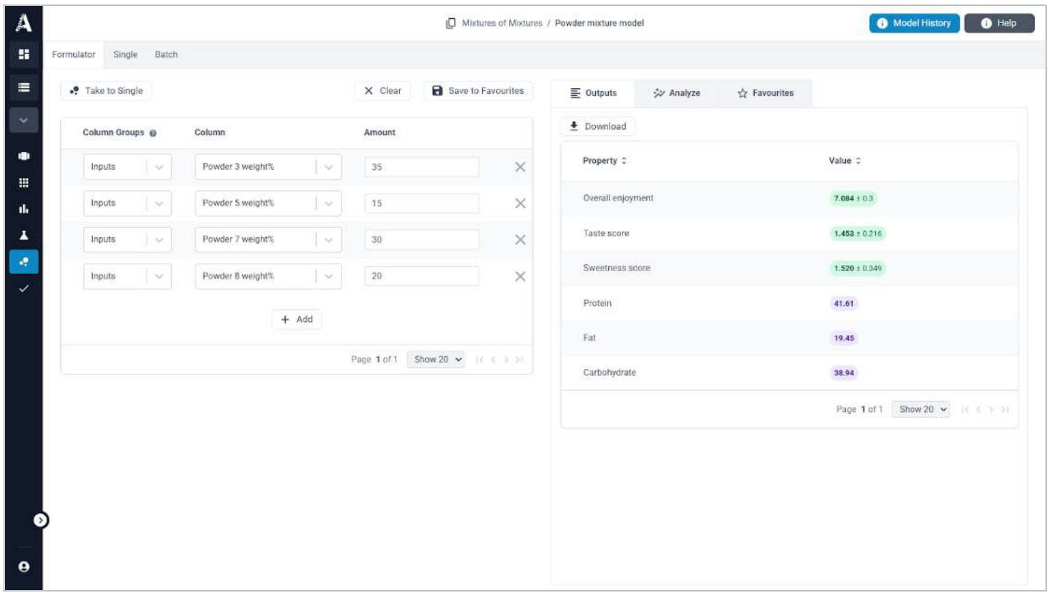


Figure 4: the Formulator page includes calculated total macronutritional profiles.

Case study 3: Recycled feedstocks

With a growing understanding of the need for circularity in material lifecycles, the use of recycled feedstocks in material blends is only increasing. For example, the use of recycled metal components as melting stock in foundries brings challenges as the underlying elemental composition of the recycled feedstock changes on a daily or even hourly basis.

If the foundry has set performance goals for the final metal product, then there is a need to rapidly select which additional raw elements will supplement or compensate for the recycled feedstock. For every new recycled feedstock Alchemite™ material optimization will in seconds propose the combination of additional raw elements to achieve foundry performance goals.

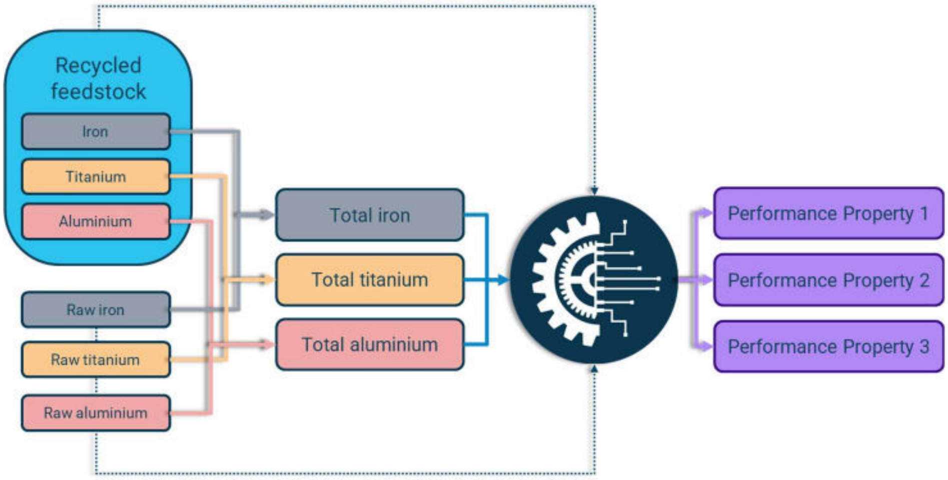


Figure 5: schematic of the combination of a recycled feedstock containing different components with raw elements to find the optimum performance.

In this mixtures-of-mixtures problem the feedstock has a variable composition, depending on which recycled material is available on a given day. This means that we can't use the *calculated columns* approach from the previous case studies, as those set a constant group of relationships. Instead, the elemental composition of the feedstock can be added as a *dependency* in the *Design Experiments* workflow. This way the content of the recycled feedstock can be adjusted on-the-fly, providing interactive access to the recycled material composition. Alchemite™ handles the evaluation of the underlying ingredient amounts automatically as part of the Design Experiments process, identifying the right combination of raw elements to achieve project goals by modelling the contribution of each element, whether it came from the feedstock or from the added raw materials.

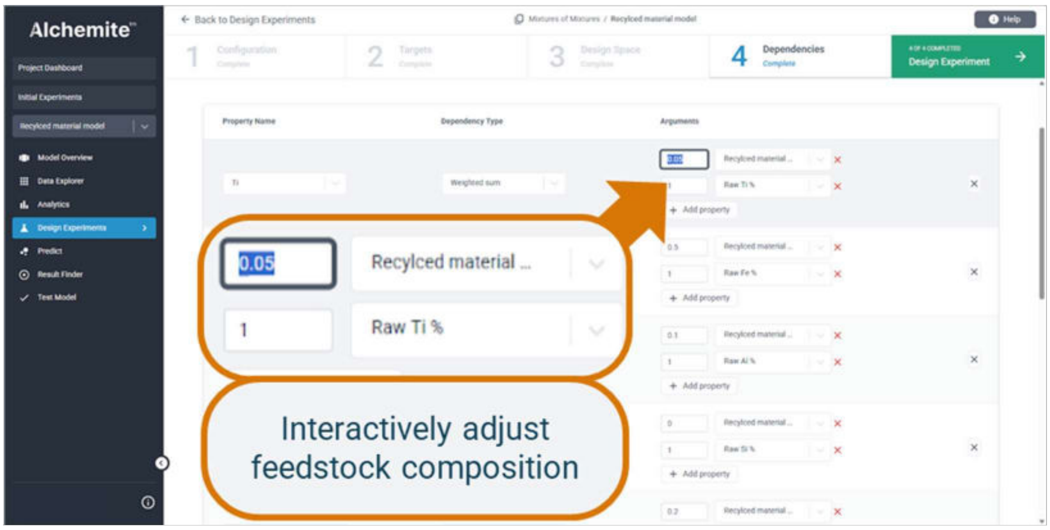


Figure 6: setting the feedstock composition dynamically on Design Experiments as new recycled feedstock becomes available.

Conclusion – Workflow benefits

The ability to mix components improves formulation workflows in three key ways:

Formulations fit for the factory: Design a formulation considering not only its component parts but also underlying ingredients. Constraints can be set on the number and type of components to ensure the formulation is readily manufacturable and can easily pass through to production.

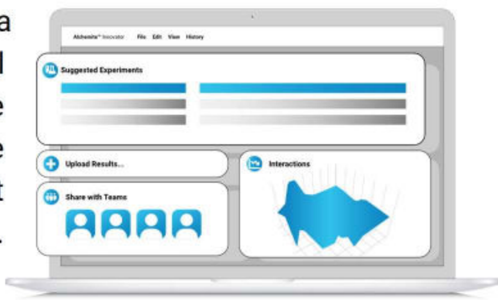
Recycling: Use of feedstocks derived from scrap is an important avenue to drive sustainability. Feedstocks vary day-to-day and require rapid adjustment of additional elements to enable consistent product performance. Therefore, a concurrent design machine learning tool that can adapt designs to incoming feedstocks is game changing.

Chemical insights: Formulation components go beyond the practical, they also represent groupings of alike ingredients. The underlying model understands the science in terms of ingredients and then communicates it in terms of families of ingredients to deliver insights into active groups of compounds, knowledge that can be transferred across a plethora of chemistries. This could, for example, connect active compounds to their catalysts within a single component.

Alchemite™

Alchemite™ Suite is a range of easy-to-use R&D tools, each focused on a key challenge for R&D managers, scientists, experimentalists, or data scientists. Give the right app to the right team member, speeding and informing their work. Then share results and collaborate across your team, creating an integrated machine learning solution for your R&D organization.

Alchemite™ Innovator combines predictive tools with a quick and easy method to design experimental programmes, for a complete project toolset. Using the powerful Alchemite™ method, you can instantly generate a machine learning model from your data, even when that data has gaps or is noisy, where other ML methods fail. Then apply the model to empower your research:



- Enrich and explore your data, filling gaps and finding key relationships
- Guide data acquisition, identifying which experiments offer the fastest route to solve your R&D problems
- Predict likely outcomes for new formulations, testing your ideas before investing in expensive experiments and focusing in on optimal solutions.

About Intellegens

The **Intellegens** vision is that machine learning will drive innovation and deliver value wherever data is used in R&D. Intellegens aims for best-in-class, easy-to-use machine learning software for data analysis in chemicals, materials, life science, and manufacturing. The Alchemite™ technology originated at the University of Cambridge and has been further developed and applied across these industry sectors by Intellegens since 2017.

www.intellegens.com | info@intellegens.com