Research Article **Open Access**

Presenting SLAMD - A Sequential Learning Based Software for the Inverse Design of Sustainable Cementitious Materials

Christoph Völker[1*], Benjami Moreno Torres[1], Ghezal Ahmad Zia[1], Rafia Firdous[2], Tehseen Rug[3], Felix Böhmer[3], Dietmar Stephan[2] and Sabine Kruschwitz[1,4]

[1]Bundesanstalt für Materialforschung und -prüfung, Unter den Eichen 87, Berlin, Germany

[2]Technische Universität Berlin, Department of Civil Engineering, Building Materials and Construction Chemistry, Gustav-Meyer-Allee 25, Berlin, Germany

[3]Iteratec GmbH, St.-Martin-Str. 114, Munich, Germany

[4]Technische Universität Berlin, Department of Civil Engineering, Non-destructive testing of building materials, Gustav-Meyer-Allee 25, Berlin, Germany

Abstract

The number of components in concrete has increased in recent decades - especially in formulations with a reduced carbon footprint. Through the type of binder, supplementary cementitious materials, activators, concrete admixtures, recycled aggregates, etc., attempts are made not only to improve the material properties but also to reduce the ecological and economic impact of concrete as the most widely used material of humankind. Cementitious materials are nanoscale materials. This is accompanied by a more inconsistent composition of raw materials, which makes an experimental tuning of formulations more and more necessary. However, the increased complexity in composition presents a challenge in finding the ideal formulation through trial and error. Inverse design (ID) techniques offer a solution to this challenge by allowing for a comprehensive search of the entire design space to create new and improved concrete formulations. In this publication, we introduce the concept of ID and demonstrate how our open-source app "SLAMD" provides all necessary steps of the workflow to adapt it in the laboratory, lowering the application barriers. The intelligent screening process, guided by a predictive model, leads to a more efficient and effective data-driven material design process resulting in reduced carbon footprint and improved material quality while considering socio-economic factors in the materials design.

Keywords

Sustainable concrete, Machine learning optimization, Inverse design techniques, Scientific software, Data-driven material design

Introduction

Concrete is the most prevalent construction material with an estimated global production of around 30 billion tons annually. However, its production significantly impacts the environment due to the high carbon footprint associated with cement manufacturing. To address this issue, there has been a recent push to develop sustainable alternatives to traditional concrete through the incorporation of various admixtures, recycled materials, and other additives.

However, the increasing number of constituents in concrete formulations and the inconsistent quality of raw materials present challenges in the development of optimal concrete formulations. The empirical materials development process for concrete is further complicated by the vast number of potential material configurations, making the identification of ideal formulations a daunting task.

This paper introduces SLAMD [1], a bespoke software that leverages ID

techniques. ID is a computational method that determines optimal material compositions to achieve specific properties. This data-driven approach mitigates the risk of overlooking relevant designs and enhances the likelihood of discovering the ideal materials for a given scenario. It explores the entire material space to uncover novel concrete formulations that surpass empirically known ones. Despite its demonstrated potential in expediting the design of cementitious materials in our previous work [2-4] and its suggested use in designing ultra-high performance concrete mixes [5], its practical application has been hindered by the profound understanding of mathematics, programming, and material chemistry it necessitates. SLAMD aims to bridge this gap by providing user-friendly graphical interfaces, making the ID process accessible to laboratory personnel, even those lacking extensive computational expertise. By transforming existing knowledge of material properties and compositions, SLAMD enables users to intuitively discover new concrete formulations. Furthermore, SLAMD integrates metrics such as carbon footprint and cost-effectiveness into the material design process, offering a more comprehensive approach to concrete formulation design. This makes it particularly useful in areas where rapid progress is crucial, such as the market adaptation of multicomponent formulas for construction materials.

This paper does not aim to benchmark ID as a methodology, but rather to introduce SLAMD as an innovative tool that brings the ID approach within the grasp of the wider building chemistry community. It elucidates the design concepts behind the software and provides a walkthrough of its workflow, aiming to empower professionals in the field to harness this potent tool in their work.

A data-driven approach can significantly impact the concrete industry by streamlining the material development process. This allows for more flexible resource utilization and the development of new markets for sustainable materials, including locally available or waste-derived ones. Diverse resource streams can broaden the supplier base and alleviate the pricing pressure created by reliance on a limited number of suppliers. Moreover, it becomes feasible to develop new concrete formulations that are not dependent on any specific raw material, facilitating supplier switching with minimal disruptions. This could ultimately lead to a relaxation of the supply chains and a reduction in price pressure, benefiting both manufacturers and consumers.

Additionally, this approach can help mitigate the environmental impact of the concrete industry by reducing the need for transportation and logistics of raw materials from distant locations, thereby minimizing the environmental footprint and enhancing sustainability. Stakeholders such as architects, engineers, and contractors can also benefit from the ability to tailor concrete compositions to specific project requirements, improving the performance of the final product and reducing the environmental impact.

Literature Review

The integration of machine learning (ML) techniques and evolutionary approaches in material design has been a growing research area, with increasing computational power allowing for the prediction of material properties for previously untested compositions [6]. Artificial Neural Networks, Support Vector Machines, decision trees, and evolutionary algorithms, have been successfully applied to predict concrete properties [7]. However, conventional ML techniques typically require large amounts of data for training, which can be costly and time-consuming to collect.

To overcome this limitation, Sequential Learning (SL) and the closely related Bayesian Experimental Design have been proposed as an iterative and adaptive algorithm in other materials science domains, for example, as discussed in the works of Reyes and Powell [8] and Lookman et al. [9]. SL was used to accelerate materials discovery, with up to 20 times faster speedup compared to random acquisition methods, for example as demonstrated by Rohr et al. [10]. Further, optimization of composition and processing for materials with desirable characteristics was studied using SL. It was found to be three times more efficient at finding optimal solutions compared to random guessing on average across multiple domains, as demonstrated by Ling et al. [11].

Previous work on building materials

In the context of building materials, SL has been used to optimize alkali-activated binders and found to be highly effective [2]. Völker et al. searched for binder formulations with good compressive strength. It has been shown to require up to 60 times less data and process more than three times as many features as conventional ML methods, indicating superior performance under real world complexity. Moreover, integrating known material parameters as a knowledge-based loss term has been proposed by Von Rueden et al. [12] to guide the search for materials with environmentally friendly properties. In a second case study Völker et al. [3] showed, that the inclusion of a-priori information in the optimization can enhance the performance of SL algorithms. The authors argue that a-priori knowledge is often available at lower cost than laboratory data. They provide evidence that eco-friendly optimization in terms of a reduced CO_2 footprint does not necessarily come at the expense of poorer mechanical properties.

To conclude, the potential of data-driven material design with SL has been demonstrated across various materials science domains, with promising applications in building materials to optimize composition and processing. As SL can be utilized with few training data and low application thresholds, which are already commonly available in research laboratories, this technology can offer a valuable solution to accelerate the development of environmentally friendly building materials.

Despite the benefits of SL, there is still a lack of awareness and expertise in integrating it into domain-specific tasks to make informed design decisions and keep track of complex data. To address this, the following chapters will provide an overview of how SL can be utilized in an ID workflow to develop sustainable materials in a data-driven fashion. The SLAMD app, which incorporates a digital laboratory and AI optimization capabilities, will also be introduced as a tool to facilitate the implementation of this approach in the design of building materials.

The ID Method

Within the realm of material design, two primary strategies are generally employed: traditional forward design and ID [13]. The forward design approach relies on both empirical studies and theoretical understanding to ascertain properties from the material's structure and composition. Its application to cementitious materials usually follows a prescription-based process, focusing on a specific formulation without broad considerations for other parameters, such as environmental impact or cost.

ID, however, takes an opposite trajectory. It begins by outlining desired properties and subsequently identifies the most suitable formulation to realize these properties. This strategy harnesses computational models and predictive algorithms, enabling the simultaneous evaluation of countless possible formulations. ID is especially significant in the realm of sustainable materials development due to its wide-ranging applicability and optimization potential [14].

Central to ID is the utility function. This mathematical construct helps streamline the material selection process by evaluating numerous factors concurrently. It examines each formulation's predicted properties, associated uncertainties, and additional information such as cost and availability, ultimately assigning each material a comprehensive utility 'score'.

Take, for instance, the forward design of a concrete mix. Traditional parameters – such as the water-to-cement ratio and the powder content – are adjusted until the mix achieves the desired workability and strength. While effective in these respects, this approach may not adequately account for other critical aspects, including cost and durability [14].

When dealing with more complex scenarios – for instance, choosing from a wide range of supplementary materials for a specific formulation – the forward design becomes less effective. Here, the utility function of ID provides an invaluable tool, offering an optimal balance between various considerations such as cost-efficiency, workability, durability, and strength.

The utility function quantifies each formulation's overall performance considering these criteria, enabling a more objective comparison. The formulation yielding the highest utility score is then selected for further investigation. In this manner, the utility function facilitates a systematic evaluation of numerous material formulations, ensuring the selection of the most suitable composition in light of the given criteria.

For a more comprehensive understanding, the works of Völker et al., 2021 and 2023 provide in-depth insights [2, 4].

The workings of ID take place in a specific space known as design space (DS). Although the concept of the DS may be foreign to construction chemists, it proves to be highly useful for formalizing the full range of possibilities in the design process. The DS, a high-dimensional vector space, represents each material formulation by its parameters or properties. Essentially, the DS encapsulates each material formulation and represents it based on its composition and processing parameters. As seen in the two-dimensional representation in figure 1, similar materials are located close to each other in this high-dimensional space. The color coding is used to represent the relative performance of the materials, with pink representing high performance and turquoise representing lower performance. This structured representation plays a central role in explicit exploration and discovery in ID and distinguishes it from traditional, forward design approaches. The right panel of figure 1 shows how the adaptive sampling of the ID approach optimizes this exploration process.
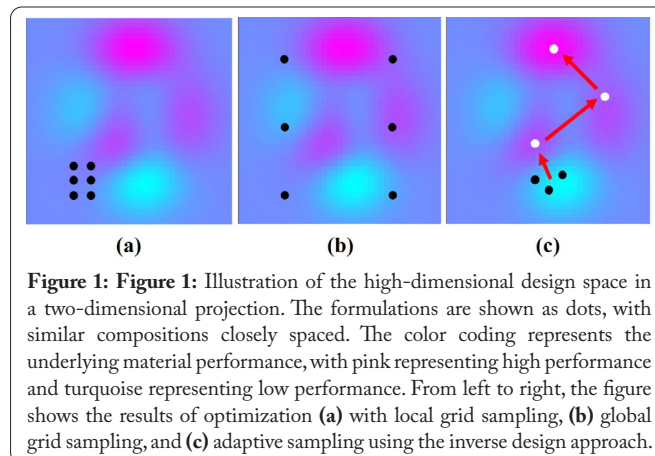

Figure 1: Figure 1: Illustration of the high-dimensional design space in a two-dimensional projection. The formulations are shown as dots, with similar compositions closely spaced. The color coding represents the underlying material performance, with pink representing high performance and turquoise representing low performance. From left to right, the figure shows the results of optimization **(a)** with local grid sampling, **(b)** global grid sampling, and **(c)** adaptive sampling using the inverse design approach.

Distinct from traditional forward design, ID employs an adaptive, iterative process. While forward design aims to collect representative samples to establish comprehensive knowledge for ideal formulation, it can falter with complex compositions, potentially missing crucial samples within a limited laboratory budget (either through fine but local sampling as in figure 1a or coarse but global sampling as in figure 1b). ID, on the other hand, focuses on the "right" data, utilizing each iteration to optimize the selection of subsequent experiments within the DS. This approach reduces experimental volume and enhances the discovery of optimal formulations.

The ID workflow, detailed in figure 2, begins by first defining a comprehensive DS that includes all possible material recipes. Variables such as cement admixtures, aggregates, powder blends, and additives create each recipe. Further refinement integrates each formulation's composition and processing information into the DS, enabling a ML model to interpolate effectively between similar mixtures.
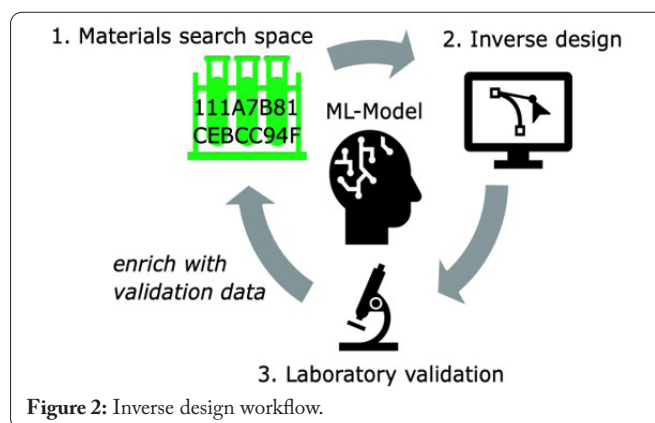
The second phase of the ID workflow incorporates ML

Figure 2: Inverse design workflow.

models, specifically non-parametric types such as Gaussian process regression (GPR) or Random Forest regression (RFR). These models can be trained even with a small set of initial laboratory validations (even fewer than ten). While the predictions of material properties within the DS may lack accuracy at this stage, the models are invaluable for identifying regions of interest. The utility function comes into play here, spotlighting the most promising formulations (so far) by considering not only the predicted performance but also model prediction uncertainties and factors like environmental footprint, resource efficiency, and cost.

Previous work provides clear guidance for the design process and has shown that the volume of training data required for data-driven design may be far less than often suggested [4]. The primary focus should be on establishing an effective ID framework that maximizes knowledge gain, rather than developing highly accurate models.

Interestingly, ID leverages model uncertainties or entropy in material designs which makes it possible to deliberately discover novel, and non-oblivious materials, striking an optimal balance between material properties and cost. Once validated, the top-performing formulations are added to the training dataset to update the predictive model in the third step. This iterative process—spanning the DS, training the ML model, and updating it—continues until the desired performance criteria are achieved.

The ID approach offers several key advantages over traditional methods. It combines the low cost and quasi-unlimited experimentation capabilities of computational methods with the empirical certainty of experimental methods, all while significantly reducing the number of expensive experiments. Additionally, the integration of cost and environmental factors into decision making has the potential to revolutionize the material discovery and optimization process by enabling the identification of innovative materials with superior properties and lower environmental impact that may not have been detected using more conventional approaches.

Introducing SLAMD

SLAMD is a software research tool that combines a Digital Lab and AI optimization to enable the discovery of new concrete recipes (see figure 3). The Digital Lab provides a framework for describing resources and processes and their socio-economic impacts and automates the creation of complex concrete recipes with detailed material knowledge. The AI optimization component integrates data from the Digital Lab and applies it to new concrete formulations.

The app is based on a 3-layer architecture using Python (Flask) on the backend. The backend utilizes libraries such as Pandas, Scikit Learn, and Lolo Machine Learning to provide data analysis and machine learning functionalities. To protect user privacy, all data is stored in a server-sided sessions and is not accessible to other users. The frontend of the software is built with Vanilla JavaScript and Bootstrap, providing a user-friendly interface. SLAMD can be deployed locally as a python application or as a web application, making it acces-

Figure 3: SLAMD landing page with navigation bar (top) and overview over modules below.
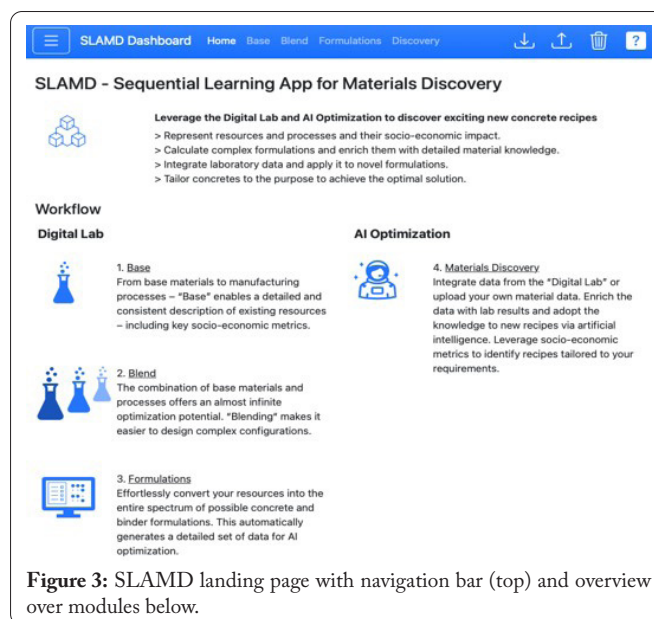
sible on any device, suitable for both personal and teaching or lab use. To install the app locally, the source code must be downloaded and executed via the python command line. This process is simple and can be done even by inexperienced users as it only requires entering a few commands in the python console. For installation details, a readme file is included in the application source code repository.

Workflow

The design of concrete recipes follows a hierarchical structure, starting with the cement content/supplementary cementitious material content and water-to-cement (W/C)-ratio and followed by the admixture content (superplasticizer, plasticizer, retarders, etc.) and aggregate content. Changes made at the top level (cement content, W/C-ratio) will affect the bottom levels (admixtures content and aggregate content). This hierarchical structure can make manual design challenging at scale, but SLAMD resolves this through an effective workflow that minimizes the need for manual intervention. The workflow consists of four successive steps: Base, Blend, Formulations, and AI-optimization each of which is described in detail below.

Base: creating base materials and processes

The first step in the Digital Lab workflow is the creation of base materials and processes. There are six different types: Powder, Liquid, Aggregates, Admixture, Process, and Custom. Each of these options has distinct properties and is treated differently in the formulation creation process. The "Custom" type allows for the specification of a new material type with no predefined properties outside of cost, relying instead on custom properties to carry information. Figure A1a in the appendix shows the user interface with the input fields for powders, including cost, composition, and additional properties. These can be extended by freely adding custom properties. All base materials and processes are structured in a table and can be edited at any point (see figure A1b).
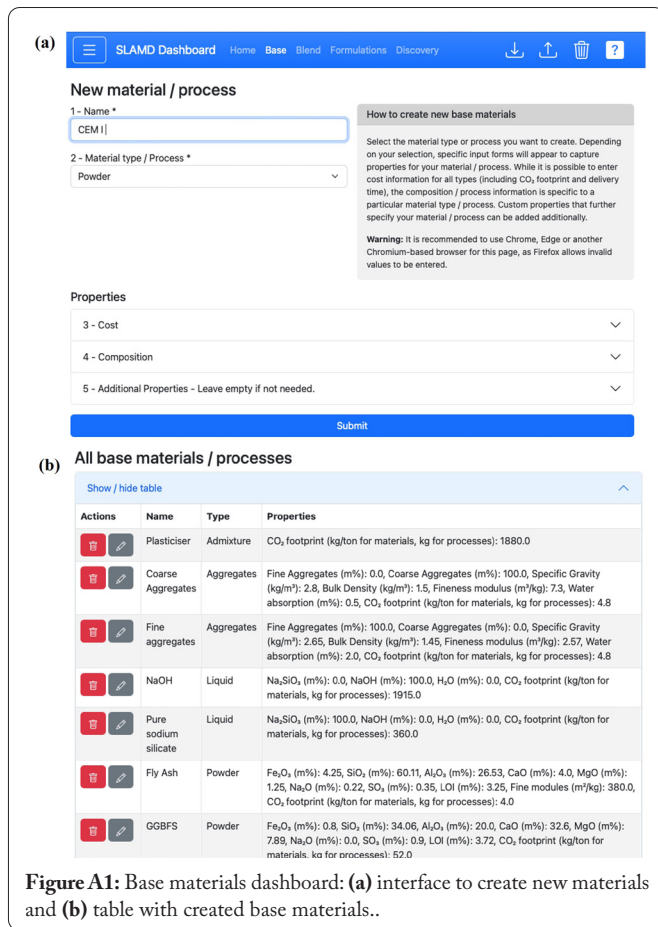
Figure A1: Base materials dashboard: **(a)** interface to create new materials and **(b)** table with created base materials..

Blend: creating blended materials

The next step is to create blended materials by selecting at least two base materials of the same type. Blended materials are mixtures of base materials, and their properties are a weighted average of the properties of the constituent base materials, except for "Delivery time," which is defined as the maximum of the base materials' delivery times. In other words, the blending function allows the user to easily augment costs and properties at the individual precursor level with the properties of many blended mixtures at once (e.g., activators, liquids, or powder blends). The composition of the blended materials is specified by the ratios of the starting materials. Multiple blends can be created at once using the "Increment" input.

Formulations: creating a concrete and binder search space

The "Formulations" step in SLAMD allows for the conversion of resources into concrete and binder formulations, creating a comprehensive DS for AI optimization. A formulation consists of selected materials and processes and must contain a powder, a liquid, and an aggregate for concrete formulations, and just a powder and a liquid for binder formulations. Additional materials and processes can be added freely. The assembly of these constituents is visually detailed in figure A2a in the appendix. The total weight of the mix of materials in the formulation needs to be specified. For example, 2400 kg could be used to lay out formulations for roughly one cubic meter of concrete. Weights and W/C (determining the ratio of liquid-to-powder) can be specified using the input

fields (compare figure A2b). They fulfil the constraint given by the total weight specified for the mixture. To simplify usability, this is enforced by appropriate autocompletion functions. The liquid-to-powder ratio is used to specify liquids. All newly generated material configurations can be edited to manually tune the formulations if needed.

AI optimization for materials discovery

The "Materials Discovery" component in SLAMD allows the user to configure the inverse design process by selecting a previously created DS or uploading a custom .csv. The user
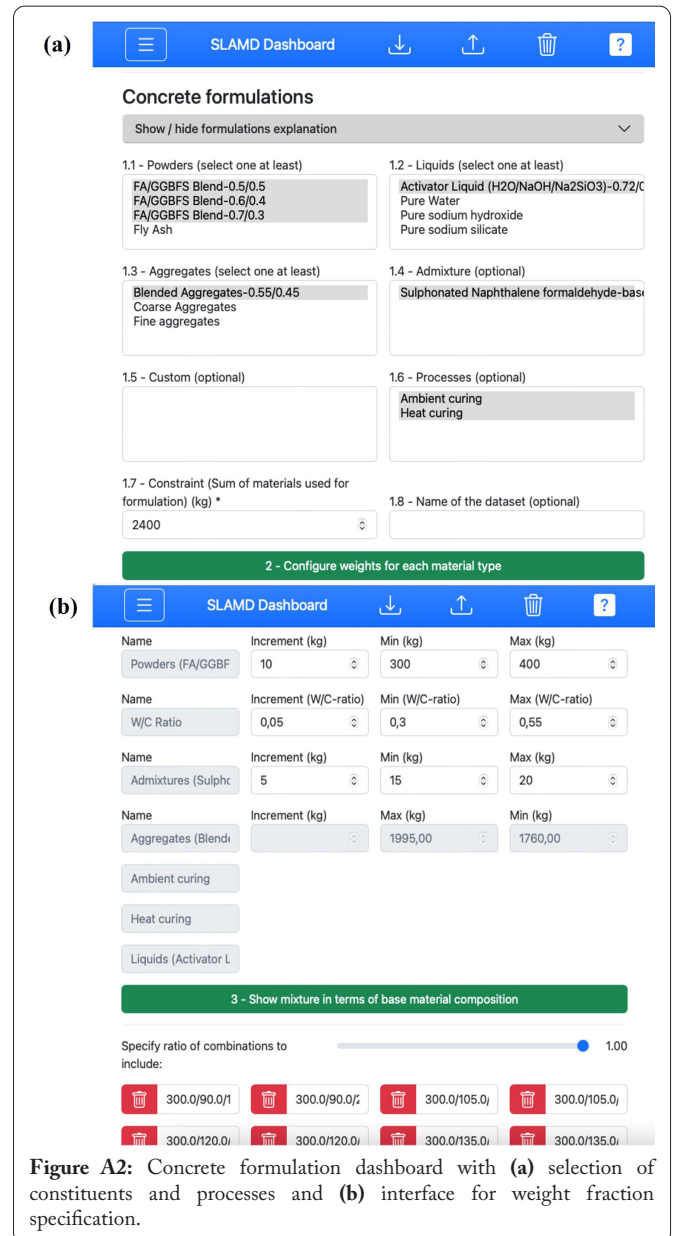

Figure A2: Concrete formulation dashboard with **(a)** selection of constituents and processes and **(b)** interface for weight fraction specification.

can delete, select, add target information, or download the DS from the user interface. Figure A3 in the appendix illustrates this process through the top panel, which displays the search space list with its various options. The "delete" option (red bin button) simply removes the DS from SLAMD. If a DS is selected via the blue selection button, the DS is picked to
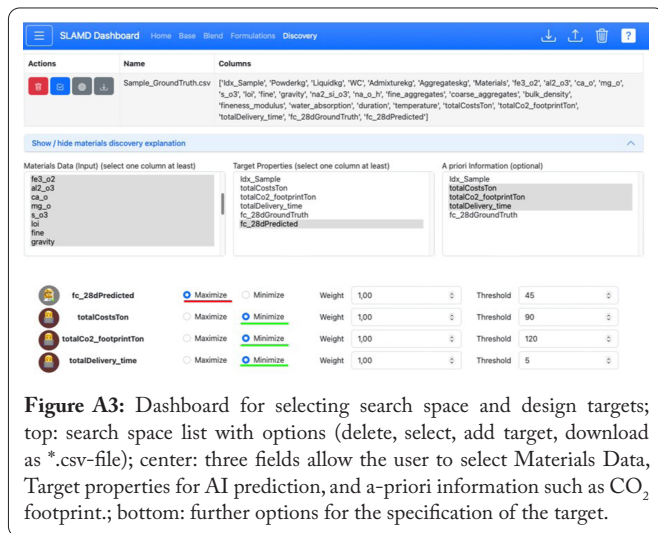
Figure A3: Dashboard for selecting search space and design targets; top: search space list with options (delete, select, add target, download as *.csv-file); center: three fields allow the user to select Materials Data, Target properties for AI prediction, and a-priori information such as CO_2 footprint.; bottom: further options for the specification of the target.

configure the ID problem. The "add target" option (bulls-eye button) opens a window where new columns can be added and gaps in the current data can be filled, making it easy to enrich the existing DS with new information from the lab. This process takes the user to a dedicated form for managing targets. Figure A4 in the appendix illustrates the user interface where new targets can be added, and existing ones can be enriched. The changes can then be saved using the fourth action button: "download".
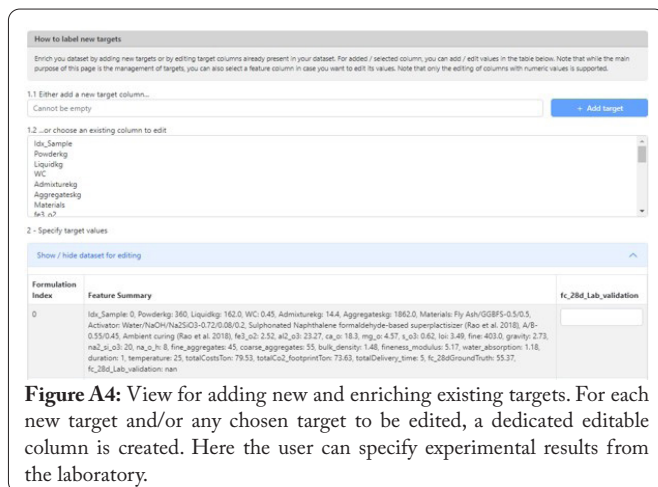

Figure A4: View for adding new and enriching existing targets. For each new target and/or any chosen target to be edited, a dedicated editable column is created. Here the user can specify experimental results from the laboratory.

Next, the user selects relevant features of the mixtures, such as chemical composition, manufacturing process, microstructure, and other key material characteristics, to define the materials input data. These inputs help define the search space for the machine learning algorithms to find new materials. A carefully chosen set of features ensures that the optimization focuses on the most critical aspects of the material, producing meaningful results (compare figure A3, center).
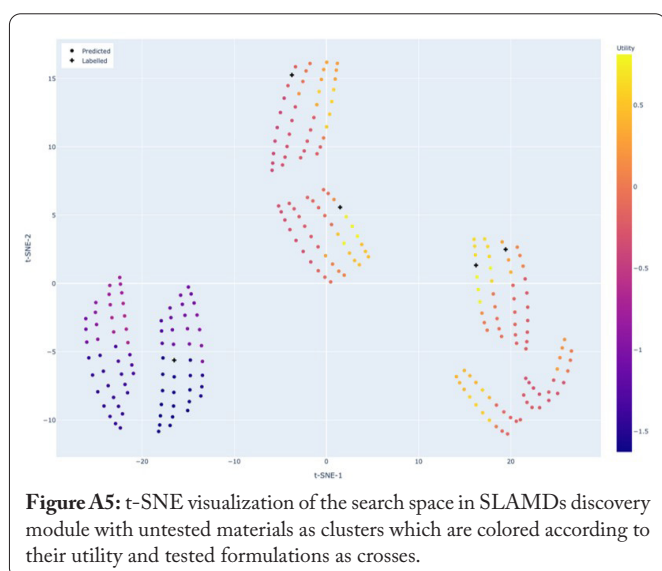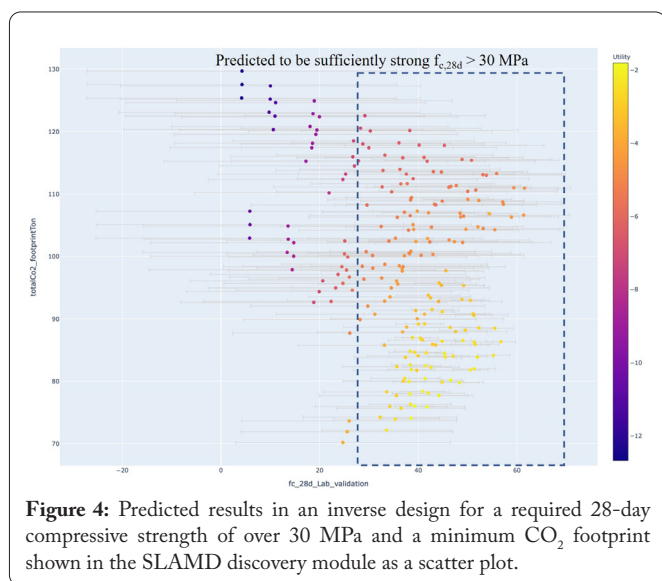
Finally, the design goals are configured, a process that is illustrated in the lowest panel of figure A3 in the appendix and is explained in more detail in the following. The estimated utility function, a mathematical representation of the trade-off between different target properties, is the key component. The utility function aggregates the estimated material performance into a single scalar value for each material formulation, allow-

ing the vast material DS to be explored and the best solution found based on the design requirements. The target properties can be optimized to maximize (e.g., strength and workability) or minimize (e.g., CO_2-footprint and costs) the utility function, depending on the design goals. The user can adjust the target weights to prioritize certain properties over others, resulting in different optimal trade-off solutions that balance conflicting objectives. Additionally, the relevance boundary can be set by specifying a threshold for each target property, such as a strength threshold beyond which improvement does not contribute to the material's utility.

SLAMD offers two prediction models, including GPR [15] and a RFR [16]. Furthermore, an option exists to add a Principal Component Analysis based dimensional reduction as a preprocessing for high-dimensional materials data. These are commonly used models implemented according to our findings in Völker et al. 2023 [4] and are selectable based on the specific characteristics of the data and the objectives of the prediction task. While GPR is suitable for smaller datasets and provides probabilistic predictions, RFR tends to perform better with larger datasets and can handle more complex interactions between variables. Although the choice of the prediction model can be an expert-level decision, our software provides default settings optimized for typical concrete formulation optimization tasks.

The utility for each material in the search space is calculated based on the specified targets and the curiosity value. A higher curiosity value prioritizes predictions with greater uncertainty, while a negative curiosity value favors predictions with low uncertainty. As described in section 3, higher uncertainties can prioritize materials with the potential for high information gain, while lower or negative uncertainties can prioritize more reliable predictions.

After the ID is configured and the prediction is made, a table lists the formulations sorted by utility and predicts target properties. The user can easily select the candidate material with the highest utility score by choosing the highest table entry for validation in the laboratory. Interactive plots further assist in the selection process. The predictions are visually differentiated via color coding based on the estimated utility. An exemplar of this is figure 4. It presents a scatter plot illustrating an ID prediction from the SLAMD exploration module for a formulation required to exhibit a 28-day compressive strength greater than 30 MPa while maintaining a minimum CO_2 footprint. In this illustrated design scenario, the emphasis is on achieving a formulation with a compressive strength surpassing 30 MPa at the 28-day mark. While many formulations meet the strength criteria, the estimated utility also factors in the CO_2 footprint. This leads to higher utilities for materials with lower CO_2 footprints. In a classical forward design, factoring in such information can be challenging. However, in ID, where a comprehensive DS contains all relevant variants of the materials, it can be easily derived, making it available for informed decision making. This makes sustainable design decisions much simpler while maintaining high materials performance.

Figure 4: Predicted results in an inverse design for a required 28-day compressive strength of over 30 MPa and a minimum CO_2 footprint shown in the SLAMD discovery module as a scatter plot.

Figure A5: t-SNE visualization of the search space in SLAMDs discovery module with untested materials as clusters which are colored according to their utility and tested formulations as crosses.

The t-SNE visualization is designed to increase the information gain of the DS. It is a mathematically derived representation of the materials input data, where similar materials are shown close together. The distinct clusters in the t-SNE visualization provide an overview of the material variants in the DS, making it easier to select a diverse set of materials for robust and informative model development. The color-coded utility helps identify the most promising candidates within each cluster. In figure A5, a t-SNE plot of a DS that contains six distinct clusters is shown, each referring to a material and processing type. Although only a few materials have labels (marked as black crosses), it is clear that the materials on the left offer little utility, helping scientists to focus on the more relevant formulations in the center and right.

Conclusion

The quest for sustainable building materials is a pressing challenge, and data-driven approaches offer a promising solution. However, these approaches can be difficult for laboratory personnel to implement and execute. SLAMD solves this is-sue by providing a user-friendly software solution for designing sustainable concrete recipes and materials. This innovative approach combines the power of ID with the precision of machine learning to identify materials with superior properties and reduced environmental impact.

SLAMD streamlines the design process and provides a comprehensive framework for creating complex materials, defining goals, and finding optimal material formulations through its interactive user interface. The effectiveness of this approach has been demonstrated in numerous case studies across various material domains, including building materials. By reducing the number of necessary experiments, SLAMD enables the discovery of non-obvious solutions and ultimately reduces the cost and environmental impact of materials development.

As technology continues to evolve, we anticipate that data-driven design tools such as SLAMD will become increasingly prevalent in the coming years. However, the successful implementation of these tools will require collaboration between researchers, engineers, and software developers to integrate them into existing laboratory workflows and effectively disseminate knowledge. If done effectively, the implementation of novel laboratory workflows can lead to a paradigm shift in materials design, revolutionizing the capabilities of laboratory personnel.

To fully embrace the opportunities of data-driven tools such as SLAMD in materials development, a change in mindset and working practices is required. This includes a shift toward a more holistic view on building materials, but also, in very practical terms, the establishment of new workflows and systems, such as electronic laboratory notebooks and databases, that facilitate the capture, storage, and sharing of data in a standardized and accessible format. This will enable the effectiveness of data-driven techniques such as inverse design to be increased across a wide range of materials and use cases.

Open science and data and software sharing is a key factor in the successful adoption of data-driven approaches. Sharing data and results can help build a more comprehensive understanding of materials and their properties, leading to better decision making and more efficient use of resources. This can also help increase transparency and accountability in the materials development process and lead to more sustainable outcomes.

The open-source nature of SLAMD also promotes collaboration and progress within the science and engineering community, contributing to improved safety and sustainability in the construction industry. The benefits of this technology are far-reaching, positively impacting architects, engineers, and suppliers alike.

In the context of sustainable concrete development, the use of data-driven techniques can already help minimize the environmental impact of the concrete industry by reducing the need for transportation and logistics of raw materials and enabling the development of new concrete formulations that are not dependent on specific raw materials. This can allow switching between suppliers with minimal disruption and ul-

Presenting SLAMD – A Sequential Learning Based Software for
the Inverse Design of Sustainable Cementitious Materials

Völker et al.

timately lead to a loosening of supply chains and a reduction in pricing pressure, which can benefit both producers and consumers.

In summary, the integration of data-driven techniques and the transition to open science and data sharing are critical to the development of sustainable concrete and other materials. By streamlining the design process and reducing the need for experimental validation, data-driven approaches can lead to more efficient and effective materials development that positively impacts the environment, industry, and society.

Acknowledgements

Conflict of Interest

None.

Appendix

None.

References

1. Sequential Learning App for Materials Discovery ("SLAMD") - Web Version. [https://github.com/BAMresearch/WEBSLAMD] [Accessed September 28, 2023]

2. Völker C, Firdous R, Stephan D, Kruschwitz S. 2021. Sequential learning to accelerate discovery of alkali-activated binders. J Mater Sci 56: 15859-15881. https://doi.org/10.1007/s10853-021-06324-z

3. Völker C, Kruschwitz S, Torres BM, Firdous R, Zia GJ, et al. 2022. Accelerating the search for alkali-activated cements with sequential learning. In FIB Conference, Oslo, Norway.

4. Völker C, Torres BM, Rug T, Firdous R, Zia GA, et al. 2023. Data driven design of alkali-activated concrete using sequential learning. J Clean Prod 418: 138221. https://doi.org/10.1016/j.jclepro.2023.138221

5. Saleh E, Tarawneh A, Naser MZ, Abedi M, Almasabha G. 2022. You only design once (YODO): Gaussian process-batch Bayesian optimization framework for mixture design of ultra high performance concrete. Constr Build Mater 330: 127270. https://doi.org/10.1016/j.conbuildmat.2022.127270

6. Li Z, Yoon J, Zhang R, Rajabipour F, Srubar WV, et al. 2022. Machine learning in concrete science: applications, challenges, and best practices. npj Comput Mater 8(1): 127. https://doi.org/10.1038/s41524-022-00810-x

7. Chaabene WB, Flah M, Nehdi ML. 2020. Machine learning prediction of mechanical properties of concrete: critical review. Constr Build Mater 260: 119889. https://doi.org/10.1016/j.conbuildmat.2020.119889

8. Reyes K, Powell WB. 2020. Optimal learning for sequential decisions in laboratory experimentation. arXiv preprint arXiv:2004.05417. https://doi.org/10.48550/arXiv.2004.05417

9. Lookman T, Balachandran PV, Xue D, Yuan R. 2019. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Comput Mater 5(1): 21. https://doi.org/10.1038/s41524-019-0153-8

10. Rohr B, Stein HS, Guevarra D, Wang Y, Haber JA, et al. 2020. Benchmarking the acceleration of materials discovery by sequential learning. Chem Sci 11(10): 2696-2706. https://doi.org/10.1039/c9sc05999g

11. Ling J, Hutchinson M, Antono E, Paradiso S, Meredig B. 2017. High-dimensional materials and process optimization using data-driven experimental design with well-calibrated uncertainty estimates. Integr Mater Manuf Innov 6: 207-217. https://doi.org/10.1007/s40192-017-0098-z

12. Von Rueden L, Mayer S, Beckh K, Georgiev B, Giesselbach S, et al. 2021. Informed machine learning–a taxonomy and survey of integrating prior knowledge into learning systems. IEEE Trans Knowl Data Eng 35(1): 614-633. https://doi.org/10.1109/TKDE.2021.3079836

13. Zunger A. 2018. Inverse design in search of materials with target functionalities. Nat Rev Chem 2(4): 0121. https://doi.org/10.1038/s41570-018-0121

14. Liao TW, Li G. 2020. Metaheuristic-based inverse design of materials–a survey. J Materiomics 6(2): 414-430. https://doi.org/10.1016/j.jmat.2020.02.011

15. sklearn.gaussian_process.GaussianProcessRegressor. [https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.GaussianProcessRegressor.html] [Accessed September 28, 2023]

16. Python Wrapper for Lolo. [https://pypi.org/project/lolopy/] [Accessed September 28, 2023]