Beyond Trial and Error: DOE Strategies for Stick Formulation Excellence

Moreira de Souza, Aline[1]; Nayyir-Mazhir, Afshan[1]; Whittaker, Katy[1].

[1] R&D, Croda Europe Ltd, Widnes, United Kingdom

Abstract

Waxes and oils are the most important ingredients for any stick formulations, such as lipstick, sunscreen, or deodorant stick. Performance attributes of sticks, such as hardness is dependent upon the combination, compatibility, types and ratio of waxes and oils. Considering the significant range of waxes available from different melting points, crystalline to non-crystalline and the vast options of emollients/oils on the market, to optimize those performance attributes several formulations via traditional empirical approach would be necessary, however, this is inefficient and time consuming, so an alternative approach such as DOE (Design of Experiment) is deemed necessary.

A total of 8 waxes and 12 emollients were used for this experiment, and a total of 50 combinations were made. A fixed amount of wax and oil was used, 20% and 80% respectively. The sticks were characterized for their hardness using a Texture Analyser and for the experimental design (DOE) and statistical evaluation a commercial software (JMP) was used.

The prediction model fitted well for hardness. Verified prediction model was obtained with raw hardness value, and a prediction profiler was built allowing to present the best combination based on the desired hardness. Verification was made using 5 formulations that were randomly chosen and had not been tested before. The formulations were characterized aiming to validate the model, which resulted in the actual value very close to that predicted by the DOE.

The use of DOE can be a great formulation tool to be broadly explored in Personal care, as it decreases the size of experiments, avoiding the trial-and-error approach. The use of predictive models helps formulators to understand the ingredient combinations, in this experiment the wax and oil interactions. Allowing a more informed decision regarding the ingredients to be used in a certain formulation.

Keywords: Design of experiment, stick formulations, predictive modelling.

1. Introduction

Formulations in stick form are not novel to the market, however there is a growing interest in their application across various domains, including for several applications sunscreens, deodorants, and colour cosmetics like lipsticks and foundations. The primary appeal of stick products lies in their user-friendly nature, allowing application without direct contact with the formulation [1].

The most prevalent type of stick is a mixture of waxes and emollients that are cast into a solid form. The efficacy of these formulations hinges on the interaction between oils and waxes, which is influenced by the properties of both components. An ideal mixture is characterized by ease of spreading and the ability to form a thin, uniform film with substantial coverage [2].

Waxes, which are esters of long-chain fatty acids reacted with fatty alcohols, are essential in stick formulations. They contribute to the desired viscoelastic properties, structural coherence, and thermal stability [3]. The intrinsic properties of each wax vary significantly, with options ranging from synthetic to plant-based waxes [4], soft to hard, crystalline to non-crystalline.

Given the wide array of raw materials available, traditional formulation approaches often rely on extensive trial-and-error to explore the interactions between different waxes and emollients. This process can be time-consuming and inefficient.

The present study aims to reduce the reliance on the trial-and-error approach in stick formulation development by utilizing Design of Experiments (DoE). Specifically, this research focuses on employing DoE to investigate the interactions between waxes and oils in terms of formulation hardness. By developing a validated predictive model, this study seeks to forecast the hardness of stick formulations based on various combinations of waxes and emollients. This methodology is expected to streamline the formulation process and enhance the precision and efficiency of developing new stick-based products.

A Design of Experiment (DoE) is a structured methodology used to investigate and model the relationship between multiple factors and their effects on one or more responses. JMP software provides a comprehensive suite of tools for creating efficient experimental designs.

Through the use of DoE, we aim to determine the formulations comprised of a blend of wax and oil by considering the chemistry of the oils and waxes as a factor. Full factorial design is used to run every possible combination of experiments while fractional factorial design selects a systematic combination of ingredients or experiments by using the results from the data collected to predict the answer where the full factorial knows the answer. In this project the use of fractional factorial design will be explored.

2. Materials and Methods

2.1 Formulations preparation

A total of 8 waxes range and 12 emollients identified by **Table I** and **Table II**, were used for this experiment, and a total of 50 combinations were made. A fixed amount of wax and oil was used, 20% and 80% respectively. The sticks were characterized for their hardness using a Texture Analyser and for the experimental design (DOE) and statistical evaluation a commercial software (JMP) was used.

Table I: List of 8 waxes used for the investigation.

	INCI Name of wax
W1	Cera Alba
W2	Euphorbia Cerifera (Candelilla Wax)
W3	Microcrystalline Wax
W4	Synthetic Beeswax
W5	C18-36 Acid Glycol Ester
W6	C18-36 Acid Triglyceride
W7	Tribehenin
W8	Synthetic Beeswax

Table II: List of 12 emollients used for the investigation.

	INCI Name of Emollient
E1	PPG-15 Stearyl Ether
E2	PPG-3 Isostearyl Methyl Ether
E3	C12-15 Alkyl Benzoate
E4	Diisopropyl Dimer Dilinoleate
E5	Caprylic/Capric Triglyceride
E6	Isostearyl Isostearate
E7	Pentaerythrityl Tetraisostearate
E8	PPG-3 Benzyl Ether Myristate
E9	Di-PPG-3 Myristyl Ether Adipate
E10	Di-PPG-2 Myreth-10 Adipate
E11	Bis-Diglyceryl Polyacyladipate-2
E12	Pentaerythrityl Isostearate/Caprate/Caprylate/Adipate

Wax (20g) and oil (80g) were added to a 250mL beaker and heated using a hot plate (IKA - C-MAG HP4 - Germany) at a temperature dependent on the melting point of the wax. The mixture was combined until homogenous using an overhead stirrer (IKA – Eurostar 20 digital – Germany) at a speed of 60-80 rpm. This process was repeated for each of the blends presented **Table III**.

To prepare the lipsticks bullets, the prepared blend of wax and oil was poured into the mould from one side carefully, then left to solidify for 5 minutes at room temperature. The excess formulation was then removed, and the filled moulds stored in the freezer for 10 minutes. Filled moulds were then left to stand at room temperature for 2 minutes before removing the sticks from the mould. Seven lipsticks were made for each blend the obtained lipsticks to test hardness.

A B C D

Figure 1: (A) blending (B) moulding (C) lipstick bullets (D).

2.2 Characterisation of sticks

Five lipsticks were used out of seven lipsticks for each blend to test hardness using the texture analyser (TA.XT. Plus Texture Analyser from Stable Micro Systems). Prepared lipstick samples were clamped within the orifice of the supporting fixture and then the lipstick was wound out to its maximum length and the supporting fixture attached to the machine bed in such a position that the hemispherical blade came down on the sample at 3mm away from the tip as shown in the **Figure 2**. The graph of typical force profile was generated, which provided the maximum force value or maximum peak which relates to the force required to break the lipstick from the main body of the sample which provided an indication of the hardness of the sample. Individual hardness values for each five lipsticks were generated and average values and standard deviations (SD) calculated. Same processes of texture analysis were repeated for all blends.

Figure 2: Texture analysis of lipstick

2.3 Use of DoE and prediction model

In this study, a screening DoE was employed to investigate the factors influencing lipstick hardness, with the response variable being the force required to deform the lipstick bullet. The objective was to optimise the performance, aiming to achieve a target hardness range 200 – 300g. The factors considered in this study were the waxes and oils, and as they have different chemistry and characteristics, they were considered categorical variables.

Full factorial design indicated a total experiment of 96 blends, where every possible combination of all levels of all factors are studied. While comprehensive, this approach can be labour-intensive and time-consuming. Therefore, screening designs are typically employed to identify and focus on the most significant factors and interactions, thereby reducing the number of experiments required.

The screening design showed a minimum number of runs (19 experiments), however that would only bring 42% χ^2 efficiency. Therefore, the runs were increased to 50 blends, resulting in a χ^2 efficiency of 87%.

All the raw hardness values obtained from the texture analyser were inputted in the JMP software. Statistical evaluation was performed to exclude any possible outlier data.

After generating the prediction model five other blends were chosen randomly and were formulated and tested using the texture analyser. The hardness results were compared with the predicted hardness values obtained from JMP software estimation to verify the prediction model.

3. Results and discussion

DoE usually works well for continuous variables or quantitative variables as opposed to categorical variables or qualitative variables such as chemistry of waxes and oils which are more descriptive in nature (descriptive observation), so here we were using the chemistry of wax and oil such as ester or ether as a categorical variable. χ^2 efficiency test is a hypothesis test that is used to determine whether there is a relationship between two categorical variables. Using the χ^2 efficiency test, it was predicted that screening the minimum number of blends would not give a good correlation and the reliability of the resultant predictive model would have a high chance of failure. Increasing the number of blends to 50 significantly increased the χ^2 result and therefore improved the likelihood of achieving a robust predictive model whilst still reducing the number of experimental tests by almost 50%. **Table III** lists the 50 blends created for this work, which was randomly created by JMP.

For this project, a multiple linear regression was used in JMP as there was more than 1 factor considering wax and oil variables to fit in a single model. So, from this, we got regression modelling, predictive modelling, and optimization. Regression modelling talks about which variables have the most influence over the response whereas predictive modelling predicts the response based on the important values of the input variables. Optimisation identifies the optimal values and the input variables for a required response. So, every modelling has their own significance. While developing these models, hardness was chosen as the Y variable and waxes and oils were chosen as construct model effects which means how hardness is dependent on different waxes and oils [5].

Table III: Experimental matrix for the 50 combinations.

Experiment number	Wax	Emollient	Experiment number	Wax	Emollient
1	W6	E2	26	W1	E6
2	W2	E3	27	W4	E11
3	W6	E10	28	W4	E6
4	W6	E4	29	W7	E5
5	W1	E10	30	W5	E5
6	W3	E2	31	W1	E8
7	W7	E6	32	W5	E9
8	W7	E11	33	W1	E4
9	W2	E8	34	W3	E8
10	W6	E7	35	W5	E6
11	W5	E1	36	W2	E7
12	W5	E12	37	W3	E12
13	W2	E2	38	W8	E11
14	W4	E12	39	W8	E7
15	W7	E10	40	W8	E8
16	W4	E4	41	W5	E7
17	W3	E3	42	W6	E9
18	W1	E5	43	W6	E11
19	W4	E2	44	W3	E7
20	W8	E10	45	W1	E9
21	W6	E1	46	W4	E9
22	W2	E1	47	W7	E1
23	W3	E4	48	W8	E12
24	W8	E5	49	W7	E3
25	W5	E3	50	W2	E9

The lipstick bullet should give structural integrity that should not crumble, melt, or deform after moulding and reaching room temperature [6]. The hardness of a lipstick usually reaches a peak after 24 hours of being moulded at optimum temperature, so all the lipsticks were tested for hardness after 24 hours of moulding.

The texture analyser measures the resisting force of the lipstick. The hardness results from this experiment could be showed by the hardness mean and graphs, however the amount of raw data would be better analysed via chemometrics, where mathematical models (statistical models) can easily show patterns where the human brain can't see it easily. Mathematical modelling can help research to be more data-driven, rather than trial-and-error approach.

The JMP prediction model was developed by inputting the individual values obtained for hardness where all the points were distributed randomly in a residual plot and predicted plot where the p-value is <0.0001 and RSq=0.76 which can be seen in the **Figure 3.**

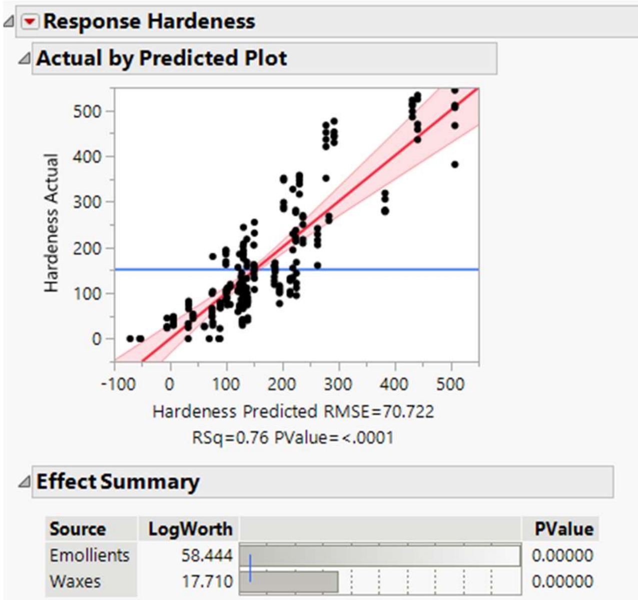

Figure 3: *Actual by Predicted plot and effect summary with individual raw data for Hardness.*

In the prediction model, the r-square value is 0.76 which shows this model fits well as the higher the value of r-square the more aligned the actual and predictive values are and therefore the more reliable the predictions. Similarly, the p-value is <0.0001 which is also good indication of fitting the model in this study as it rejected the null hypothesis [7]. There were also RMSE value which was also reduced by removing outilers, so the RMSE value also helps to find the reliability of the module [8].

Table IV: Estimation showing impact of waxes and oils on hardness value.

Scaled Estimates

Nominal factors expanded to all levels

Term	Scaled Estimate	Std Error	t Ratio	Prob>|t|
Intercept	151.92042	4.623531	32.86	<.0001*
Waxes[Wax 1]	51.534403	13.00482	3.96	<.0001*
Waxes[Wax 2]	44.797861	13.22336	3.39	0.0008*
Waxes[Wax 3]	43.678722	14.94665	2.92	0.0038*
Waxes[Wax 4]	-44.36288	13.08384	-3.39	0.0008*
Waxes[Wax 5]	38.537381	12.05848	3.20	0.0016*
Waxes[Wax 6]	22.055068	11.98247	1.84	0.0670
Waxes[Wax 7]	-102.4861	13.17953	-7.78	<.0001*
Waxes[Wax 8]	-53.75447	13.18135	-4.08	<.0001*
Emollients[Emollient 1]	40.119025	15.86959	2.53	0.0122*
Emollients[Emollient 2]	-46.34511	16.16986	-2.87	0.0046*
Emollients[Emollient 3]	-120.6542	16.4242	-7.35	<.0001*
Emollients[Emollient 4]	-74.45603	16.0217	-4.65	<.0001*
Emollients[Emollient 5]	-103.6302	15.93179	-6.50	<.0001*
Emollients[Emollient 6]	-101.6636	15.89091	-6.40	<.0001*
Emollients[Emollient 7]	28.515068	14.30276	1.99	0.0474*
Emollients[Emollient 8]	-65.89527	17.73073	-3.72	0.0003*
Emollients[Emollient 9]	-66.66039	14.16163	-4.71	<.0001*
Emollients[Emollient 10]	89.084853	15.9042	5.60	<.0001*
Emollients[Emollient 11]	333.92617	15.87288	21.04	<.0001*
Emollients[Emollient 12]	87.659721	17.10693	5.12	<.0001*

Table IV shows the estimation of the waxes and oils impact on the hardness results of the lipsticks in the mixture of wax and oil, whether the hardness value decreases or increases in the presence of them with the representation of the bar chart including p-value. It shows multiple linear regression estimates for the parameters (waxes and oils) individually by giving the p-values. From these p-values, we can say which kind of wax and oil have a significant impact on the hardness. We can see p-values in different colours in the **Table IV**. The yellow color indicates greater significance while red colour lesser significance and black indicates no signifiance in the variables. Here, W7 and W8, play a significant role in reducing the hardness of the lipstick which can be seen from the scaled estimates figure where bar diagrams lie towards the left side. While W1, W2, W3 and W4 were responsible for increasing the hardness of the lipstick as bar diagrams were pointed towards the right side. E2, E3, E4, E5, E6, E8 and E9 are shown to decrease the hardness with high significance whereas emollients E10, E11 and E12 have the ability to increase the hardness.

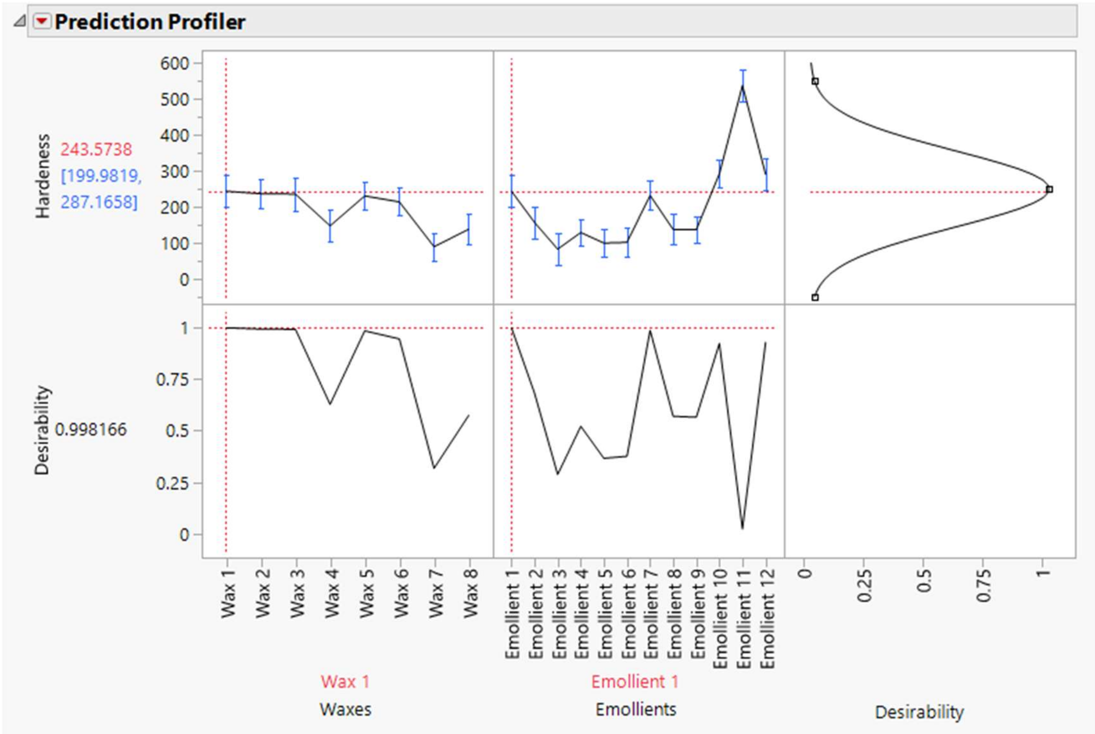

Figure 4: *Prediction profiler with raw hardness data.*

The prediction profiler in JMP is a great way to interact with complicated statistical models. With the profiler, we can interactively generate predictions from our model as we change settings of individual factors and visualize interactions like how the effect of one factor for our response can change at different levels of another factor. Here, it finds optimal settings for our factors which is waxes and oils, based on goals we have set for our response i.e., hardness.

The prediction model was verified by choosing five blends randomly showed by JMP prediction model which had not been tested experimentally (see **Table V**). Those five blends were made, and texture analysis completed. The actual hardness results obtained were compared with the predicted hardness results obtained from the JMP prediction model. The actual hardness values and predicted hardness values were similar for all five blends done for verification of model which validates the prediction model with more accuracy.

Table V: Predicted and actual hardness values for model validation.

Blend	Hardness (g)	
	Prediction (DoE)	**Actual**
W1 and E3	82 (+/-45)	52 (+/-7)
W8 and E9	31 (+/- 41)	51 (+/-9)
W3 and E11	529 (+/- 47)	620 (+/-90)
W7 and E5	0	0
W4 and E3	0	0

Actual hardness value '0' indicates that the lipstick bullets did not form, this also matched the '0' values from the prediction model for the same blends. For instance, the prediction model predicted the hardness value of lipstick formed by the blend of W7 and E5 was -50.6g, from which we can assume that there were no lipsticks formed because -ve hardness value is not possible for the lipstick bullets which are placed in the texture analyser (there must be some +ve value for hardness). All the blends which did not form lipsticks were taken as '0' hardness value. Whilst running the experiments, the blend could not form lipstick in real experiments and value was taken as '0'. Likewise, all other four blends showed similar experimental value as shown by the prediction profiler in similar way. The correlation of data from the texture analyser and prediction model completely correlated with each other, and no significant difference between actual and predicted results at $p<0.001$. Therefore, this study shows the great finding in terms of investigation of wax and oil interaction as it can predict the value of hardness of blends for lipstick without experimenting them.

4. Conclusion

The prediction model fitted well for hardness. Verified prediction model was obtained with raw hardness value after removing possible outliers from the data. Making five blends also showed a good correlation between the texture analyser results and prediction model hardness data. JMP predicted the positive and negative effect of waxes and oils on hardness. This prediction model can also predict the value of hardness of that combination of wax and oil which have never been tested before. It is also capable of predicting the combination of wax and oil which can give the ideal value of hardness specified by this study i.e., 200-300g. So, from this prediction model, the investigation of wax and oil interactions became easier. So, the JMP prediction model indicated the possibility of using the data for formulating guidelines to help formulators to use and understand wax and oil interactions.

This study could also be useful for any stick formulations such as sun care stick and deodorants, not just for lipstick formulations. The investigation was for only one wax and one oil combination, but in future, the study can be conducted for more than one wax and oil combination and at different ratios. As initial step of lipstick formulation, we can also move one step forward using the same strategy used in this study in order to test the impact on the hardness with the addition of pigments or UV filters.

The use of DOE can be a great formulation tool to be broadly explored in Personal care, as it decreases the size of experiments, avoiding the trial-and-error approach. The use of predictive models helps formulators to understand the ingredient combinations, in this experiment the wax and oil interactions. Allowing a more informed decision regarding the ingredients to be used in a certain formulation. It can also provide information regarding which combination of waxes and oil to use to reduce or increase hardness.

Additional DOE’s can be complemented to this work to narrow down even further the optimum ratio of waxes and oils to achieve the desired end features. The use of statistical tools can show the positive and negative impact of each ingredient in the characteristics the formulators are measuring, for this project it was possible to have information on the impact of certain emollients in combination with any waxes, via predictive modelling, without having measured all the possible combinations. It’s also a valuable tool to use as it can significantly save the formulator a lot of time.

Acknowledgements

Throughout this project, numerous individuals have contributed their expertise in various domains. We acknowledge Katy Whittaker for her proficiency in data analysis and JPM software, and Afshan Nayyir-Mazhir for her support in the project's execution and her keen interest in understanding the individual effects of ingredients in stick formulation.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

1. Schueller R, Romanowski P. Beginning Cosmetic Chemistry. 2009. 254.
2. Beri A, Pichot R, Norton IT. Physical and material properties of an emulsion-based lipstick produced via a continuous process. International Journal of Cosmetic Science. 2014;36(2):148-58.

3. Branna, Tom. The Physical Chemistry of Cosmetic Formulations. Happi. 2008. Available from https://www.happi.com/the-physical-chemistry-of-cosmetic-formulatio/
4. H. de Clermont-Gallerande et al. Substitution of synthetic waxes by plant-based waxes in lipsticks: OCL 2022, 29, 19.
5. Statistics Knowledge Portal [Internet]. Jmp.com. 2013 [cited 2022 Aug 03]. Available from: https://www.jmp.com/en_nl/statistics-knowledge-portal.html
6. Sango D, Binder D. Lip Care Product Formulation Strategies. Handbook of Formulating Dermal Applications. 2016 Dec 12:513.
7. Dahiru T. P-value, a true test of statistical significance? A cautionary note. Annals of Ibadan postgraduate medicine. 2008;6(1):21-6.
8. Wang WJ, Lu YM, Iop, editors. Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. 5th International Conference on Mechanical Engineering, Materials Science and Civil Engineering (ICMEMSCE); 2017 Dec 15-16; Kuala Lumpur, MALAYSIA. BRISTOL: Iop Publishing Ltd; 2018.