PREDICTING THE RECIPE SPACE

Leveraging Parameterless Self-Organizing Maps (PLSOM) and Geometrical Polynomial Fitting (GPF) to predict biopolymer composites characterization with small datasets

Gabriella Rossi[1], Arianna Rech[2], John Harding[3], Paul Nicholas[1], Anders Egede Daugaard[2], Martin Tamke[1], Mette Ramsgaard Thomsen[1]
[1] CITA, Royal Danish Academy (DK) [2] Danish Polymer Center, DTU (DK) [3] University of Reading (UK)

Royal Danish Academy
Architecture Design Conservation
royaldanishacademy.com
CITA
Centre for IT and Architecture

1. Introduction

Problem statement

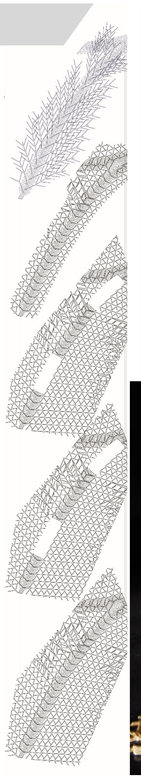
Biopolymer composites are an interesting class of materials to be explored using robotic 3d printing for architectural applications. They provide the opportunity to incorporate biodegradable waste-stream materials into the built environment. Mixtures and recipes can be tuned across the print to satisfy specific performance requirements within the global design.

However, when considering gradual and continuous grading responding to multiple objectives, rather than discrete compositions, two limitations arise:

1. *High dimensional recipe-performance space:* Multi-ingredient variation to multi-performance response mapping becomes exponentially complicated.
2. *Large quantity of experimental samples:* Physical in-lab characterization of continuous recipe permutation becomes exponentially unfeasible due to material and time costs.

Our approach

The project develops an experimental methodology to predict, with sufficient accuracy, the physical performance of all ingredient permutations within a recipe space, using a small physical dataset of lab-samples. By leveraging the associative positioning of Self-Organizing Maps we are able to geometrically fit a Polynomial model which outperforms state of the art predictive models. The low-dimensional mapping also allows us to develop an intuitive interface to navigate the ingredient-performance response and plugs directly within a computational design workflow.

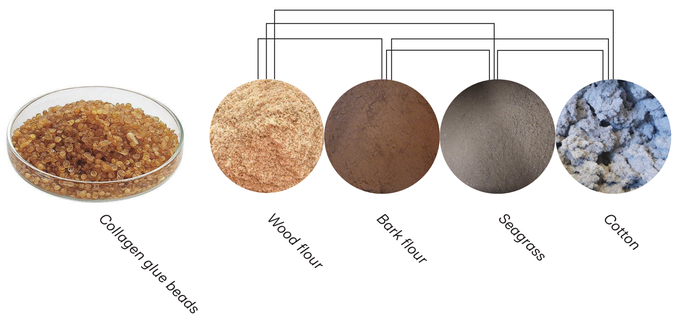



Radicant, CITA (2022)
Living Prototypes Exhibition, AEDES Gallery, Berlin

2. Methods

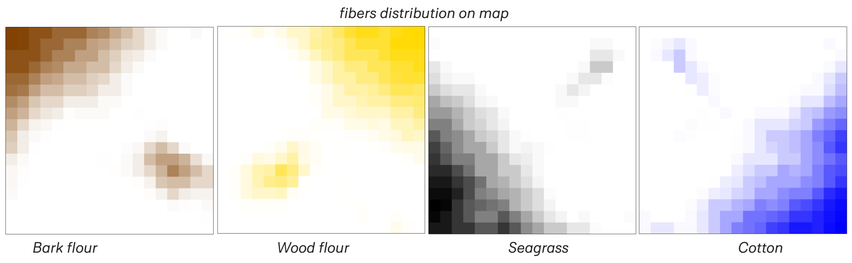
The recipe space

We consider a recipe for a collagen-glue based biopolymer which is mixed with 4 different waste stream fibers: wood flour, bark flour, seagrass and cotton. The Fibers can occupy at most 30% of the material weight, and can at most interact 2 at a time. By working at a granulatiry increment of 3%, a traditional Design Of Experiment approach would yield a Full Factorial matrix of 312 different recipe combinations.



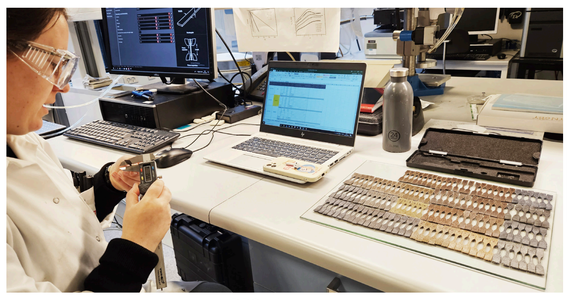
Training the PLSOM

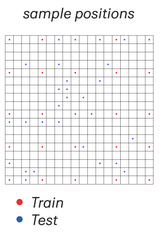
We implement Erik Berglund and Joaquin Sitte's modified version of Teuvo Kohonen's Self-Organizing map. While the SOM is known to struggle with hyperparameter search, the PLSOM with PCA initialization is able to converge , deterministically. We use a Bayesian optimizer to find the best Neighborhood Range β for a given Full Factorial recipe matrix, while maximizing the spread of the combinations over the map nodes, avoiding overcrowding of the nodes, and reducing quantization and topographic error. As a result we obtain an 18x18 map which project our 4D recipe space onto 2D.



Data collection

The map is then used to determine the dataset composition. The training set is composed of 25 recipe combinations that originate from a 5x5 grid sampling of the map nodes. The testing set is composed of 30 radomly selected nodes on the map. Sets of 6 dog bone samples per recipe are compression-molded, and subjected to physical and structural testing in the lab. We report measurements of Youngs Modulus, Tensile Strength, Elongation, Shrinkage and Weight Loss.


DPC laboratory



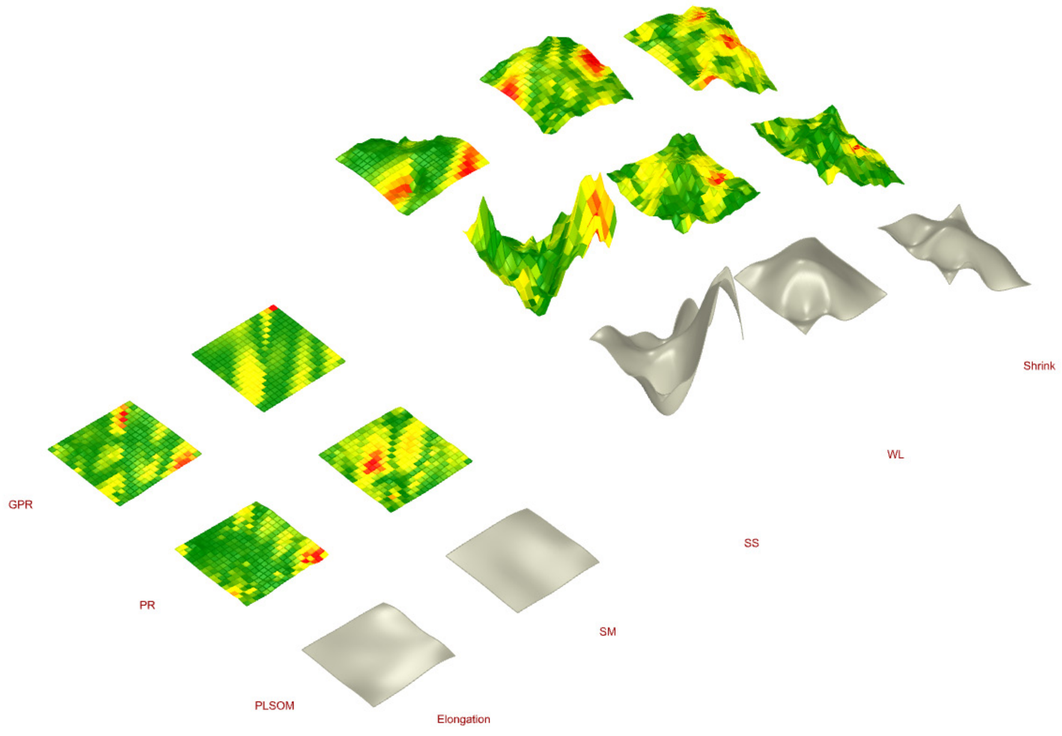
3. Results

Predictive modelling and benchmarking

We use Geometrical Polynomial Fitting (GPF) to create a Response Surface (RS) for each of the measured performance metrics. The sampled recipe "grid point" is constructed with Z being the mean measure, these points are then interpolated into a 3rd degree polynomial curve and a network surface with position continuity is constructed. Additionally we build the standard deviation cloud above and below the RS.

This GPF is benchmarked against the state-of-the-art models for small datasets: SK Learn's Kernel Ridge Regression at 3rd degree polynomial (PR) and a Gaussian Process Regression (GPR). We evaluate 2 metrics:

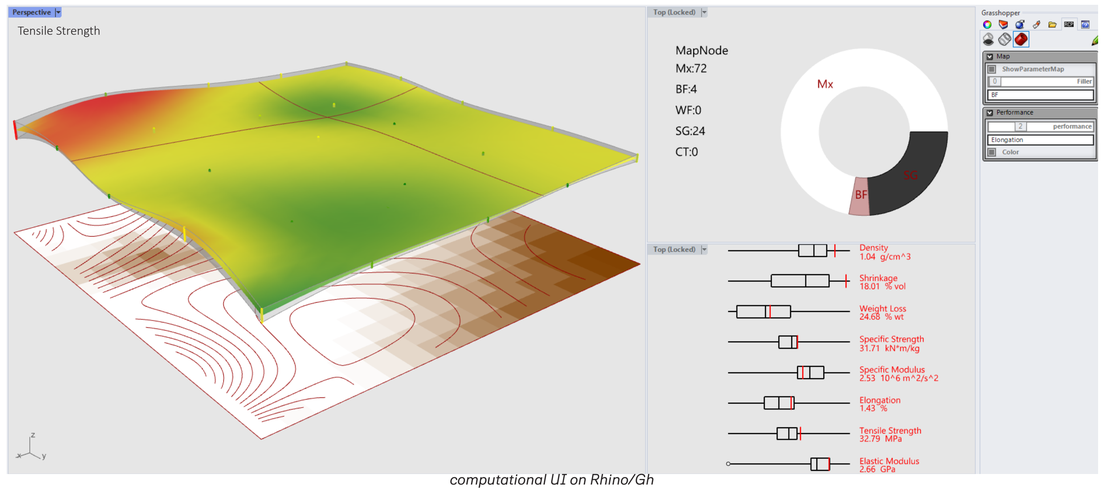
1. *The accuracy of the test data prediction using GPF compared to the ground truth:* 94%, 97% and 77% of respectively Specific modulus, specific strength and elongation predictions fall within the standard deviation of the real measurement.
2. *The accuracy of the train data prediction using PR and GPR compared to the ground truth:* we observe that the GPR fails to capture variation of the data and considers is as noise, while the PR tends to overfit by producing a rough response surface.



Overview of the different performance Response Surfaces obtained by the different predictive models. The color gradient is the distance to the PLSOM surface.

Explorative UI interface

The advantage of utilizing a PLSOM is that it allows to intuitively explore the high dimensional recipe space directly in the design environment. We develop a UI interface where just by hovering over the trained map, the designer can immediately observe the recipe composition, its different performances with respect to statistics of the ensemble, as well as uncertainties related to the data, irrespective of whether the recipe has been physically tested or not. This user-friendly approach also allows for exploration of the performance landscape, multi objective masking beyond the scope of just obtaining one optimized recipe formulation.



computational UI on Rhino/Gh

4. Conclusions

Collecting enough physical samples for dataset composition is one of the main hurdles of using Machine Learning workflows in Material Science. In this poster we showcase an approach which allows to predict the performance of a combinatorial recipe space for biopolymer composites using a relatively low number of experimental lab samples, to an accuracy that falls within the bounds of measured standard deviation among the samples. The approach leverages dimensionality reduction via the PLSOM which allows us to specifically select the data samples to be physically produce which can then be fit into a Response Surface via GPF. Next steps will explore the potential of using the full output of the map as a base dataset for training a neural network, in order to extrapolate performance predictions to recipes that fall outside of the variational bounds of the map.

Project funding:
DANMARKS FRIE FORSKNINGSFOND
Grant n 9131-00034B
Project collaborators:
CITA
DTU
University of Reading

Project team:
Mette R. Thomsen(PI)
Anders E. Daugaard
Arianna Rech
Ayoub Lharchi
Carl Eppinger
Gabriella Rossi
Hasti Valipour
John Harding
Konrad Sonne
Martin Tamke
Paul Nicholas
Ruxandra Chiujdea

Presenter:
Gabriella Rossi
PhD Fellow, CITA
garo@kglakademi.dk

Symposium on Human-Machine Teams for Design and Sustainable Construction
September 13-15, 2023 | TUM