# Prediction of higher order drug combinations to treat cancer

**Auteur :** Grosfils, Amandine
**Promoteur(s) :** Geurts, Pierre
**Faculté :** Faculté des Sciences appliquées
**Diplôme :** Master en ingénieur civil biomédical, à finalité spécialisée
**Année académique :** 2020-2021
**URI/URL :** http://hdl.handle.net/2268.2/11469

# Prediction of higher-order drug combinations to treat cancer


Master thesis conducted by
**Amandine Grosfils**
with the aim of obtaining the degree of Master in Biomedical Engineering

Under the supervision of
Prof. Juho Rousu (Aalto university)
Prof. Pierre Geurts (ULiège)

University of Liège - Applied sciences
Academic year 2020-2021

# Abstract

Finding a treatment for cancer is a major challenge of our time. In the ongoing research, combination therapies (the use of several drugs together) are of high interest. In comparison with the use of a single drug, combinations of synergistic drugs (*i.e.* drugs that are more effective together than alone) can be as effective while allowing to overcome the drug resistance, to reduce the doses at which the drugs are used and consequently decrease their toxic effect and multiply their targets. However, the space of all potentially effective combinations is too large to experimentally test all of them and assess their effectiveness. This is known as the combinatorial explosion problem. To overcome that, the identification of interesting combinations requires the help of computational tools. In recent years, machine learning models have been successfully used in biomedical applications. They are typically used in order to determine which combinations would be interesting to be experimentally tested. Since some models aiming at predicting the responses of *pairwise* combinations (*i.e.* two drugs in the combinations) already exist, there are only a few machine learning models able to predict responses of *higher order* drug combinations (the order of a drug combination is defined as the number of drugs in the combination). In addition to the response of a drug combination (typically expressed as a growth percentage), the synergy score of this combination is of high interest. The synergy score allows to answer the question: how much are those drugs more effective together than individually?

This work is a step towards the use of machine learning to predict the effect of *higher order* (order larger than two) cancer drug combinations. It has been made in collaboration with Aalto University (Finland), where a machine learning tool called ComboFM has been developed. ComboFM is able to efficiently predict pairwise responses of cancer drugs. The goal of this work is to extend the use of ComboFM to the predictions of *higher order* drug combinations. To that end, we propose to combine ComboFM with another model, called the Dose model. The Dose model computes the responses of *any* order drug combinations, based on all the pairwise responses existing in the combination.

This thesis investigates how those two models can be combined together in order to predict responses of *higher order* drug combinations while decreasing the amount of required experimental data (pairwise responses). This combination of models gives rise to several issues that are tackled and investigated.
Firstly, it appears that the Dose model should be slightly adjusted in order to be suitable to the data used in ComboFM. The Dose model uses a classical Hill function to fit the data and this function can be modified. The way of modifying this Hill function in order to better fit the data is thus analyzed.
In the second part of the experiments, the way of using ComboFM is investigated. Indeed, there are a lot of different possibilities of using ComboFM before using the Dose model (which pairwise responses do we want to predict using ComboFM?). Different of those possibilities are thus tested and analyzed.
The third part of our experiments focuses on the assessment of the performance through a sensitivity analysis aiming at understanding how the two models interact with each other.
The Dose model is not the only way to predict responses of higher order drug combinations,. Indeed, there exist different analytical formulas to do so and thus several of those analytical formulas will be computed and compared.
In addition to the predictions of the response of a combination, scientists are interested in synergy scores of the combination. There exist different synergy metrics. Several of them will be computed from the predictions made with the proposed model, and will then be compared.
Note that in addition, the impact of using ComboFM and the Dose model instead of the Dose model only is assessed. The last part of experiments tries to increase the order of the combinations in order to push the analysis one step further.

The experiments made in this work show that ComboFM and the Dose model can efficiently be combined, as long as the parameters of both models are optimized specifically for this application.

# Acknowledgements

Firstly, I would like to thank Prof. Pierre Geurts, my supervisor at the university of Liège who helped me to find a master thesis that combines my interest in machine learning applied to the biomedical field and my wish to collaborate with an international lab. He put me in contact with Prof. Juho Rousu from Aalto university (Finland). I also would like to thank him for his advises and his supervision.
Thanks to Prof. Rousu, I had the chance to work on a subject that takes up my interests in both machine learning and a biomedical application. Pr. Rousu hosted me (virtually) in his team, where I felt welcome, and helped to understand the topic of cancer combination therapy and also guide me during this work. While it was not possible to be there during my master thesis, he offered me the opportunity to do an internship this summer to visit Helsinki and stay in his lab. I thank him also for that reason.
I also would like to thank Pr. Rousu's team that I learnt to know during their lab meeting every week. In particular, I thank Tianduanyi Wang for answering my questions and helping me get used to the Finnish cluster that probably saved me a lot of computational time.

Finally I thank my family and friends for the moral support during this period which was even more appreciated because of the 100% online nature of this work, due to the pandemic.

# Contents

# Chapter 1 Introduction

## 1.1 Motivation

Cancer refers to a group of diseases characterized by an uncontrolled multiplication of abnormal cells[2]. This illness is one of the principal cause of death in all countries of the world. According to the global cancer statistics of 2020[3], 19.3 millions of new cases of cancer were diagnosed and 10 million deaths were counted around the world, during the past year[3]. The most common type of cancer is the female breast cancer, representing 11.7% of the diagnosis and the most lethal type of cancer is the lung cancer, representing 18% of the deaths. The probability of having a cancer during his life is now around 46.9 % for a man and around 36.6% for a woman[2][10]. More importantly, it is expected that the incidence rate of cancer will increase in the coming years. Indeed, in 2040 we expect 28.4 millions of new cancers which represents an increase of 48% in comparison with 2020. These numbers show the importance of treating cancer.[3]

However, it is of common knowledge that treating cancer is one of the biggest medical challenges of our time. Indeed each cancer is different because it depends on the location, the considered treatment and the causes of the cancer, among other things[12]. The causes of cancer are especially difficult to identify. It has been established that less than 10% of cancers are hereditary, while the main causes are environmental factors such as smoking, radiations, chemical substances, virus.[4] As a consequence, it appears now[12] that combination therapies are often required[20]. The interest for combination therapies goes back to the end of the 1950s[17]. Such therapies are very promising for several reasons. They firstly allow to overcome the recurring problem of mono-therapy resistance. They also allow to reduce the toxic effect of the drugs because when we use a combination of drugs, we can usually decrease the dose of each drug in comparison with mono-therapy. Furthermore, those treatments are more efficient because they have a multi-targeting effect. Indeed, each drug in the combination can have its own target and by combining them, we can multiply the targets[20].
However, experiments must be done in order to know the effects of the combinations of drugs.
Ideally, one would be able to identify the most efficient combination of drugs for cancer type as a pre-clinical stage[18]. Dealing with different cancer types means dealing with different cell lines. This represents a major challenge for at least two reasons. Firstly, we do not know the order of the best combination (meaning how many drugs we should use), neither which drugs to combine. Secondly, the response to a drug combination also depends on the cell context (meaning that each combination should be tested in different cell lines)[18].
All together, each drug combination should be tested at different doses (for each drug in the combination), on different cell lines. It is easily understandable that experimental screening is unfeasible in this case. As an example, if we consider one particular cell line and 10 different drugs, each of them at 10 different doses, one should perform $10^{10}$ experiments. It is not feasible to perform so many experiments in a reasonable time. This is also not efficient, it would take $10^{10}$ experiments to maybe find a very small number of interesting combinations. This is known as the *combinatorial explosion problem*[5].

This is where computational tools and machine learning models can be very interesting. The purpose of using machine learning in this context is to identify some drug combinations that could be efficient in a particular context, before doing any experimental measurements[27]. That would allow to perform a pre-research and make experimental measurements only on potentially interesting combinations. That would highly decrease the number of experiments and increase the research efficiency (and speed).
However, the use of machine learning in this field is quite recent[27]. This is largely due to the lack

of sufficient and high quality training data, which constitutes a major challenge to overcome.

In the best scenario, the scientific community would end up building a model/tool allowing physicians to choose the best treatment for each patient [27], leading to a personalized and thus more efficient treatment. We are still far away from that tool but researchers are moving forward.
Note that finding the most effective drug combinations is not sufficient in order to find the best treatment against cancer. Indeed, this work focuses on the effectiveness of the drugs against targeted cancerous cell but it does not consider their toxic effect against healthy cells.

## 1.2 Work structure

This master thesis has been made in collaboration with the Aalto university (Finland), aiming at extending a developed tool called ComboFM [27]. ComboFM is able to efficiently predict the responses of pairs of drugs. The goal of this thesis is to use ComboFM as a baseline to predict the effect of higher order combinations (meaning with more than two drugs).

This work is organized as follows:

The first part of this work (Chapter 2) will set the context by explaining the different needed notions and explaining the different models and formulas used.
The second part (Chapter 3) explains the different possibilities in order to achieve the goal (predicting the effect of higher order drug combinations using ComboFM) and which one has been chosen as well as the reasons behind this choice.
The third part (Chapter 4) will detail the data on which the work is realized, as well as the challenges that represents such data.
The implementation of the chosen working hypothesis will be detailed in the next part (Chapter 5). This choice rises some challenges. Those challenges will be detailed, as well as the chosen solutions to overcome them.
The last part (Chapter 6) of this work will present the different experiments, as well as their results and analysis.
Finally, possible further investigations and experiments will be presented, with the conclusion of this work (Chapter 7).

Note that the main scripts and data files needed to run the main experiments of this work are available on GitHub : `https://github.com/AmandineGrosfils/Master-Thesis`.

# Chapter 2 Background

This chapter aims at providing the baselines of this work, as well as the general principles and tools that are used in the subsequent chapters.
Several models and methods that can/could give a response (or a part of a response) to our research question will be explained. In particular, this thesis relies mainly on two models: ComboFM and the Dose model. Both models will thus be explained in details (see Sections 2.3 and 2.4).

## 2.1 General principles

This section is an overview of the terms/concepts used in this work.

### 2.1.1 Machine learning

**Supervised machine learning model** [21] is a process in which a model is built in order to find a relationship between the set of input $\mathbf{x}$ and the set of output $\mathbf{y}$. Once the relationship between both has been defined, the model can be used in order to predict the output corresponding to unseen input.
The parameters of the model ($a$, $b$ for example in the case of a linear regression $y = ax + b$) are learned in order to maximize the performance of the model (meaning the accuracy of its predictions).

If the output of a model is continuous, the problem is called a regression problem. This is opposed to classification problems, where the number of possible output values is limited (equal to 2 in the case of binary classification).
In this work, supervised machine learning methods will simply be referred to as machine leaning methods, because unsupervised learning will not be used.
Regarding the performance of the machine learning model, one has to distinguish the performance on the learning set (that is maximized in order to build the model) and the performance on unseen data which is measured via the generalization error. The most important is the performance in generalization, in order to obtain an efficient model for various data. To evaluate those performance, different metrics are used and are detailed below (see Table 1 for a summary).

### 2.1.2 Evaluation metrics

**Root-mean-square error (RMSE)** [22] is a commonly used method of evaluation. The formula is the following: $RMSE = \sqrt{\sum_i^n (y_i - \widehat{y_i})^2}$, where $y$ is the true output, $\widehat{y}$ is the predicted one and $n$ is the number of instances in the datasets. The value of the RMSE is always greater or equal to 0. The lower the RMSE, the better the predictions. The best case is thus a value of 0, meaning the values of $\hat{y}$ perfectly match the values of $y$. Note that the range of the RMSE is proportional to the range of the values which makes the comparison between two RMSE made on different data quite complicated.

**Mean absolute error (MAE)** [22] is another measure of error between two sets of data. The computation is made using the following formula: $MAE = \frac{\sum_i^n |y_i - \widehat{y_i}|}{n}$ with $n$ being the number of elements in the datasets, $y$ the true output and $\widehat{y}$ the predicted one. As the RMSE, the MAE is a scale-dependent measure. The two measures are insensitive to the direction of the errors (the sign has no importance since there is a power two in RMSE and a absolute value in MAE). For both measures, the lower the value the better. The main difference between the two measures lies in the fact that the RMSE gives more weight to large errors, while MAE gives the same weight to all the errors.

**$R^2$ [9]** is another value used to evaluate the predictions of a model. The formula is the following: $R^2 = 1 - \frac{\sum_i^n (y_i - \widehat{y_i})^2}{\sum_i^n (y_i - mean(y_i))^2}$, where the sum is over instances in the test set. This coefficient is always bounded between -1 and 1, which makes it easier to use across different scales of data. The ideal value is 1, meaning that the predictions perfectly fit the data. The lower the value of $R^2$, the lower the quality of the predictions.

**Pearson correlation coefficient [22]** is a commonly used coefficient to compare two vectors of data, $x$ and $y$. The coefficient is computed using the formula: $Pearson = \frac{cov(x,y)}{\sigma_x \sigma_y}$, where $cov(x, y)$ is the covariance between $x$ and $y$ and $\sigma$ are the standard deviations. The value is always bounded between -1 and 1. A value of 0 means that there is no linear correlation between the two vectors, $x$ and $y$. A value of 1 or -1 means that there is an exact (respectively increasing or decreasing) linear relationship between the vectors. Note that having a correlation coefficient equal to 1 does not mean that the two vectors of data contain the same values.

**Spearman correlation coefficient [22]** can be seen as a modified version of the Pearson coefficient. While the Pearson coefficient assumes that the two datasets are normally distributed, the Spearman coefficient does not make any assumption on the distributions of the two vectors of data. The Spearman coefficient is computed as follows: $r_s = \frac{cov(r_{gX}, r_{gY})}{\sigma_{r_{gX}} \sigma_{r_{gY}}}$ where $cov(r_{gX}, r_{gY})$ is the covariance of the rank variables and $\sigma_{r_{gX}}, \sigma_{r_{gY}}$ are the standard deviations of the rank variables $r_{gX}$ and $r_{gY}$. In order to obtain the rank variables, the values of the dataset are sorted in descending order. The highest value has a rank of 1, the second one a rank of 2 and so on.
Like the Pearson coefficient, the Spearman coefficient is bounded between -1 and 1. A value of 0 means that there is no monotonic relationship between the two vectors of data, while a value of -1 or 1 means that there is a perfect monotonic relationship. The difference with the Pearson coefficient is that the Spearman coefficient looks for a monotonic relationship, which does not have to be linear. If there is a perfect monotonic relationship but not linear, the Spearman coefficient will be equal to 1 (in absolute value) while the Pearson coefficient will be lower than 1.

| Metric | Formula | Range of value |
|---|---|---|
| $RMSE$ | $RMSE = \sqrt{\sum_i^n (y_i - \widehat{y_i})^2}$ | Scale-dependent |
| $MAE$ | $MAE = \frac{\sum_i^n \lvert y_i - \widehat{y_i} \rvert}{n}$ | Scale-dependent |
| $R^2$ | $R^2 = 1 - \frac{\sum_i^n (y_i - \widehat{y_i})^2}{\sum_i^n (y_i - mean(y_i))^2}$ | -1 to 1 |
| Pearson | $Pearson = \frac{cov(x,y)}{\sigma_x \sigma_y}$ | 0 to 1 |
| Spearman | $r_s = \frac{cov(r_{gX}, r_{gY})}{\sigma_{r_{gX}} \sigma_{r_{gY}}}$ | 0 to 1 |

Table 1: Summary of the metrics used in this work, as well as their range of value.

As it has been mentioned, it is very important to assess the generalization error of a model in order to optimize its performance. There exist different techniques to do so. Two of them are explained below.

**Cross-validation** [22] is a widely used method to validate and optimize a model (see model selection part of FIGURE 2). The idea is to divide the dataset in different folds and then use one of the fold as a test set and all the others as the learning set. The method allows to test different parameters for the model and find the ones that give the best performance. In this work, a 5-folds cross-validation will be used. In addition, a test set can be kept apart in order to predict the outputs of data that is never part of the learning set. This method has been used in this work.

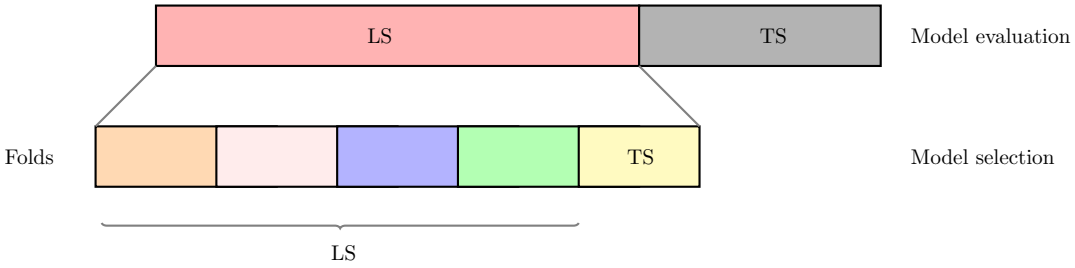



Figure 2: Schematics of a cross-validation and a separated test set. The dataset is divided into the learning set and the test set. The learning set in then divided in 5 folds. In each loop (the number of loops is equal to the number of folds) of the cross-validation, each folds is once the test set while all the others constitute the learning set.

**Nested cross-validation** is a more elaborated cross-validation. The general principle is the same but the division into different folds goes one step further. Each fold of the cross-validation is divided into $p$ other folds. The cross-validation is thus doubled. As an example, a $5 \times 10$ nested cross-validation is typically used.
The nested cross-validation is generally used when hyperparameters of the models need to be optimized [14]. Hyperparamaters differ from classical parameters in the fact that their value is used to control the learning process. In contrast, the value of the classical parameters is obtained via the learning step [13]. Using the classic cross-validation (without a separated test set) instead of the nested one can introduce a bias into the model towards the dataset. This is due to the fact that, in the non-nested cross-validation, the same data is used in order to both tune the model parameters and evaluate its performance. This creates a risk of overfitting. Using the nested cross-validation allows to avoid this bias [14].

In order to evaluate the performance of a model in generalization while avoiding overfitting, one has typically two options. The first one is to use a cross-validation added to a test set. The second one is to use a nested cross-validation, as explained above. In the ComboFM paper [27], a nested cross-validation was used. In this work, we will use a cross-validation and a separated test set.

### 2.1.3 Machine learning regression models

Different methods of machine learning are used in this work, the main one being higher-order factorization machines. In order to introduce how this method works, let us start by reminding some more basic machine learning methods.

**Linear regression** [7] [26] is a very simple model. From the input $\mathbf{x} \in \mathbb{R}^n$, the prediction $\hat{y}(\mathbf{x})$ is the following:

$$\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i \tag{1}$$

where $n$ is the number of features in the input data, corresponding to the number of weights $w_i$ to be learned. This very simple model allows to make the predictions in linear time, $\mathrm{O}(n)$, while the learning step has a time complexity of $\mathrm{O}(n^3)$. However it does not handle the interactions between the different features, meaning that each weight is associated to an individual feature. To overcome this drawback and capture features interactions, one can introduce weights that are combining the features:

$$\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_i x_j w_{ij} \tag{2}$$

where $\mathbf{w}$ is a matrix of weights $w_{ij}$. This formula corresponds to a polynomial regression. Adding the $w_{ij}$ weights allows to capture the interactions between two features at a time. The complexity (both in time and space) of the model is increased to $\mathrm{O}(n^2)$ for the evaluation of the model. The main drawback of using this polynomial regression is that it is not convenient when dealing with sparse data, meaning learning the $w_{ij}$ is difficult. One possibility to overcome this serious drawback is the use of factorization machines.

**Factorization machines (FM)** [6] [26] are supervised learning approaches, based on the polynomial regression presented above (Equation 2). The difference comes from the modeling of the weight parameters. In FM, those weights are modeled using factorized parameters, which allows to rewrite the equation:

$$\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \mathbf{v_i}, \mathbf{v_j} \rangle x_i x_j \tag{3}$$

where

$$\langle \mathbf{v_i}, \mathbf{v_j} \rangle \quad = \quad \sum_{f=1}^{k} v_{if} v_{jf} \tag{4}$$

which is the dot product of the two features vectors, of size $k$, where $k$ is an hyper-parameter of the model called the rank of factorization.
The use of a factorized form to estimate the weights decreases the number of estimated parameters, compared to polynomial regression from $O(n^2)$ to $O(nk)$.
There exist different implementations in order to learn the weights of factorization machine. Typically, gradient-based optimization methods are used.

**Higher order factorization machines (HOFM)** [16] [27] are an extension of factorization machines. It allows to extend the formula to an arbitrary order $m$ of factorization (while this order is equal to $m = 2$ in the case of simple factorization machines). The predictions follow the same idea:

$$y(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{i,j} x_i x_j + ... + \sum ... \sum w_{i_1,i_2,...,i_m} x_{i_1} x_{i_2} ... x_{i_m} \tag{5}$$

where $m$ is the order of factorization. In that case, one can capture the interactions between $m$ features at a time. The order of factorization must be carefully chosen in order to give satisfying results while limiting the number of weights to estimate.
The factorization form of the weights gives:

$$w_{i_1,i_2,...,i_m} = \langle \mathbf{v_{i_1}}, \mathbf{v_{i_2}}, ..., \mathbf{v_{i_m}} \rangle \tag{6}$$

where $\langle \mathbf{v_{i_1}}, \mathbf{v_{i_2}}, ..., \mathbf{v_{i_m}} \rangle = \sum_{s=1}^{m} v_{i_1 s} v_{i_2 s} ... v_{i_m s}$ is the generalized inner product, computed by taking the sum over features of the product of the factors.

### 2.1.4 General principles of drug interactions

In this work, we are interested in *cancer* drugs. Responses of cancer drugs are typically expressed as a percentage of growth of the targeted cells, depending on the dose of the drug(s). For each considered dose, there is one corresponding response. We thus have a so-called drug-response curve. The corresponding machine learning problem is a regression problem, if we want to predict the mono-therapy response of a drug for one particular dose.
However, the responses of drugs (both antibiotics and cancer drugs) can be expressed in a binary way: either it is efficient or not efficient. This corresponds to a classification problem of machine learning. Typically, the research in antibiotics focuses on this qualitative response: the microbe is or is not drug resistant. One can do the same for cancer drugs: the cancer tissue is or is not drug resistant. This type of response will not be considered in this work, simply due to the used dataset (see CHAPTER 4 for more information concerning this dataset).
Below is an overview of the different responses and interactions between the drugs that are used in this work.

**Mono-therapy response** [31] is the response of a single drug, expressed in growth percentage in this work. It is represented as a dose-response *curve*, meaning for each dose of the considered drug, we have the corresponding response.

**Pairwise response** [31] is the response of a combination of two drugs. By considering cancer drugs, it is represented as dose-response *surface*. For a given combination, each drug can be taken at different doses. This gives rise to dose-response matrices.

In general, the response of a combination of $n$ drugs is represented in a *n+1* dimensional space.

In order to fully understand how we can characterize a combination of drugs and the mechanisms of interactions between the different drugs, several definitions are needed. Let us consider a combination of three drugs, A, B, C as represented on FIGURE 3.

**Pairwise interactions** [31] are all the interactions between only 2 drugs. Inside a combination of 3 drugs for example, we find 3 pairwise interactions (corresponding to the interactions between drugs A-B, B-C and A-C on FIGURE 3).

**Emergent interaction**[31] is the interaction between all the drugs of the mixture. There is only one emergent interaction in a given mixture of drugs (corresponding to A-B-C in the case of three drugs).

**Net interactions**[31] are the set of all the interactions. It does include the pairwise interactions and the emergent interaction (corresponding to A-C, B-C, A-C and A-B-C) but also all the others if there are at least four drugs in the combination. If we consider the quadruplet of drugs ABCD, the interactions between ABC, BCD and ACD are neither pairwise interactions neither emergent interactions, they are thus called other interactions.

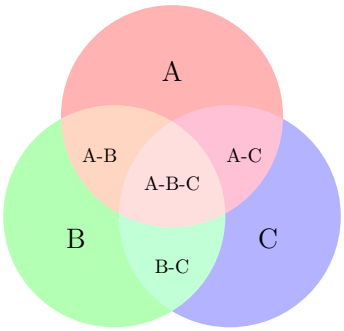



Figure 3: Representation of the different interactions between drugs in a mixture of 3 drugs A, B and C.

**Hill equations/curves** were initially used in biochemistry. They refer to the binding of ligands to macromolecules (such as proteins). It is expressed as a function of the ligand concentration. In general the binding of a new ligand is enhanced if there are already other ligands on the macromolecule (this phenomenon is known as cooperative binding). Typically Hill curves are sigmoid curves. Their shape is thus very convenient to also represent dose-response curves, see FIGURE 4 for an example of a Hill curve fitting mono-therapy data points. The type of data used is important to understand this curve. Indeed, in general one would expect the response to be a increasing function of the stimulus (dose). It is indeed the case here because the response is expressed as a growth percentage of the targeted cells. The lower the response, the greater the impact of the drug. As a consequence, the effect of the drug increases with the considered dose. See SECTION 4.1 for a detailed explanation of the data used in this work.

### 2.1.5 Synergy quantification

A synergy combination of drugs is defined as a combination having a greater effect than the sum (or product or other formulas, depending on the synergy model used) of the separate effects of the drugs. Opposed to synergistic combination, we have antagonistic combinations. Their effect is thus lower than the combination of the separate effects of the drugs. Having a synergistic or antagonistic combination reflects the presence of interactions between the drugs.
Synergy quantification is the actual goal of predicting the responses of drug combinations. One wants to discover synergistic combinations, and answer the question: are those drugs more effective together than alone?

In general, the synergy or antagonism quantification is computed by taking the difference between the expected effect by assuming no synergy and the measured or predicted effect of the combina-

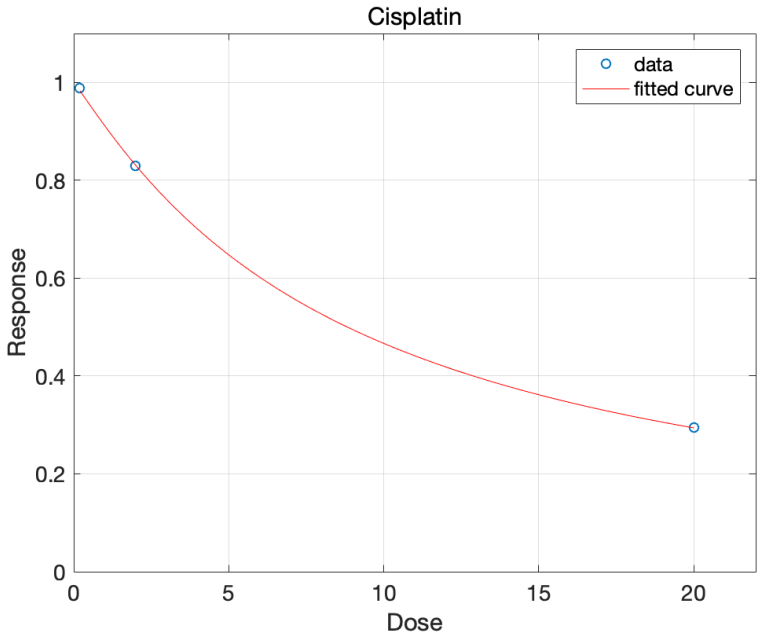



Figure 4: Example of a Hill curve fitting mono-therapy data.

tion. If the measured/predicted effect is greater than the expected one, we talk about synergistic combination. If it is lower, we talk about antagonistic combination. In both cases, the greater the difference, the stronger the interactions (that can be positive or negative) between the drugs.
However, there exist several definitions of a synergistic combination. The three main principles when dealing with synergy quantification are[11]:

- Dose equivalent principle (DEP): the principles states that if the effect of reducing one drug dose can be counterbalanced by adding a second drug at a constant ratio, the two drugs are said to be additive. There is no synergy between them.
- Multiplicative Survival principle (MEP): it considers a probabilistic approach. It states that, in case of no-synergy, the probability for the combination to be ineffective is equal to the product of the probabilities of each drug alone. This formulation was firstly introduced by Bliss in 1939[15] and gave rise to the Bliss independence formula that will be later presented.
- Highest single agent (HSA): it states that there is no synergy if the effect of a combination is equal to the effect of the most effective drug of the combination.

Different synergy scores derived from those general principles will be used in this work and are thus presented below. The notation is the same for all the scores: $g_i$ is the effect of a single drug $i$ and $g_{ij}$ is the effect of the pair of drugs $i$ and $j$.

**Bliss score** corresponds to the application of the MEP principle. The score is thus: $S_{ij} = g_i g_j - g_{ij}$, where $g_{ij}$ is the observed response and $g_i g_j$ is the expected one. If $S_{ij} > 0$, it means that $g_i g_j$ is greater than $g_{ij}$. According to how the data used in this work is defined, *i.e.* the lower the response value the greater the effect (see SECTION 4.1), a lower $g$ means a more effective drug (and thus a higher response). As a consequence, if $g_i g_j > g_{ij}$, then the observed response is greater in term of effectiveness (but lower in value) than the expected one and so we can talk about synergy. If $S_{ij} < 0$, the observed response is lower than the expected one and so the combination is antagonistic. It the score is equal to 0, the two drugs are said to be independent.

**HSA score** simply corresponds to the HSA principle. The formula is the following: $S_{ij} = min(g_i, g_j) - g_{ij}$. Note that the most effective drug is the one having the lowest value of $g$ because it corresponds to a higher growth inhibition (SECTION 4.1.2). That explains the *min* in the formula.
The interpretation is the same as the one of the Bliss score. If the score is positive, we consider the pair to be synergistic. If it is negative, we consider the pair as antagonistic.

Note that those two simple scores can very easily be extended to higher order drug combinations, by extending the product and the *min* term respectively.

**ComboScore[27]** can be seen as en extension of the Bliss score. However there is a significant difference. While we have one Bliss score for each dose combination, we have only one ComboScore for each drug combination.
The idea is the same as the one of the other synergy scores. One compares the difference between the expected response and the measured (or predicted) one. The difference is that ComboScore takes the sum of the differences for each dose combination. Let us consider two drugs, A and B, respectively taken at doses $D_A$ and $D_B$, the ComboScore is computed as follows:

$$\text{ComboScore(AB)} = \sum_{D_A, D_B} g_{AB}(D_A, D_B) - Z_{AB}(D_A, D_B) \tag{7}$$

where $g_{AB}(D_A, D_B)$ is the response of the pair of the 2 drugs, while $Z_{AB}(D_A, D_B)$ is the expected response:

$$Z_{AB}(D_A, D_B) = \begin{cases} \min[g_A(D_A), g_B(D_B)] \text{ if } g_A(D_A) \leq 0 \text{ or } g_B(D_B) \leq 0 \\ \frac{1}{100}(\widetilde{g_A}(D_A) * \widetilde{g_B}(D_B)) \text{ otherwise} \end{cases} \tag{8}$$

where $\widetilde{g_A}(D_A) = \min(g_A(D_A), 100)$ which truncates the response of a drug to 100, meaning that the expected growth with a drug can not be higher than the one without any drug (the control). The minimum between $g_A$ and $g_B$ if one of them is negative ensures that if one of the drugs kills the cells, the combination does too (see SECTION 4.1 for further information about the growth percentages in the dataset used in this work).

If the ComboScore is greater than 0, the combination is theoretically synergistic. However, the greater the score, the stronger the synergy of the combination. As a consequence, a threshold higher than 0 is often chosen to define synergystic combinations.

Note that the ComboScore can easily be extended to higher order drug combinations. Consider a $m$ order combination as example. Under the Bliss assumption the term $Z_{AB}$ will become:

$$Z_{AB...M}(D_A, D_B, ...D_M) = \begin{cases} \min(g_A, g_B, ..., g_M) \text{ if } g_A \leq 0 \text{ or } g_B \text{ ... or } g_M \leq 0 \\ \frac{1}{100}(\widetilde{g_A} * \widetilde{g_B} * ...\widetilde{g_M}) \text{ otherwise} \end{cases} \tag{9}$$

**Loewe additivity model[8]** is another commonly used model to assess the synergy of a pair of drugs. Let us consider two drugs: A and B. $d_A$ and $d_B$ are the doses at which the *combination* of A and B gives the given effect $e$. We further define $D_{Ae}$ and $D_{Be}$, the doses at which drugs A and B

produce the effect $e$ *alone*. The Loewe additivity model is defined as follows:

$$\frac{d_A}{D_{Ae}} + \frac{d_B}{D_{Be}} = 1 \tag{10}$$

The model is usually represented in a 2D space, as shown in FIGURE 5. When EQUATION 10 is satisfied for all the effects $e$, we have the so-called isobole curve. The synergy or antagonism of the combination is assessed by the deviation from the isobole. If the dose-response curve of the drug pair is *below* the isobole, the pair is synergistic. Indeed, in that case the required doses to obtain a given effect $e$ are lower than the ones if the drugs were additive. On the contrary, if the dose-response curve is *above* the isobole, the pair is considered to be antagonistic. In that case, the required doses are higher than the ones if the drugs were additive.

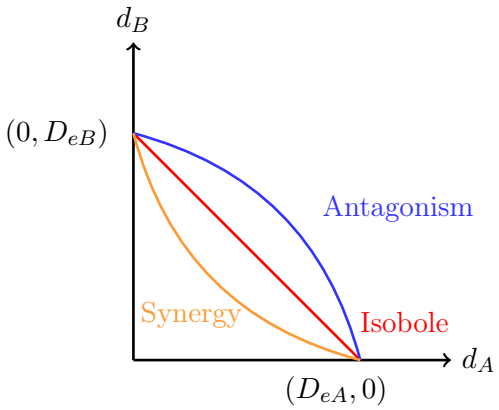



Figure 5: Representation of the Loewe additivity model for the pair of drugs A-B.

## 2.2 Related work

This section aims at explaining how responses of high order drug combinations can be predicted using different analytical formulas or machine learning methods. Note that analytical formulas must be distinguished from machine learning models, because no parameters are learned when using them.

### 2.2.1 Analytical formulas

There exist several formulas modeling the responses of combinations of drugs. The general idea is to predict the response of a combination with a formula involving responses of lower order combinations. Those formulas are not actual machine learning models because the idea is to measure responses of lower order combinations and then use the formulas. However, there is no learned model and no training step. Some of those formulas make the assumption that only pairwise responses are enough to predict the response of high order combinations (any order greater than two). This hypothesis will be developed later (see SECTION 3). The different formulas and their hypothesis are summarized in TABLE 2.
Note that the response of a drug $i$, noted $g_i$ always depends on the dose at which it is considered: $g_i(D_i)$. To ensure the simplicity of the notations, we will simply write $g_i$.

**The Bliss Formula**[34] is the most basic formula. It simply assumes that the response of a drug combination is the product of all the mono-therapy responses of the drugs in the combination, taken

at their respective dose: $g_{1,...,n}(D_1, D_2, ..., D_n) = g_1(D_1)g_2(D_2)...g_n(D_n)$.

**The Dose model[33]** is an extension of the Bliss formula and uses only mono-therapy and pairwise responses to predict the responses of high order drug combinations (any order greater than two). The basic idea is to use effective doses instead of actual doses in the combination: $g_{1,...,n}(D_1, D_2, ..., D_n) = g_1(D_{1eff})...g_n(D_{neff})$
The effective doses aim at reflecting the interaction between two drugs. It means that the presence of a second drug has an impact on the effect of the first drug (and vice versa). This is modeled via the effective doses (each real dose has its corresponding effective dose). If the effective dose of the first drug is higher than the real one, it means that the second drug enhances the effect of the first drug. If it is lower, the second drug limits the effect of the first one. If the effective dose is equal to the real one, the second drug does not impact the effect of the first one (no interaction between the drugs).

The main hypothesis of the dose model is that pairwise interactions are enough to fully characterize the responses of a drug combination of any order. The Dose model is detailed in Section 2.4.

**The Pair model[34]** is the following: $g_{1,...,n} = (\prod_{ij} g_{ij})^{1/(n-1)}$. It assumes that one can rely only on pairwise data to predict responses of higher order combinations. This model has an interesting property, called the Bliss conservation property: if the pairs are assumed to be Bliss pairs ($g_{12} = g_1 g_2$), then the formula is the same as the Bliss model.

**The Iserliss formula[34]** is another formula. The formula for a triplet of drugs is the following : $g_{123} = g_1 g_{23} + g_2 g_{13} + g_3 g_{12} - 2g_1 g_2 g_3$. For a quadruplet of drugs, it gives: $g_{1234} = g_{12} g_{34} + g_{13} g_{24} + g_{14} g_{23} - 2g_1 g_2 g_3 g_4$. This formula makes the same assumption as the Pair model: one can rely only on pairwise data to predict responses of higher order combinations. As the Pair model, the Iserliss formula satisfies the Bliss conservation property. To the best of our knowledge, this formula has not been tested on cancer drugs yet but seems to be very efficient for antibiotics (especially for triplets and quadruplets)[33].

**The regression model[34]** is the last formula and is the following for a triplet of drugs: $g_{123} = \frac{g_{12} g_{23} g_{13}}{g_1 g_2 g_3}$. As the last two formulas, it assumes that pairwise data are enough for higher order predictions. It also satisfies the Bliss conservative property.

## *Side note*

All the presented formulas require to know those mono-therapy responses at the doses of the drug combinations. This is not very convenient because it does require quite a lot of measurements. One can avoid this by fitting a curve (typically a Hill curve) to some mono-therapy data points. By doing that, one can then have the mono-therapy response of a particular drug for any desired dose. Machine learning techniques can thus decrease the number of required measurements.

All the formulas presented above can be divided in two parts: from one hand we have the Bliss formula and the Dose model and on the other hand we have the Pair model, the Iserliss formula and the regression model. The difference remains in the needed measurements to implement the formula. The Bliss formula needs only mono-therapy data. The Dose model needs mono-therapy data and pairwise data to get its parameters but not especially at the doses of the higher order combinations (because data interpolation is performed).
On the contrary, all the other formulas require the pairwise data at the doses of the higher order

combination. That makes them more complicated to use because, once again, the main problem in this field is data availability.

| | Formula | Hypothesis |
|---|---|---|
| Bliss | $g_{1,...,n}(D_1, ..., D_n) = g_1(D_1)...g_n(D_n)$ | No interaction between drugs |
| Dose model | $g_{1,...,n}(D_1, ..., D_n) = g_1(D_{1eff})...g_n(D_{neff})$ | Pairwise interactions are sufficient |
| Pair model | $g_{1,...,n}(D_1, ..., D_n) = (\prod g_{ij})^{1/(n-1)}$ | Pairwise interactions are sufficient |
| Iserliss formula | $g_{123}(D_1, D_2, D_3) = g_1g_{23} + g_2g_{13} + g_3g_{12} - 2g_1g_2g_3$ | Pairwise interactions are sufficient |
| Regression model | $g_{123}(D_1, D_2, D_3) = \frac{g_{12}g_{23}g_{13}}{g_1g_2g_3}$ | Pairwise interactions are sufficient |

Table 2: Summary of the different formulas and their hypothesis to predict the responses of higher order drug combinations.

### 2.2.2 Machine learning models

By contrast with analytical formulas, machine learning models are able to predict the responses of a drug combination at any doses. Indeed, the models learn a relationship between the doses (and all the other input data, depending on the model we consider) and the responses. Once the model is learned, one can thus obtain the predicted responses for any unseen combination of drugs (the combination itself can be unseen or just the doses of the combination can be unseen). While machine learning methods seem to be much more interesting than simple formulas, they are also much more complicated to obtain. Indeed, the main challenge of learning a model is the learning data which has to be in high quantity and of high quality. This section describes some of the existing machine learning models used in this context.

**DECREASE**[24] (Drug Combination RESponse prEdiction) is a machine learning model able to efficiently predict missing entries in a pairwise dose-response matrix. The input to the DECREASE model is incomplete dose-response matrices (including the mono-therapy data for each drug in the pair), it thus only considers pairs of drugs. The algorithm is constructed in 3 steps:

- Outliers detection: it is possible that some measurements (of mono-therapy responses or of pairwise responses) may be wrong and considered as outliers. The algorithm detects them in order not to use them as training data. All the detected outliers are removed from the data and now considered as unknown.
- Full matrix prediction: the model predicts the missing entries of the dose-response matrices, as well as the outliers that have been removed in the first step. The predicted values are percentage of inhibition, between 0 and 100. The model uses a weighted non-negative matrix factorization (NMF) algorithm to make its predictions.
- Synergy scoring: now that the dose-response matrices are complete (the missing entries have been predicted in the previous step), DECREASE can use different synergy models to assess the synergy score of each pair of drugs. It allows the detection of synergistic or antagonistic combinations of drugs.

DECREASE has been tested for both cancer and non-cancer drugs. It achieved great performance in terms of synergy score prediction. However, the model is for now only able to predict the responses of pair of drugs and has not been extended to higher order combinations yet.

**DeepSynergy**[29] is the first use of deep learning in order to predict drug synergy. Deep learning is defined as a field of machine learning, based on artificial neural networks. DeepSynergy is able to predict synergy scores for pairs of drugs. The considered synergy score is based on the Loewe additivity model (Section 2.1.5). As input, the model uses the chemical descriptors of the two considered drugs as well as genomic information of the considered cell line.
DeepSynergy has been trained on a dataset composed of 23 062 combinations of 2 drugs in a cell line. The dataset comprises 39 different cell lines, in a total of 7 different tissue types.
The model has been compared to four other machine learning models. DeepSynergy performed better than the other models. It gives good results while predicting synergy score on novel pairs of drugs. However, the results are significantly less good while predicting synergy scores on novel pairs *and* novel cell lines (and this is the case for all the models tested). This can be due to the small size of the used dataset.

**Dr.VAE**[30] (Drug Response Variational Autoencoder) is a deep generative model. The model uses a Variational Autoencoder to map the expression of the genes before and after the effect of a drug (via a latent embedding), with a (logistic regression) classifier. This classifier learns from the latent embedding of the VAE, and predicts probability of drug responses. Dr.VAE is able to predict the responses of single drugs (mono-therapy responses), which makes it less interesting than other models, especially in the context of this work.

**TreeCombo**[25] is another machine learning model used to predict the synergy score of pairs of drugs. TreeCombo uses extreme gradient boosted trees, which sequentially combines tree-based models. TreeCombo predicts the synergy score based on the Loewe additivity model (Section 2.1.5), in the same way as DeepSynergy. Furthermore, it uses the same features as input, namely the chemical descriptors of the drugs and the gene expression level of the cell lines. The model has also been tested on the same dataset as DeepSynergy. TreeCombo showed 10% better performance than DeepSynergy.
Furthermore, TreeCombo is said to be more interesting because its predictions are more easily interpretable.

**ComboFM**[27] is a Machine Learning method able to efficiently predict pairwise responses of cancer drugs. It uses Higher Order Factorization machines to learn the parameters of non linear regressions.
It is the very first building block of this work and will thus be explained in details in Section 2.4.

### *Side note*

All the models presented above are, at best, able to predict pairwise responses but no higher order drug combinations. This highlights a little bit more the goal of this work and its motivation.
Furthermore, ComboFM is one of the few models predicting responses of drug combinations, and not only synergy scores. This constitutes a significant advantage of this model.

## 2.3 Dose Model

### 2.3.1 General principle

The Dose Model has been firstly introduced in 2016[33]. This model was built to predict the response of high order drug combinations (for any number of drugs), by taking as input the mono-therapy data of each drug and pairwise data of each pair of the drugs in the combination, and giving as output the response of the drug combination at different doses. This model was originally designed

for antibiotics but it can also be used on cancer drugs. The Dose model is based on the Bliss independence model, which has been described above (Section 2.2.1). The Dose model improves the Bliss independence model by modifying the hypothesis: only pairwise interactions between drugs have an impact on the overall effect of the drug combination (instead of considering that mono-therapy data contain enough information). The Dose model modifies the Bliss model by introducing the effective dose of each drug. The formula becomes:

$$g_{1,...,n}[(D_{1eff},...(D_{neff})] = g_1(D_{1eff})...g_n(D_{neff}) \tag{11}$$

where the effective doses model the pairwise interactions. They are computed by dividing the real dose by a Michaelis-Menten-like factor:

$$D_{ieff}(D_i) = D_i \prod_{i \neq j} (1 + a_{ij} \frac{D_{jeff}/D_{0j}}{1 + D_{jeff}/D_{0j}})^{-1} \tag{12}$$

where $a_{ij}$ is called the interaction parameter and represents the impact of the drug i on the effective dose of drug j, and $D_{0j}$ is the halfway point of drug j (the doses at which the effect if 50% of the maximum effect). The model thus relies on the assumption that pairwise interactions are sufficient to predict the response of high order combinations.

The Dose model works as a succession of different steps, each of them being an estimation/compuation of a parameter.

**Mono-therapy responses** The first step is to obtain the expression of the mono-therapy response of each drug: $g_i$. The mono-therapy $g_i(D_i)$ data is fitted using a Hill curve (estimation of the parameters $D_{0i}, n_i$):

$$g_i(D_i) = \frac{1}{1 + (D_i/D_{0i})^{n_i}} \tag{13}$$

**Pairwise response** This step aims at computing the interaction parameters $a_{ij}$ and $a_{ji}$ for each pair of drugs. Those parameters are derived by taking into account the known mono-therapy responses and pairwise responses (*i.e.* the responses of a pair of drugs). Based on the mono-therapy responses (Equation 13), one can write and resolve the equations of the effective dose (in the pairwise combination) of each drug:

$$\begin{aligned} g_{12} =& g_1(D_{1eff})g_2(D_{2eff}) \quad \text{with} \\ D_{1eff} =& D_1(1 + a_{12}\frac{D_{2eff}/D_{02}}{1 + D_{2eff}/D_{02}})^{-1} \\ D_{2eff} =& D_2(1 + a_{21}\frac{D_{1eff}/D_{01}}{1 + D_{1eff}/D_{01}})^{-1} \end{aligned} \tag{14}$$

**Effective dose in the whole combination** Once all the model parameters have been estimated $(D_{0i}, n_i)$ and computed $(a_{12}, a_{21})$, one can resolve Equation 12 and then Equation 11 for each dose of interest (the ones we want to predict). For each real doses of the drugs in the combination, one can obtain the corresponding effective and thus the response.

### 2.3.2 Performance of the Dose model

The Dose model has been tested with different order of combinations (up to 10) and different types of drugs[33].

Firstly, the model has been used to predict the responses of triplets and quadruplets of antibiotics[33]. The dataset was composed of 6 triplets and 2 quadruplets, with a total of 1 384 dose combinations. The $R^2$ of those predictions is comprised between 0.85 and 0.93, which corresponds to good performance.

Secondly the Dose model has been tested on three anticancer drugs (the same 3 drugs that will be investigated in this work, see SECTION 5.2.1). The three drugs have been tested at 8, 7 and 5 different doses respectively, which gives 280 different points.
The $R^2$ on the predictions of the all three-drug interaction matrix is equal to 0.82. It is once again corresponding to good performance.

In a second paper on the Dose model[34], even higher order combinations have been tested. The order of the tested combinations goes up to ten drugs (antibiotics).
Three to ten drugs combinations of antibiotics for *E. Coli* were predicted. The $R^2$ of those predictions is always between 0.88 and 0.96. The RMSE is between 0.04 and 0.07 . Knowing that the response is always between 0 and 1, a maximal error of 0.07 is totally satisfying. This represents 7% of error. This range of error seems acceptable in such a context. Indeed, knowing that a combinations of drug is efficient at 85% or 78-92% does not change much the conclusion. The reverse observation is the same, if a combination is efficient at 10% for example (meaning it is not effective), it is not very important to know if it is 3% or 17%.
Finally, the model has been used to predict the effects of tuberculosis drugs. 9 drugs have been selected to form 4 triplets, 4 quadruplets and 1 quintuplet. The $R^2$ on the predictions is always greater than 0.9 and the RMSE around 0.06.

The Dose model has thus proved that it can achieve a very good level of performance and is reliable to predict the effect of higher order drug combinations, even when this order goes up to 10 drugs.
The performance on cancer drugs are satisfying but only triplets have been tested. We thus do not know if the Dose model is efficient for combinations of cancer drugs of order greater than 3.

**Summary** The Dose model takes as input the mono-therapy data of each drug and pairwise data of each pair of the drugs in the combination. As an output, the model is able to predict the response of the drug combination at different doses.
As an example, if one wants to have the response of a triplet of drugs A,B,C, one would need to have data about the 3 mono-therapy responses in addition to responses data of the 3 pairs: AB, BC and AC. It is not mandatory that the doses at which the input data are taken correspond to the doses at which we want to predict the responses of higher-order combination. This is a big advantage of the Dose model.

## 2.4 ComboFM

Machine learning tools applied to the problem of higher order drug combinations have been developed much more recently than the analytical formulas presented above (see SECTION 2.2.1). The main reason of this late development is the lack of training data. Indeed, an efficient machine learning tool needs a lot of data to train. Collecting enough data is a real challenge due to the number of

different drugs, cell lines and the high heterogeneity of cancer cells.
The following of this section aims at developing the ComboFM model[27].

### 2.4.1 General principle

ComboFM[27] is a new machine learning framework, able to efficiently predict the effect of pairs of cancer drugs. ComboFM uses high order tensors to model the cell context-specific drug interactions. Even if ComboFM is not the only machine learning tool able to predict drug responses[24] [30], it has several important advantages.
The main one remains in the prediction in itself. Indeed, while most of the other computational tools such as DeepSynergy and TreeCombo (see Section 2.2.2) directly predict drug combination synergies (*i.e.*, if the effect of the combination is greater than expected), ComboFM predicts the actual response (*i.e.*, the effect of the combination) as a percentage of growth. This is a very interesting point, because discovering synergistic combinations constitutes the final goal of drug combination experiments. However, there exist different synergy quantification methods (Section 2.1.5). These synergy quantification methods are based on the responses of the drugs and combinations. Therefore, predicting the responses and not a synergistic classification allows to choose the synergy quantification method and to compare the methods. ComboFM uses the NCI ComboScore as a synergy classifier. The NCI ComboScore has been introduced with the NCI-ALMANAC dataset and is based on the Bliss independence score.
The second main advantage is the fact that ComboFM is able to integrate a lot of information. Indeed, in addition to the needed inputs of the model, ComboFM can handle additional inputs. Those additional inputs are generally describing the drugs and cell lines. Typically, chemical and genomic descriptors are used. In this context, the genomic descriptors used are gene expression profiles of the cancer cell lines. They were obtained using the *rcellminer* R package. Concerning the chemical descriptors of the drugs, molecular fingerprints were used. Molecular fingerprints are series of bits, representing the absence (0 value) or the presence (1 value) of particular substructures in the molecule. The longer the series of bits, the more precise the molecular description. In this work, fingerprints of type 'E-state' were selected. It consists of 79 bits, corresponding to the E-state atom types, as defined in the *rcdk* R package. The E-state fingerprint is based on electrotopological state (E-state) indices, which encodes information about associated functional group, graph topology, and the Kier-Hall electronegativity of each atom[19].
Finally, ComboFM is one of the few computational tools able to predict one response for each dose combination (and not only one prediction for each pair of drugs).

### 2.4.2 Detailed method

The general principle of ComboFM is the use of higher-order factorization machine (HOFMs) (see Section 2.1.1 for further information). The FM models the interactions between the different modes of a high order tensor. The tensor is indexed by the different drugs, their concentrations and the cell lines (and potentially additional information).

ComboFM is able to predict responses of pairs of drugs in three different scenarios (as shown in Figure 6[27]. Note that in all the scenarios, ComboFM knows the mono-therapy responses of the drugs in different cell lines. This means that even more challenging scenarios could be tested. As an example, one could remove the mono-therapy measurements in some cell lines.

**Data representation** In order to integrate all the information regarding the interactions between the drug combinations in the cell lines (depending also on the doses of each drug), ComboFM uses a

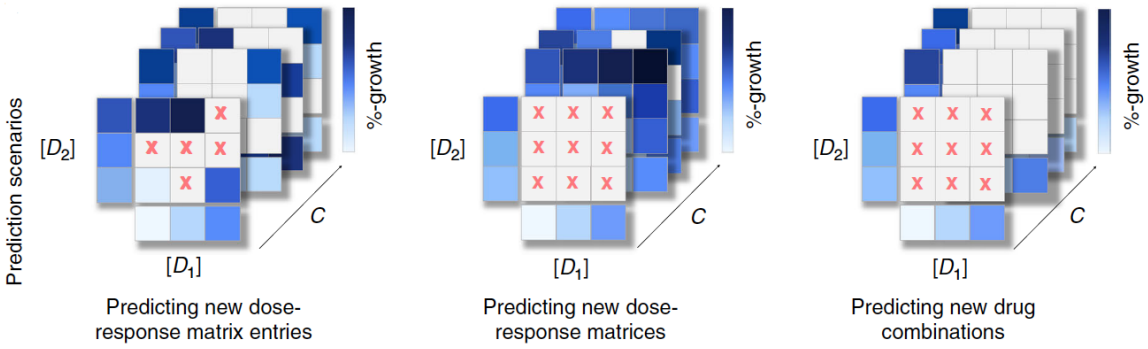



Figure 6: Representation of the 3 different scenarios of ComboFM. Image taken from [27]

fifth-order data tensor, referred to as $\mathbf{X}$ (FIGURE 7). For the learning step of the machine learning model, the $\mathbf{X}$ tensor is flattened as a 2D array (FIGURE 8). In that 2D array, each row vector $\mathbf{x}$ corresponds to a single entry of $\mathbf{X}$. $y_i$ is the responses associated to the $i^{th}$ row of $\mathbf{x}$. Given this response, ComboFM models the relationship between the data using higher-order factorization machines (HOFMs). The data is thus represented in a one-hot encoding way.
Note that each entry of the flattened array $\mathbf{x}$ is duplicated. This is necessary in order to ensure that the pair of drug A-B is considered as the same as the pair of drug B-A.

**Higher-order factorization machines (HOFMs)** Higher-order factorization machines learn a non-linear regression model from $\mathbf{x}$ to the output $y$, see SECTION 2.1.1 for further information. It is done by estimating regression weights $w_{i_1,...,i_t}$ for each combination of input features $x_{i_1,...,i_t}$, where $t$ is the order of the interactions (set to 5 in this case). One advantage of using factorization machines instead of polynomial regression is that the weights are estimated using factorized parametrization (instead of estimating them separately). In higher-order factorization machines, the weights are coupled via the multiplication of the latent factors learned (FIGURE 9). This difference constitutes an advantage because it avoids computational and statistical problems that could arise from the direct estimation of the total weight tensor $\mathbf{W}$. Furthermore, coupling the weights allows to learn efficiently on very sparse tensors, which is typically the case in such a context.

In ComboFM, the weights of the model are learned by minimizing a regularized mean squared error:

$$min\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i(\mathbf{x}_i))^2 + \frac{\beta_1}{2}||\mathbf{w}||^2 + \sum_{t=2}^{m}\frac{\beta_t}{2}||w_{i_1,i_2,...,i_m}||^2 \tag{15}$$

where $\beta_1, ..., \beta_m > 0$ are the regularization parameters. ComboFM uses a uniform set of regularization parameters ($\beta_1 = ... = \beta_m$) in order to limit the numbers of parameters to optimize.

In ComboFM, the optimization of the weights (and thus the learning of the models) are implemented using *TFFM* [28]. *TFFM* is a TensorFlow implementation, able to handle factorization machines of any order. In order to optimize the weights, TFFM uses gradient-based optimization and a logistic loss.

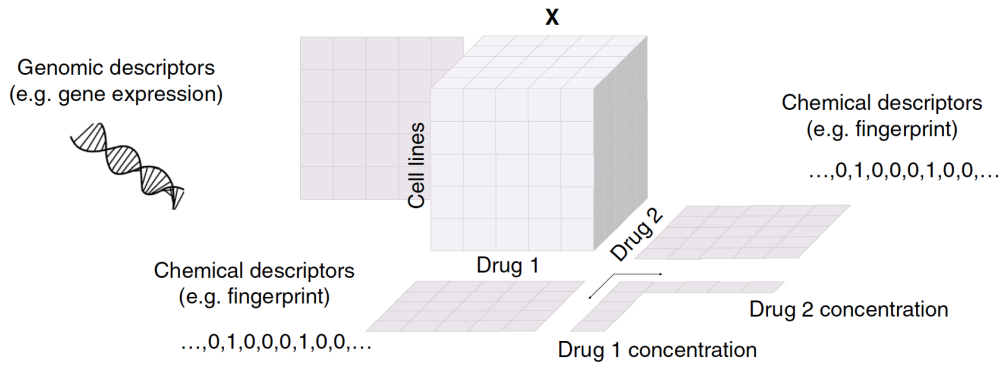



Figure 7: Representation of the fifth-order tensor $\mathbf{X}$ of ComboFM. Image taken from [27].

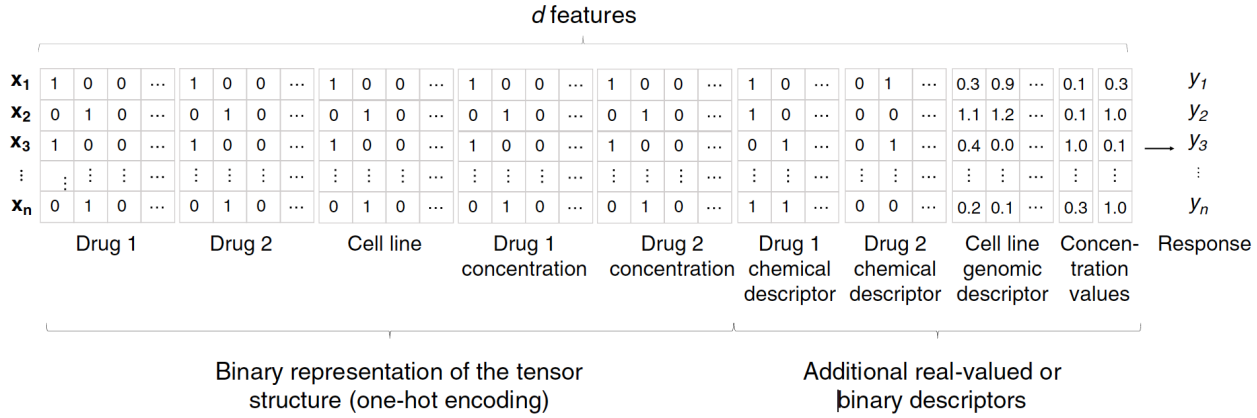



Figure 8: Representation of the flattened array $\mathbf{x}$ of ComboFM. Image taken from [27].

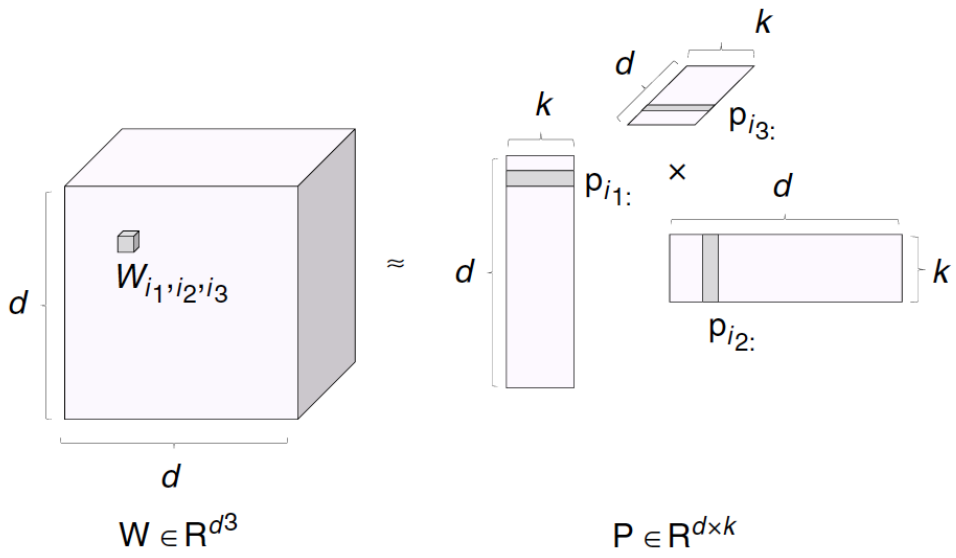



Figure 9: The parameters of the models $W_{i_1,i_2,\ldots,i_t}$, where $t$ is the order (equal to 3 in this example) and $i_1, i_2, \ldots, i_t$ are the features of input data are learned by the high order factorization machines. $k$ is the rank of factorization while $d$ is the total number of features of input data. Image taken from [27]

### 2.4.3 Performance of ComboFM

The performance of ComboFM has been assessed[27] by using a nested cross-validation, made of 5 inner folds and 10 outer folds. As it has been explained in Section 2.1.2, the nested cross-validation is used in order to avoid the risk of overfitting while evaluating the model.
The performance was evaluated by taking the average on the 10 outer folds of the cross-validation. The three possible scenarios of ComboFM have been tested, see Table 3 for the performane. The performance has been assessed on a subset of the NCI-ALMANAC dataset (see Chapter 4 for further information regarding the NCI-ALMANAC dataser). This subset was composed of 50 anti-cancer drugs approved by the FDA. 617 different combinations of two of those drugs were considered, across 60 different cell lines. Those cell lines are originated from 9 different tissue types[27]. In total, this constitutes a dataset of 333,180 drug combination response measurements and 222,120 monotherapy response measurements.
ComboFM is able to reach great performance in the three scenarios. The poorer performance is, as expected, achieved in the last scenario since there are less data to train the model.

| | RMSE | Pearson | Spearman |
|---|---|---|---|
| New dose-response entries | 9.86 | 0.97 | 0.91 |
| New dose-response matrices | 10.39 | 0.91 | 0.9 |
| New pairs of drugs | 13.04 | 0.95 | 0.88 |

Table 3: Summary of the performance of ComboFM, in the three different scenarios. The values are the averages performance on the 10 outer folds of the nested cross-validation.

### 2.4.4 Summary

ComboFM is an effective tool to predict the responses of pairs (and only pairs) of drugs. It is very suitable for the goal of this work because of the nature of its predictions. Indeed, the dose-response predictions (rather than synergy predictions) are required in order to use the Dose model, as it wil be explained in Section 5.

## 2.5 Conclusion

This chapter aimed at explaining the needed concepts in terms of machine learning and drug interactions that will be used in the rest of this work, as well as the different models used. Related works were also presented. Finally, the two building blocks of this work namely the Dose model and ComboFM have been detailed.

# Chapter 3 Predicting effect of high order combinations of drugs

This chapter will explain several possibilities to predict the responses of higher order drug combinations using ComboFM, as well as their corresponding hypothesis.

## 3.1 Working hypothesis

The goal of this work is to manage to predict the effect of high order drug combinations (meaning combinations of order higher than 2), by using ComboFM. It turns out that there are mainly two point of views regarding the interactions between the drugs in a combination. These two points of view will be developed in this section and the rest of this work will focus on the first one.

Before going into the details of these hypotheses, one should keep in mind the following definitions of the different types of interactions inside a mixture of drugs. Below is a reminder of those important definitions:

- Pairwise interactions are all the interactions between only 2 drugs. Inside a combination of 3 drugs for example, we find 3 pairwise interactions.
- Emergent interaction is the interaction between all the drugs of the mixture. There is only one emergent interaction in a given mixture of drugs.
- Net interactions are the set of all the interactions. It does includes the pairwise interactions, the emergent interaction but also all the others if we consider a combination of order greater than 3.

Let us illustrate those types of interaction with two examples.
Firstly, one can consider a mixture of 4 drugs: A, B, C and D. In that case, there are 4 pairwise interactions: A-B, B-C, C-D and D-A. There is only one emergent interaction: A-B-C-D. The net interactions are all the already cited ones *and* the other interactions: A-B-D, B-C-D and A-C-D.
Let us do the same illustration with a triplet of drugs (Figure 3): ABC. In that case we have 3 pairwise interactions: A-B, B-C, C-D. Once again, there is only one emergent interaction: A-B-C. In this case there is no other interaction, the net interactions are only the pairwise and the emergent ones.

### 3.1.1 Hypothesis 1

The first point of view is the one adopted by the Dose model but also by all the analytical formulas presented in Section 2.2.1. It states that emergent interactions are very rare. That implies that pairwise interactions contain enough information in order to be able to efficiently predict the effect of higher order combinations of drugs. This hypothesis has been validated by the performance of the different formulas presented, especially the Dose model. However one has to keep in mind that those formulas have been mostly tested with antibiotics. To the best of our knowledge, studies using the Dose model on combinations of more than three cancer drugs have not been published. There is thus no guarantee that those formulas are valid for 'very' high order combinations of anti-cancer drugs. Indeed, the contexts in which anti-cancer drugs are used are way more complicated than the ones in which antibiotics are used. As a consequence, it is possible that those simple formulas are not able to capture all this complexity.

Making this hypothesis allows the use of all the formulas and models presented in Section 2.2.1. Note that the Bliss model does not need this assumption to be valid. Indeed it does not consider pairwise data at all, but it considers only mono-therapy data. The assumption of the Bliss model is that there is no interaction between the drugs.

The purpose of this work is to investigate the validity of this hypothesis by making predictions of the responses of higher order combinations of cancer drugs based on only pairwise (and mono-therapy) data.

### 3.1.2 Hypothesis 2

On the contrary, the second hypothesis states that emergent interactions are not rare at all and contain a lot of information that is needed in order to predict responses of higher-order drug combinations[31]. This statement comes from a study made on antibiotics[31]. It has to be kept in mind because the same conclusion may not be valid for cancer drug and no corresponding study has been made yet, to the best of our knowledge. The study was performed on 8 different antibiotics, giving 251 two-drug combinations, 1512 three-drug combinations, 5670 four-drug combinations and 13 608 five-drug combinations. The type of interactions present in the combinations between the drugs was investigated by computing the expected response with and without considering the emergent interactions. This analysis gave rise to the conclusion that the frequency of higher-order interactions (meaning not pairwise interactions) increases with the number of drugs in the mixture. It means that there exist emergent interactions and modeling them is important to accurately capture what is happening between the different drugs of a combination.

Under such an assumption, all the models and formulas presented in Section 2.2.1 can not be used. By extending this statement to cancer drugs, the only solution to predict the responses of higher order drug combination using ComboFM, is to explicitly extend the machine learning model. It is of course theoretically feasible but it implies dealing with a very high order tensor (the rank of the tensor in ComboFM would not be 5 but much higher). It would lead to a very costly computation in terms of time and resources. Furthermore, this is not the main faced obstacle if we want to extend ComboFM. Indeed, the major problem is the lack of data regarding high-order cancer drug combinations that would be needed to train the model. The built tensor would mostly be empty, which would make the learning very difficult.

## 3.2 Conclusion

Ideally the goal is to be able to reduce the number of needed measurements in order to predict the responses of a combination. It is the case when using the hypothesis 1 but not with hypothesis 2. Thus the second hypothesis is to use as a second option, in the event that the first hypothesis is not verified.
As already mentioned, this work focuses on the first hypothesis which implies the use of ComboFM in combination with another model (the Dose model). The purpose is thus to investigate the validity of this first hypothesis by making predictions of responses of higher order cancer drug combinations based on only pairwise and mono-therapy data. The way we decided to combine those two models will be explained in Chapter 5.

# Chapter 4 Data

This chapter aims at explaining the data used in this work, its interpretation and the resulting constraints. It is important to understand the data in order to use it correctly. Firstly, the interpretation of the responses in the dataset is detailed. Secondly, a small statistical analysis of the dataset used is performed.

## 4.1 NCI-ALMANAC

### 4.1.1 General overview

The *National Cancer Institute* (from the US National Institutes of Health) released in 2017 one of the largest databases regarding pairs of cancer drugs, called the **NCI-ALMANAC**[1]. The dataset includes 105 different cancer drugs, tested in 60 different cell lines. The cell lines are the so-called NCI-60 cell lines. Each drug is first tested alone, at different doses. Then, more then 5,000 pairs of drugs (approved by the Food and Drug Administration) are tested in the NCI-60. All together, the ALMANAC includes more than 3 million data points.

### 4.1.2 Detailed description of the dataset

**About the responses**

In order to measure the effect of a drug, one usually measures the growth of the cells, with and without the presence of drug(s). Based on those measurements, one can compute the percentage of growth in presence of the drug, with respect to the growth without the presence of the drug. There are two different computations of the percent growth available in the NCI-ALMANAC, one considering the zero time measurement and the other one not considering it. Those two ways of computing the response will play an important role in this work, see Section 5.2.2.
The difference between both is that considering zero time measurement allows to detect lethality. Indeed, lethality happens when the percent growth is a negative value, meaning that cells have been killed (there are less cells at the end). The growth percentages are computed based on three mean optical density measurements:

- $T_z$ (zero time measurement): is the measured growth before adding any drug.
- C (control growth): is the endpoint measurement without any drug, two days after adding the drug(s) in the test well.
- $T_i$ (growth under test concentration): is the endpoint measurement of the reference (control) well, two days after adding the drugs.

Then, the percent growth with time zero is computed via the following equations[23]:

$$\begin{aligned} &\text{If } T_i \geq T_z: \\ &\qquad \text{PercentGrowth} = 100\frac{T_i - T_z}{C - T_z} \\ &\text{Else:} \\ &\qquad \text{PercentGrowth} = 100\frac{T_i - T_z}{T_z} \end{aligned} \tag{16}$$

The percent growth without time zero is computed via the equation:

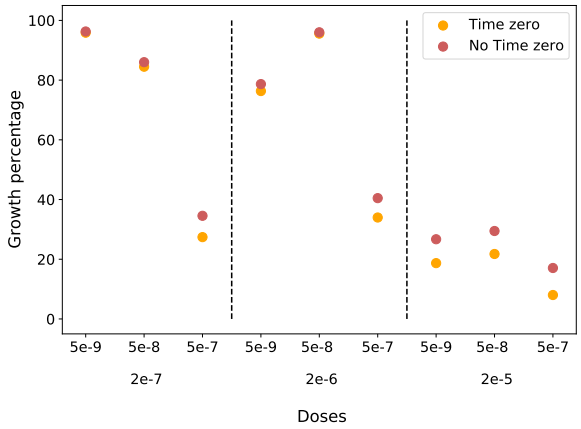
$$\text{PercentGrowthNOTZ } = 100\frac{T_i}{C} \tag{17}$$

One can easily see that the PercentGrowth can be negative while the PercentGrowthNOTZ can not be negative. The interpretation is thus different and data with zero time gives more information by detecting lethality.

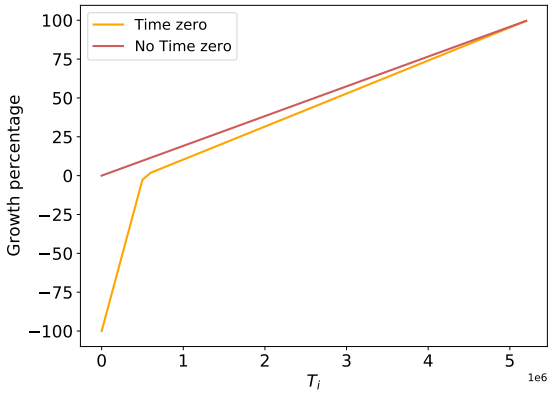
The interpretation of the different growth percentages is the following[23]:

- A value equal to 100 means no growth inhibition. Indeed, in that case we would have $T_i = C$ which means that adding a drug did not change the natural growth of the cells.
- A value greater than 100 means that the drug has a negative effect. Indeed, this means that the targeted cell line grows more when there is no drug than when there is a drug. In that case one has $T_i > C$ meaning that growth is larger than the one without any drug.
- A value between 0 and 100 means growth inhibition. A value of 60 for example means that $T_i$ is equal to 60% of C. Thus, the drug inhibited the growth at 40%.
- A value of 0 means no net growth. Note that a value of 0 can only be obtained with time zero measurements, when $T_i = T_z$.
- A negative value means lethality. As an example, if the percent growth is equal to -40, that means that we have $T_i$ is equal to 60 % of $T_z$. There are thus less cells compared to the zero time, which means lethality.
- A value of -100 means that all cells are dead.

See Figure 10 for an example. Figure 10 a) shows the two different growth percentage data points that are available in the NCI-ALMANAC dataset and Figure 10 b) shows the theoretical curves of growth percentage based on Equations 16 and 17. Both figures concern one pair of drugs in one particular cell line.
The larger the percentage, the more similar the two measures are. As we are getting closer to the threshold value ($T_i = T_z$), the difference between the percentage with and without time zero becomes more and more important.




(a) Pairwise growth percentages (with and without time zero measurements).




(b) Comparison of the theoretical curves of growth percentages (with and without time zero measurement).

Figure 10: Example of data in the NCI-ALMANAC dataset for the pair cisplatin and doxorubicin in the A549 cell line.

### Statistical analysis of the dataset

The NCI-ALMANAC is a very large dataset and this section aims at explaining its content in a more detailed way. It is interesting to have a clear view of the data (see TABLE 4 for a summary):

- In total, the dataset contains 3 686 475 drug responses among 105 different cancer drugs, in 60 cancer cell lines. These drug responses comprise mono-therapy data and dose-responses matrices of pair of drugs.
- The different pairs of drugs (5 355 pairs in total) have been tested across the 60 cell lines, see the distribution on FIGURE 12. Nearly all the pairs have been tested in more than 50 cell lines, while a third approximately has been tested in all the cell lines.
- The dose-reponse matrices in the NCI-ALMANAC dataset can be of two different sizes : 5×3 or 3×3 (corresponding to the number of doses at which the drugs have been tested), depending on the screener (organisms) who made the measurements. As shown on FIGURE 11, the vast majority of matrices are of size 3×3.
- Concerning mono-therapy data, all of the 105 drugs are represented in each of the 60 cell lines. The number of data points may differ from one drug to another, as it will be explained in SECTION 5.2.

| # lines of data | 3 686 475 |
|---|---|
| # of cell lines | 60 |
| # of drugs | 105 |
| # of pairwise matrices | 311 604 (3x3: 300 091; 5x3: 11 513) |
| # of mono-therapy data | 812 961 |

Table 4: Summary of the NCI-ALMANAC dataset.

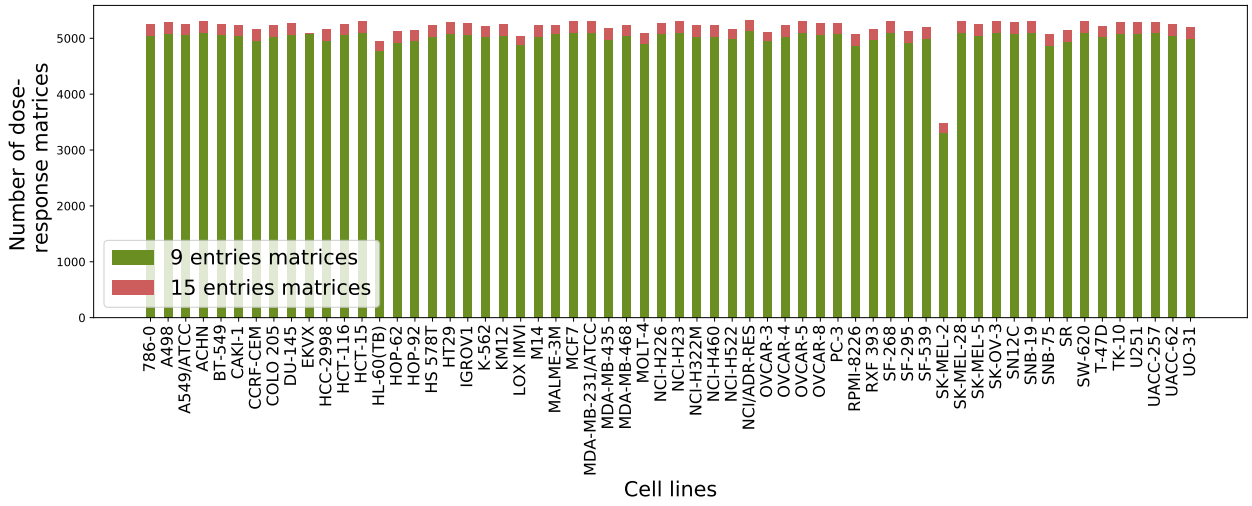



Figure 11: Number of dose-responses matrices in each cell line in the NCI-ALMANAC dataset.

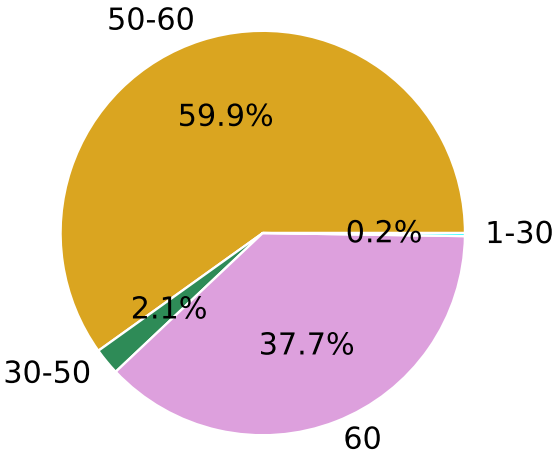



Figure 12: Pie chart representing in how many cell lines the responses of each pair of drugs have been measured.

### 4.1.3 Subset used in ComboFM

In this work (*i.e.* in all the experiments presented in Chapter 6), ComboFM is used on a subset of the NCI-ALMANAC dataset. This part explains the difference between both datasets and the reasons behind them.
The summary of the subset of the NCI-ALMANAC used in ComboFM is shown in Table 5 and Figure 13.

| # lines of data | 4 591 714 |
|---|---|
| # of cell lines | 60 |
| # of drugs | 104 |
| # of pairwise matrices | 299 973 (all 3x3) |
| # of mono-therapy data | 1 891 957 |

Table 5: Summary of the subset of the NCI-ALMANAC dataset used in this work.

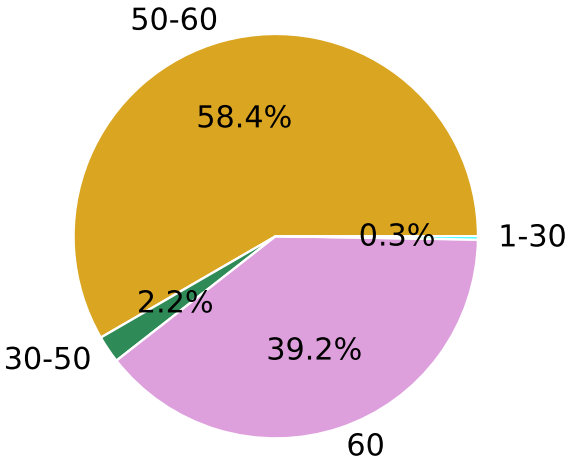



Figure 13: Pie chart representing in how many cell lines the responses of each pair of drugs have been measured, in the dataset used in ComboFM.

The main difference between the two datasets is the fact that all the 5×3 matrices present in the NCI-ALMANAC dataset have been removed. The reason is that ComboFM considers all the dose-responses matrices of the same size. If there are 5×3 matrices, all the 3×3 matrices will be considered as having missing data. This is not convenient. Furthermore, there are very few 5×3 matrices compared to the number of 3×3 so ignoring them does not have a big impact on the quantity of data used to learn the model.
Note that the matrices of interest for the experiments done in this work (see Section 5.2.1) are all 3×3 matrices, which allows to ignore the 5x3 matrices. Those pairs of interest are present in 60, 52 and 43 cell lines, respectively.

The second difference is that it looks like there is more data in the subset, especially regarding mono-therapy data. It is in fact not the case. Let us take an example to understand this difference in the number of lines in the two datasets.
Consider two pairs of drugs: A-B and B-C. Each drug is taken at 3 different doses: $D_1$, $D_2$ and $D_3$. We thus have 9 data points for each dose-response matrix. Additionally, there are 3 mono-therapy data points for each drug. The NCI-ALMANAC dataset encodes only once each mono-therapy data, while the dataset used in ComboFM encodes the mono-therapy data as part of the dose-response matrix (see FIGURE 15 in CHAPTER 5). For this example, the lines of mono-therapy data in the two datasets are presented in TABLES 6 and 7. We clearly see that in the second table, the mono-therapy data of drug A is duplicated for each matrix. This explains the big difference in the number of lines between the two datasets. This data duplication is necessary for ComboFM to consider the dose-response matrices separately.

Concerning the drugs, one of them has been removed from the NCI-ALMANAC by the authors of ComboFM. It has no impact because, once again, it does not concern the drugs nor cell lines of interest.

| Drug 1 | Concentration 1 | Drug 2 | Concentration 2 |
|---|---|---|---|
| A | $D_{1A}$ | / | / |
| A | $D_{2A}$ | / | / |
| A | $D_{3A}$ | / | / |
| B | $D_{1B}$ | / | / |
| B | $D_{2B}$ | / | / |
| B | $D_{3B}$ | / | / |
| C | $D_{1C}$ | / | / |
| C | $D_{2C}$ | / | / |
| C | $D_{3C}$ | / | / |

Table 6: Example of mono-therapy data in the **complete NCI-ALMANAC dataset.**

| Drug 1 | Concentration 1 | Drug 2 | Concentration 2 |
|---|---|---|---|
| A | $D_{1A}$ | B | 0 |
| A | $D_{2A}$ | B | 0 |
| A | $D_{3A}$ | B | 0 |
| B | $D_{1B}$ | A | 0 |
| B | $D_{2B}$ | A | 0 |
| B | $D_{3B}$ | A | 0 |
| A | $D_{1A}$ | C | 0 |
| A | $D_{2A}$ | C | 0 |
| A | $D_{3A}$ | C | 0 |
| C | $D_{1C}$ | A | 0 |
| C | $D_{2c}$ | A | 0 |
| C | $D_{3C}$ | A | 0 |

Table 7: Example of mono-therapy data in the **subset of the NCI-ALMANAC dataset used in ComboFM.**

## 4.2 Conclusion

This chapter aimed at presenting the available data to work in the context of predicting the response of anti-cancer drugs. Furthermore, the interpretation of the data is presented and is important in order to understand to results of the experiments that will follow. Finally, the subset of the complete NCI-ALMANAC dataset that is used in this work has been presented.

# Chapter 5 Combining ComboFM and the Dose model

The purpose of this work is to use ComboFM in order to predict the effect of high order combinations of cancer drugs. By considering the hypothesis 1 (explained in Section 5), one can also use the Dose model to make such predictions. This chapter thus explains how the two models (ComboFM and the Dose model) can be combined in order to predict the responses of higher order anti-cancer drug combinations.

## 5.1 Procedure

This section will explain the followed procedure in order to combine the two models. As a reminder, the Dose model takes as input the mono-therapy responses of each drug (at different doses) as well as the pairwise responses of all the pairs of the drugs (at different doses too).
Combining ComboFM with the Dose model means using predictions of ComboFM as input of the Dose model, instead of using the data from the NCI-ALMANAC dataset.
Let us illustrate this with an example. Consider that we want to predict the response of a triplet of drugs: A, B and C. In order to predict those responses with the Dose model, we need the mono-therapy responses of A, B and C, as well as the pairwise responses of the pairs AB, BC and AC. This is where ComboFM is interesting. Indeed, ComboFM is able to predict pairwise responses. We can thus use ComboFM to predict the pairwise responses AB, BC, AC that are needed in the Dose model, instead of using the ones from the dataset. However, the pairs of drugs are considered at different dose combinations (as explained in Section 4.1.2). This implies that we have different responses for each pair, each one corresponding to one combination of doses. As a consequence, we can choose to use ComboFM to predict *all* the pairwise data or only *some of them*. This defines different scenarios of the use of ComboFM. The impact of this choice will be investigated in Section 6.5. See Table 8 for a summary of the input and output of the models. Depending on the considered scenario of ComboFM, the number of pairwise responses between the drugs of the combinations (AB, AC, BC in the considered example) coming from the NCI-ALMANAC dataset used in the Dose model will vary. If we predict all the pairwise responses with ComboFM, there is no value regarding those pairwise responses coming from the NCI-ALMANAC dataset.

Figure 14 represents a schematics of the combination of the two models in order to predict responses of higher order drug combinations. In this example, one diagonal of the dose-responses matrices is predicted by ComboFM while the other entries are considered as known. The entries that are considered as known are used as input in both models.

To sum up, here are the steps to follow in order to use the combination of ComboFM and the Dose model :

- Define the scenario of use of ComboFM. It means that we need to choose which entries of the dose-response matrices we want to predict (*i.e.* those are in the test set) and which ones we consider as known (*i.e.* those are in the training set).
- Use ComboFM to predict the chosen entries.
- In the input file of the Dose model, replace the values of the NCI-ALMANAC dataset by the ones predicted with ComboFM in the previous step.
- Use the Dose model and predict the response of the higher-order combinations of anti-cancer drugs.

| | Input | Output |
|---|---|---|
| ComboFM | • Mono-therapy (A,B,C)<br>• (Pairwise responses (AB, AC, BC))<br>• Pairwise responses : other<br>• Additional info | Pairwise responses (AB, AC, BC) (depending on the scenario) |
| Dose model | • Mono-therapy responses (A,B,C)<br>• (Pairwise responses (AB, AC, BC)) | Predictions of triplet ABC (any dose we want) |
| Dose model after ComboFM | • Mono-therapy (A,B,C)<br>• Pairwise responses (AB, AC, BC)<br>• (Pairwise responses (AB, AC, BC)) | Predictions of triplet ABC (any dose we want) |

Table 8: Summary of the input and output of the different models used in this work, in the example case where we want to predict the responses of the triplet of drugs ABC. Color representation : orange: data from the NCI-ALMANAC dataset; blue:predictions from ComboFM.

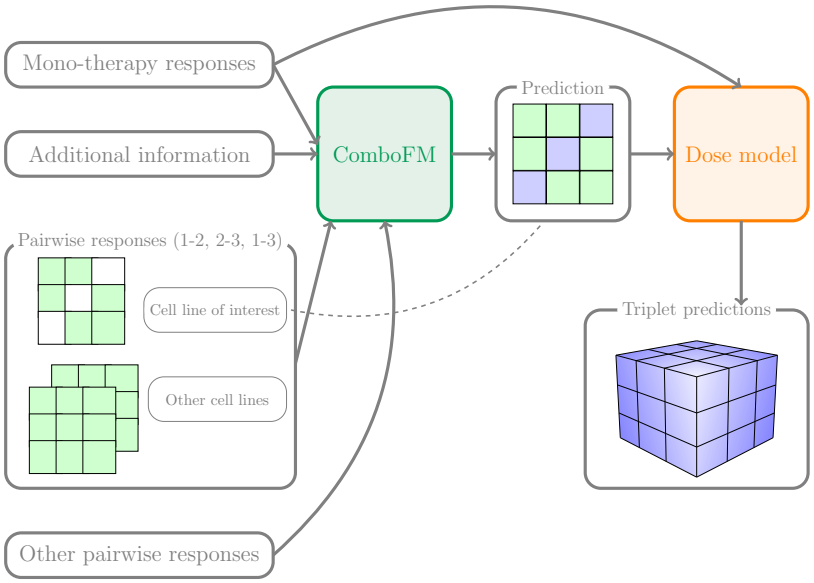



Figure 14: Schematics of the combination of ComboFM and the Dose model. This example corresponds to a combination of three drugs and corresponds to the scenario two of the experiments. ComboFM predicts some entries of the matrices of interest (depending on the chosen scenario) and the Dose model uses those predictions as input. Color representation: white = unknown; green = known; blue = predicted.

## 5.2 Challenges

The combination of the two models is not straightforward. There are some challenges that need to be tackled in order to ensure the consistency between both models but also respect the hypothesis of each model. This subsection explains those challenges and the choices that have been made to overcome them.

### 5.2.1 How to assess the predictions?

As it has already been mentioned, the main challenge in this work is validation data. There is not much experimental data concerning the responses of high order combinations of cancer drugs. Furthermore, consistency between training and testing data must be ensured. It means that the same cell line must be considered.

ComboFM is based on the NCI-ALMANAC dataset. The needed validation data is thus concerning cancer drugs which are included in the NCI-ALMANAC (and more precisely, included in the subset of the NCI-ALMANAC, as presented in Section 4.1.3). Furthermore, the number of cell-lines included in the NCI-ALMANAC is also limited. All together, we need experimental data about high order combinations of drugs included in the NCI-ALMANAC, in a cell line also included in the NCI-ALMANAC. These requirements make the research quite complicated and the possibilities quite limited.

The ideal case is when detailed validation data is available. Detailed data is when we have a sufficient number of measured doses (each drug of the combination at a given dose)-responses (in the form of growth percentage, as explained in Section 5). In that case, the Dose model can predict the growth percentage for each dose. Then the predicted valued can be directly compared to the experimental one.
It is the ideal case but also the least common one.

The Dose model has mainly been tested on antobiotics but the authors also tested the model in a triplet of three cancer drugs, in one cell line. Those three drugs and this cell line are available in the subset of NCI-ALMANAC dataset used. Those drugs and this cell line will be referred to as the drugs and cell line *of interest* for the rest of this work. The three drugs are: Cisplatin, Doxorubicin (hydrochloride) and Paclitaxel (also known as Taxol) and the cell line is A549(/ATCC), being a lung cancer cell line.
The authors of the Dose model made the measured responses of the triplets at each combination of the doses presented in Table 9 available. There are 7 different doses for doxorubicin, 8 for cisplatin and 5 for paclitaxel. This gives 280 different dose combinations of the triplets.
Note that those doses are presented in $\mu M$ while the NCI-ALMANAC dataset gives the concentrations in $M$. As a consequence, one must be careful to correctly convert the data.

| | 1.37e-2 | 4.10e-2 | 1.23e-1 | 3.7e-1 | 1.11 | 3.33 | 10 | 20 |
|---|---|---|---|---|---|---|---|---|
| Doxorubicin | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | |
| Cisplatin | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Paclitaxel | ✓ | ✓ | ✓ | ✓ | ✓ | | | |

Table 9: Doses (in $\mu M$) of each drug in the triplet.

However, having this data does not correspond to the ideal case explained above. Indeed, the experiments were not made using the same protocol as the ones of the NCI-ALMANAC dataset. As a consequence, it is difficult to assess the validity of the comparison between those data and the predictions. Furthermore, we do not know if the formula used to express the response is the same as the ones used in the NCI-ALMANAC. We can however note that all the responses are bounded between 0 and 1, which can correspond to the growth percentage without time zero of the NCI-ALMANAC (Equation 17).

When there is no (or not enough) validation data in order to evaluate the new model, one can use another method. The goal of this work is to investigate whether we can combine ComboFM with the Dose model, instead of using only the Dose model. The predictions of the new combined model can thus be compared to the ones of the Dose model only. The predictions made by the Dose model alone are thus used as ground truth.
This can be done even when there is no validation data at all. It does not allow to know if the predictions are accurate but it allows to evaluate the accuracy of the new combined model, in comparison with the Dose model alone.

### 5.2.2 Dose model

#### Data in general

The first challenge remains in the data used in the Dose model. As it has already been explained in Section 4, the NCI-ALMANAC dataset contains two ways of interpreting the response of a drug/drug combinations: the growth percentage with time zero measurement, and the growth percentage without time zero measurement (Equations 16, 17). The first one can be negative while the second one can not be negative.
As it has been explained, the first step of the Dose model is the fitting of mono-therapy data with a Hill curve. Typically, a Hill curve has the following shape: $y(x) = \frac{1}{1+(x/a)^b}$ with $a, b \geq 0$. The Hill curve is thus bounded between 0 and 1 (and as a consequence can not be negative).
The first consequence is that the data from the NCI-ALMANAC must be divided by 100.
The second and most important consequence is that one can *not* use the percentage with time zero in the Dose model. Indeed, a Hill function is not able to fit negative values. Thus, using negative values will surely give a bad fitting result. That means that, in the Dose model, we have no other choice than using the measurement without time zero.
However, ComboFM uses measurements *with time zero*, since it gives more information. The predictions of ComboFM thus correspond to the growth percentage of Equation 16. In order to use the Dose model after the use of ComboFM, one has to convert data:

$$\begin{aligned} \text{if} \quad ComboFM_{prediction} > 0 : & \\ & T_i = \left( \frac{ComboFM_{prediction}}{100} (C - T_z) \right) + T_z \\ \text{else} : & \\ & T_i = \left( \frac{ComboFM_{prediction}}{100} T_z \right) + T_z \\ \text{PredictionNOTZ} = \frac{T_i}{C} & \end{aligned} \tag{18}$$

Note that the values of $C$ and $T_z$ are available in the NCI-ALMANAC, there is one value for each dose-response matrix of drug pairs. One will then obtain a positive value, and this value is then compatible with the Dose model.

**Mono-therapy data**

Another challenge remains in the fitting of mono-therapy data with a Hill curve. In the NCI-ALMANAC, the experiments for each 3×3 dose-response matrix can be seen as a 4×4 matrix, including the mono-therapy responses (FIGURE 15). For each experiment, one thus has 3 data points for mono-therapy response of each drug. In this work, three drugs are of interest, each of them being in two dose-response matrices. That gives a maximum of six mono-therapy points for each drug.

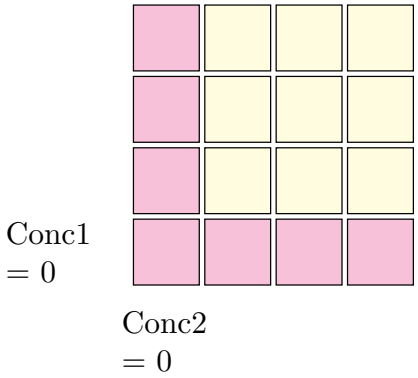



Figure 15: Representation of the dose-response matrices available in the NCI-ALMANAC. Color description: pink = mono-therapy data; yellow = pairwise data.

However, if several dose-responses matrices have been measured inside the same study, they will share mono-therapy data. In the end, one has six or three data points for mono-therapy data for the drugs of interest. It is a very low number of points and this may impact the quality of the fitting. It is important to understand that the NCI-ALMANAC dataset groups different studies and screeners (organisms that made the measurements). Let us take an example to understand the impact of this. Consider two pairs of drugs, A-B and A-C. Let us assume that drug A is taken at the same doses $(D_{1A}, D_{2A}, D_{3A})$ in the two dose-response matrices and that those two response-matrices have been made in different studies. It implies that we have two different values for the mono-therapy response of drug A, for each of the three considered doses.
This is illustrated in FIGURE 16, by taking Cisplatin as an example. One can clearly see that the variance of mono-therapy data for one dose is surprisingly high. As a consequence, there exists no single curve able to fit all the points. This implies that a choice must be made regarding the mono-therapy data points that are used. The choice made in this work is detailed in SECTION 6.2

## 5.3 Pairwise data

In this work, pairwise data is exclusively based on the NCI-ALMANAC dataset. For each pair of drugs, the dataset provides the responses for 9 combinations of doses (3 different doses for each drug). This gives a 3×3 dose-response matrix for each pair. The doses of the three drugs in the dose-response matrices of interest are summarized in TABLE 10. This means that we have, at most,

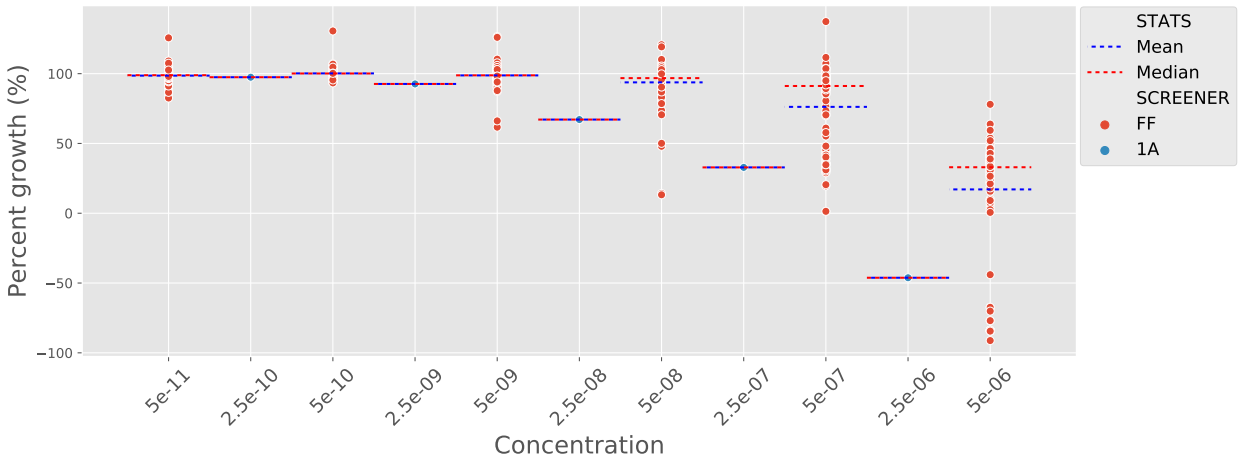



Figure 16: All mono-therapy data for doxorubicin available in the NCI-ALMANAC dataset. The data is grouped by screener.

$3 * 3 * 3 = 27$ data points predicted by ComboFM (corresponding to the scenario in which we use ComboFM to predict the entire dose-response matrices).

| | Dose 1 | Dose 2 | Dose 3 |
|---|---|---|---|
| Doxorubicin | 5e-9 | 5e-8 | 5e-7 |
| Cisplatin | 2e-7 | 2e-6 | 2e-5 |
| Paclitaxel | 3e-9 | 3e-8 | 3e-7 |

Table 10: Doses of the drugs in the three dose-responses matrices of interest. The doses are expressed in $M$.

## 5.4 Conclusion

This chapter explained how we can combine the two models of interest in this work: ComboFM and the Dose model. Even though the two models can theoretically be easily combined, this combination rises some challenges that were explained, as well as the chosen solutions to overcome them.

# Chapter 6 Experiments

This chapter will present the experiments that have been made and their results. As it has been explained in Section 5.2.1, the experiments are mainly done on a triplet of drugs.
This chapter is divided in different sections. The first one explains the questions the experiments are designed to answer. Secondly, the parameters and settings that are common to all the experiments are detailed. The validation procedure is then described. Finally, the different experiments and their results are presented.

## 6.1 Goal of the experiments

The experiments have been designed in order to address the following questions:

- Is the Dose model suitable in this context? The model is based on a fitting of the data using a Hill function. The idea is to evaluate the goodness of the Hill function in this context. Slightly modified Hill functions will be tested and compared.
- What is the best learning set-up of data when using ComboFM (meaning which entries of the matrices of interest should be predicted and which ones should be considered as known)? In order to do that, different scenarios (a scenario is defined as a division of the pairwise matrices between training and testing sets) will be tested.
- How sensitive are the models? We will simulate the impact of noisy predictions of ComboFM. This will give an idea of the sensitivity of the Dose model to the predictions of ComboFM.
- How does the Dose model compare to other analytical formulas presented in Section 2.2.1?
- What is the impact of using ComboFM in combination with the Dose model on synergy predictions? In order to answer this question, different synergy scores will be computed and compared. The idea is to see if small changes in the predictions of the Dose model or ComboFM have a big impact on the synergy prediction.
- What about order combinations of drugs greater than 3? No measured responses are available in that case. As a consequence, one can not measure the accuracy of the predictions but one can evaluate the impact of using ComboFM before using the Dose model. This will be done on a quadruplet and a quintuplet of drugs.

Note that for the last four parts, mainly one scenario of the second part will be considered.

## 6.2 Parameters setting and tuning

For all the different parts, there is a general set-up which remains constant. This general set-up is the following:

- Cross-validation: the same cross validation (CV) has been performed in ComboFM. This is a 5-fold CV, aiming at the optimization of two parameters: the regularization parameter and the rank of factorization (see Section 2.4). The tested values for the regularization parameter are $10^2, 10^3, 10^4, 10^5$ and the tested values for the rank of factorization are $25, 50, 75, 100$, as it has been made in the ComboFM paper [27]. Note that a test set has been kept apart from the cross-validation such that it is never used in the training set. The different scenarios modify the training and test sets.

- Final model of ComboFM: the regularization parameter and the rank of factorization of the final model are chosen according to the results of the cross-validation. For each value of each parameter, the mean of the Pearson coefficients (between the predicted responses and the real ones) is computed over the 5 folds. The chosen value of parameter is the one corresponding to the highest mean of Pearson coefficients[27].
- The data used in all the experiments comes from the NCI-ALMANAC dataset. As a reminder, a subset of the NCI-ALMANAC has been used, as it is explained in Section 4.1.3. As a consequence, the response of a drug is always expressed as a growth percentage and the interpretation of this growth percentage is the one explained in Section 4.1.2.
- Epochs: 100 epochs have been used in each training inside the cross-validation, while 200 epochs have been used for the final training[27].
- The learning rate of the optimizer has been set to 0.001[27].
- The order of the factorization machine has been set to 5, as the rank of the input tensor of ComboFM. This value has been chosen according to the ComboFM paper[27].
- Transition to the Dose model: once ComboFM has made its predictions, the predictions are converted in order to have the percentage of growth without the time zero so that the values are non-negative and expressed as percentage between 0 and 1 (as explained in Section 5.2). It has been done in each of the experiments presented.
- Mono-therapy data of the Dose model: Each mono-therapy point of a drug is often duplicated, depending on the number of different studies in which this drug has been evaluated. In this work, it has been decided to keep the mono-therapy data coming from the studies which made the dose-response matrices of interest (see Table 11). Paclitaxel and doxorubicin have been assessed in two different studies but at the same doses, it means that one has six mono-therapy points for those drugs, at 3 different doses (since the dose-responses matrices are 3×3). On the other hand, cisplatin has been assessed in only one study for the three dose-response matrices of interest. We thus only have three mono-therapy data points for this drug (at three different doses).

| Pair | Study |
|---|---|
| Doxorubicin-Paclitaxel | CD040CP38 |
| Cisplatin-Paclitaxel | CD086CP130B |
| Cisplatin-Doxorubicin | CD086CP130B |

Table 11: Studies in which the dose-response matrices of interest have been assessed in the NCI-ALMANAC dataset.

## 6.3 Validation procedure

As a reminder, in all the following experiments, the predictions will be made for a triplet of drugs (Paclitaxel (also known as Taxol), Doxorubicin and Cisplatin) in one particular cell line of lung cancer (A549). Measured responses of this triplet in 280 combinations of doses are available, as explained in Section 5.2.1[33].
However, the followed protocol is not exactly the same as the one followed in the NCI-ALMANAC dataset. That means that predicting those responses and then only comparing the predictions to the measurements may be risky. Indeed, we can not be sure that those data are comparable and we thus

expect that predictions may not fully match the measurements. In order to perform a more rigorous validation, the predictions will also be made once using only the Dose model (and the mono-therapy and pairwise data coming from the NCI-ALMANAC, without using ComboFM). The performance (in terms of $R^2$ and $RMSE$) of those predictions (see TABLE 12) will be used as ground truth. Note that those reference values will be slightly modified with the first part of the experiments.
The goal of this work is to see if we can use ComboFM in combination with the Dose model. Thus, the interesting point is to compare the performance when we use this combination of models versus only the Dose model. The performance in itself does not matter so much in this context.

| $R^2$ | RMSE |
|---|---|
| 0.4833 | 0.1638 |

Table 12: Performances of the Dose model using only data from the NCI-ALMANAC dataset (without using ComboFM at all). Those results are the ones obtained with the original Hill function. They will be improved in SECTION 6.4

For the two last parts of experiments, the measured responses for the triplet can not be used as validation data. In that case, the validation procedure is modified. The predictions made using only the Dose model will be directly compared to the predictions (and not the performance of the predictions) made by using ComboFM and the Dose model, as it has been explained in SECTION 5.2.1. The two sets of predictions will be compared based on the $R^2$, the $RMSE$ and the Spearman and Pearson correlation coefficients.

## 6.4 Experiments part 1: Hill function

This first experiment aims at addressing the first question, namely 'Is the Dose model suitable in this context?'
As a reminder, the Hill function used in the Dose model in order to fit the mono-therapy data has the following shape: $y(x) = \frac{1}{1+(\frac{x}{a})^b}$. It is thus upper bounded by 1, because $a$ and $b$ are positive.
However, the percentage without time zero can be higher than 1 (see EQUATION 17).
Intuitively, the growth percentage should not exceed 1 (meaning 100 %). Indeed, such a value means that the drug has a negative effect: the growth with the drug is higher than the one without the drug. One would thus logically consider that those values are due to noise in the measurements.
Whether it is noise or not, it does not seem optimal to fit data that is sometimes higher than 1 with a function that can not be higher than 1. There are several ways to deal with this. Different solutions will be investigated in this part.

### 6.4.1 Analysis of the data

Before going into the different proposed solutions, let us take a look at which entries of the dose-response matrices of interest are concerned by this.
In the data of interest, there is one drug (doxorubicin) having mono-therapy data points higher than 1 (two out of the six points). Those two points correspond to the same dose, the lowest one. The two values are 1.0209 and 1.0239. The fact that those values correspond to the lowest dose and that the values are really close to 1 confirms the intuition that this is simply noise. Indeed, by definition if we use a very low dose of a drug, we expect to have a response close to 1 (and a response equal to 1 if we do not use any drug). As a reminder the response is expressed as a growth

percentage in regards to the targeted cells. A value of 100% (or 1) means that the growth under the presence of the drug is the same as the one without any drug.
Note also that there are 4 of the 9 points of the pair doxorubicin-paclitaxel points that are higher than 1 (see FIGURE 17 for a visual representation). This helps to understand the impact of the fitted Hill function on the score. Indeed, if the two fitted functions for doxorubicin and paclitaxel are bounded by 1, the product of them will never exceed 1 and thus the fitting of pairwise data can not be optimal.

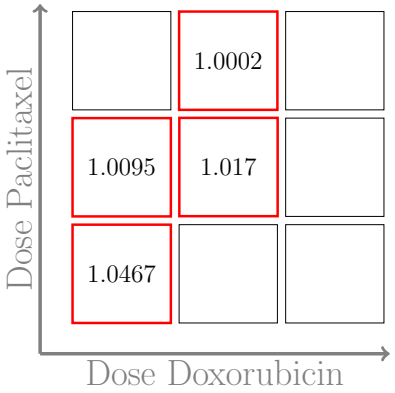



Figure 17: Representation and values of the entries higher than 1 of the dose-response matrix for the pair doxorubicin-paclitaxel.

### 6.4.2 How to deal with it?

This part explains the different tested solutions in order to deal with those values higher than 1. Note that these experiments are done only using the data coming from the NCI-ALMANAC dataset, without using the predictions of ComboFM. Considering only those data should be sufficient given the way the Dose model is built. As a reminder, the fitting using the Hill function is only done on mono-therapy data. Those functions are then used to fit pairwise responses by assuming that the pairwise response is the product of the mono-therapy responses (meaning a product of the fitted Hill functions). Since ComboFM is only used to predict pairwise responses (and not mono-therapy responses), the data on which the Hill functions will be fitted is not impacted by the use of ComboFM.

#### Truncating the data

Firstly, one can consider that values higher than 1 are simply noise. As a consequence, we can consider that any value larger than 1 should actually be equal to 1 and we can thus truncate the data.

#### Changing the Hill function

Another way to overcome this problem is to modify the Hill function, so that it becomes more consistent with the data. There are different possibilities to do so and several of them have been tested. As a reminder, the shape of the Hill function is $y(x) = \frac{t}{1+(\frac{x}{a})^b}$ with $t = 1$. The value of $t$ can be modified in order to change the upper bound of the curve.

**t = max(data)**

The first possibility is to allow the Hill function to be higher than 1 only when it is necessary and to the highest value only. The fitting with a Hill function has been set to:

$$\begin{aligned} &\text{if } \max(x) > 1: \\ &\qquad t = \max(x) \\ &\text{else}: \\ &\qquad t = 1 \\ y(x) &= \frac{t}{1 + (\frac{x}{a})^b} \end{aligned} \tag{19}$$

This choice of modification of the Hill function used in the Dose model can be motivated by the "International Union of Pharmacology Committee on Receptor Nomenclature and Drug Classification"[32]. The relationship between the dose $[A]$ of drug A and the effect $E$ is defined as follows:

$$\frac{E([A])}{E_{max}} = \frac{[A]^{n_H}}{[A]^{n_H} + [A]_{50}^{n_H}} \tag{20}$$

where $E_{max}$ is the maximal effect of the drug A and $[A]_{50}$ is the drug concentration for which the effect is 50% of $E_{max}$. The equation can be rewritten as:

$$\frac{E([A])}{E_{max}} = \frac{1}{1 + (\frac{[A]_{50}}{[A]})^{n_H}} \tag{21}$$

this equation has the same shape has the one used in the Dose model, except the fact that Equation 21 is normalized by the maximal effect. This suggests that the Hill function used in the Dose model makes the assumption that $E_{max} = 1$, which is not true when we work with the NCI-ALMANAC dataset.

**t = fixed value**

Choosing the maximum value of the data may not be the best option. In order to investigate this, several values of $t$ between 0.95 and 1.2 have been tested. In this case, $t$ is fixed and always the same regardless of the values of data.

The results (in terms of the $R^2$ score) of those different experiments are shown on Figures 18 and 19. In particular, Figure 19 shows the fitted Hill functions for the only drug having mono-therapy data points higher than 1 (which is doxorubicin).

The performance highly depends on the chosen value. Truncating the data does not seem to be the best option. Intuitively, this was not expected since those values higher than 1 were supposed to be due to some noise. However, the fact that keeping those values gives better results tells us that there is some information on those data that is important to predict the triplet responses.
By looking at Figure 18, the best Hill function for this data seems to be the one with $t = 1.15$. It tends to show that the classic Hill function (with $t = 1$) is not optimal for the NCI-ALMANAC dataset. Choosing $t = max(data)$ gives a slightly better result than $t = 1$. The difference is not high because the maximum value is 1.0239, which is close to 1. Then, increasing the value of $t$ gives

better results until the value is too high (for $t = 1.2$).

Let us now take a look at FIGURE 19 where one can see the impact of $t$ on the fitted Hill curve. Choosing a value of $t$ lower than 1 is the worst option. Indeed, the concavity of the fitted Hill curve for this value of $t$ is the other way around and it is not suitable for dose-responses data.
One can clearly see that the curve corresponding to $t = 1.15$ is not the one that best fits the data (see the first data point where $Dose = 0.005$, which is quite far below the fitted curve). The curve that best fits the data is the one with $t = max(Data)$. This observation tells us that in order to obtain the best performance in terms of triplet prediction, we do not need to obtain the best fitting of mono-therapy data points. This means that using a value of $t = 1.15$ may overfit the triplet data points and is thus maybe not a good choice. However, we do not have enough validation data to verify whether this overfits the data or not.
We would probably need more higher-order drug combinations data to evaluate the overall best value of $t$. Furthermore, it is possible that the optimal value of $t$ also depends on the order of the combination.

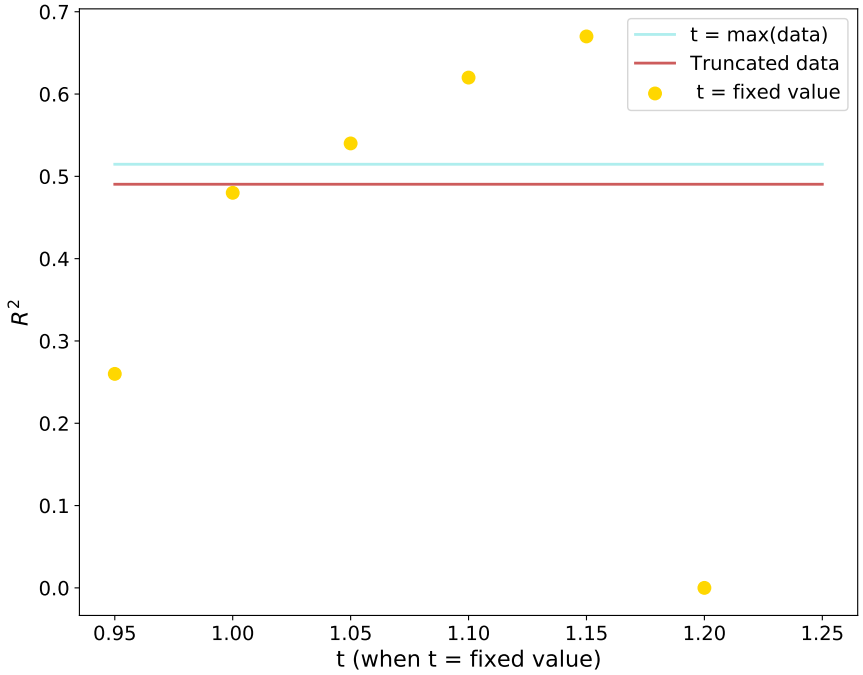



Figure 18: Performance of the Dose model depending on the chose Hill function/data set-up.

### 6.4.3 Conclusion

To conclude, this experiment highlights the fact that the current Dose model using a classic Hill function may not be totally suitable for the NCI-ALMANAC dataset, in order to predict responses of higher-order drug combinations. This can be surprising because of the values of data and their interpretation (those values higher than 1 could/should only be noise).
Consequently, in the rest of experiments, the chosen value of $t$ is $t = max(data)$ so that potential

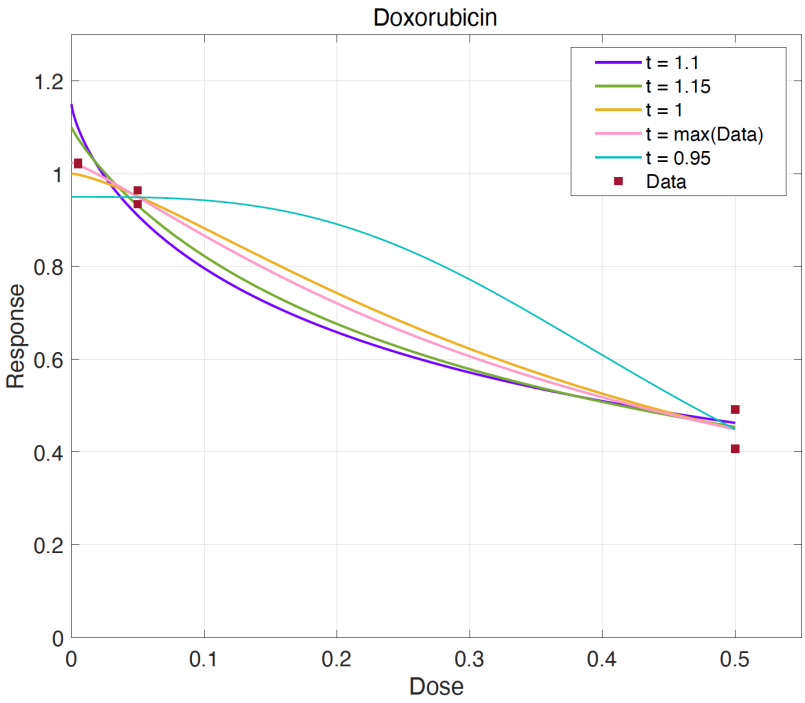



Figure 19: Result of the fitting of mono-therapy data for doxorubicin for different values of $t$ (Equation 13).

overfitting should be limited while allowing values higher than 1. As a consequence, the new reference values for the performance of the Dose model on the triplet are presented in Table 13.

| $R^2$ | RMSE |
|---|---|
| 0.5147 | 0.1587 |

Table 13: Performance of the Dose model using only data from the NCI-ALMANAC dataset (without using ComboFM at all). Those results are the ones obtained with the modified Hill function ($t = max(data)$)

## 6.5 Experiments part 2: evaluation of different scenarios

### 6.5.1 Set up of the scenarios

This second part of experiments aims at answering the following question : what is the best learning set-up of data when using ComboFM? There are indeed different possibilities regarding the predictions made by ComboFM. In order to answer this, different predictions scenarios of ComboFM will be tested. Let us now describe the specific set up for each scenario. The scenarios are different in the selected entries for the test set (and thus in the training set). The goal of testing different scenarios is to evaluate if the Dose model (and/or ComboFM) is sensitive to the number (and/or the positions) of the predicted entries of the dose-response matrices.

As a reminder, we have three matrices of interest in this work. Those matrices are of size 3×3. In all the tested scenarios, the three matrices will follow the same scenario (meaning the same entries will go in the test set). At least one of the entries should be put in the test set (which gives at least

three entries to predict in total since all the matrices follow the same scenario). At most, the entire matrices can be predicted which gives a test set of 27 points.

There exist a lot of different possible scenarios to test, and thus choices have to be made. The tested scenarios have been chosen according to one very important idea: it has to be realistic. The purpose is to decrease the number of experiments a lab has to do. It is thus very important to match the different possible conditions of a real lab. If we try to consider real-life conditions, the lab will decide two things : how many entries of the matrices they want to measure and then which ones. It means that it is very unlikely that a lab will randomly choose which entries of a matrix he wants to predict (and measures the rest).
Furthermore, it has been shown that some patterns are more interesting than others[24], especially diagonals.

Once the scenarios have been selected, one has to be really careful regarding the set-up of the cross-validation. It *has* to match the prediction scenario. Let us take an example (see FIGURE 20 for a representation of this example): the chosen scenario is the prediction of one diagonal of each matrix of interest (the same diagonal for the three matrices), based on the other entries of those matrices, and other complete matrices. Saying that the cross-validation set-up must match the prediction set-up means that, in each step of the cross-validation, the model has to predict *only* diagonals of matrices. As a consequence, the test sets of the cross-validation are only composed of diagonals of matrices, while the training sets are composed of the rest of the entries (complete matrices except the diagonal) and complete matrices.
In the training set used to make the predictions, we find the rest of the matrices of interest and all the other complete matrices.

Note that **all** the mono-therapy data are **always** in the training set (for the cross-validation and the final model). ComboFM assumes that mono-therapy data is known.

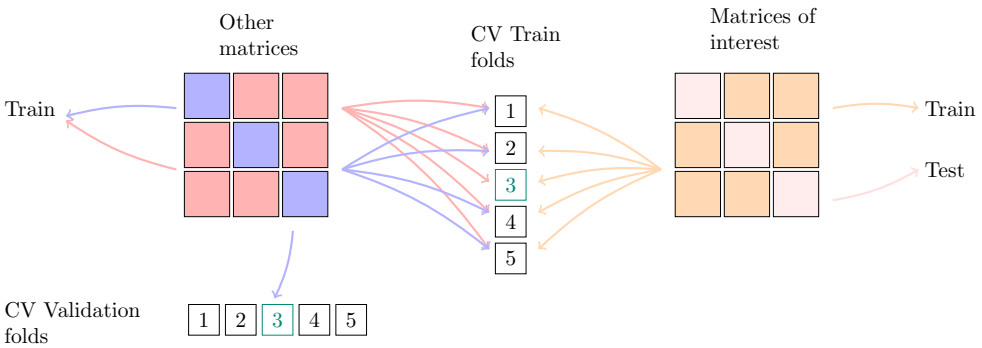



Figure 20: Example of one (5 folds) cross-validation set-up. The diagonal of the matrices of interest are kept apart in the test set. The other entries of those matrices are put in every train sets of the CV. For the other matrices, the diagonal is put in one validation fold of the CV (randomly chosen, the third one for the example). The other entries are put in all the training folds *except* the one chosen as test set. The entire matrices (not of interest) are also put in the final train set.

The chosen scenarios are the following, corresponding to FIGURE 21:

- **Scenario 1**: diagonal 1 in the test set.
- **Scenario 2**: diagonal 1 in the training set (meaning the predicted entries are all the entries but the diagonal).
- **Scenario 3**: diagonal 2 in the test set.
- **Scenario 4**: diagonal 2 in the training set.
- **Scenario 5**: four corners in the test set. That corresponds to entries "Min-Max", "Max-Max", "Min-Min" and "Max-Min".
- **Scenario 6**: four corners in the training set.
- **Scenario 7**: two opposed corners in the test set, corresponding to "Min-Max" and "Max-Min" entries.
- **Scenario 8**: two opposed corners in the training set, corresponding to "Min-Max" and "Max-Min" entries.
- **Scenario 9**: other two opposed corners in the test set, corresponding to "Min-Min" and "Max-Max" entries.
- **Scenario 10**: other two opposed corners in the training set, corresponding to "Min-Min" and "Max-Max" entries.
- **Scenario 11**: the entire matrices in the test set.

Each scenario has its mirroring scenario (except the last one).

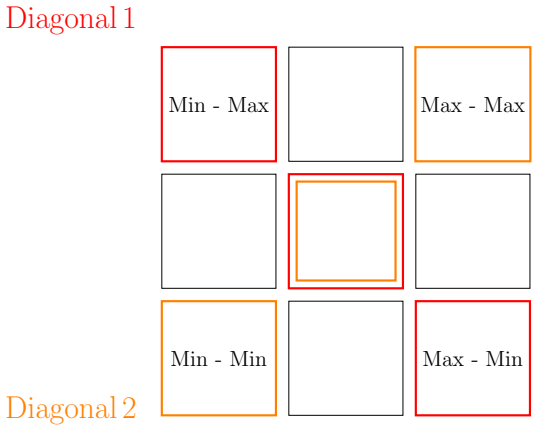



Figure 21: Representation of the different names of the entries of the matrices.

### 6.5.2 Comparison of the scenarios

The results of the 11 scenarios presented will be analyzed. Firstly, the performance of ComboFM (as a reminder ComboFM is used only to predict pairwise responses that are then used as input of the Dose model) will be detailed.
Then the performance of the triplet predictions of the Dose model will be detailed. Remember that

this performance has to be compared with the performance of the Dose model only and we do not expect better results than that (see TABLE 13). The goal is to identify which entries of the matrices are important in order to correctly predict the triplet responses.

The experiments on the 11 scenarios have been done 5 times, using 5 different random seeds in ComboFM. That allows to assess the stability of ComboFM, and its impact of the predictions made by the Dose model.

#### Results of ComboFM

**Cross-validation results.** The results of the cross-validation performed in ComboFM are summarized in TABLES 14 and 15 (see TABLES 27, 28, 29, 30, 31 in Appendix for the detailed results). As a reminder, the cross-validation is based on the Pearson correlation coefficients between the predicted values and the real ones. The chosen parameter is the one having the highest mean of Pearson coefficients among the 5 folds of the cross-validation. As each scenario has been ran with 5 different random seeds, the tables below present the means and standard deviation of the means of Pearson coefficients. Furthermore, the last column shows the standard deviations of those means across the different values of the parameters. The standard deviations have been normalized by N-1, N being the number of observations (5 for the 5 random seeds in this case, and 4 for the last columns). There are several interesting observations to make based on those tables:

- The scenario does not seem to highly impact the values of the Pearson coefficients. As a reminder, the set-up of the cross-validation matches the ones of the test set. ComboFM does not seem to be highly impacted by the chosen scenario.
- The random seed does not impact much the values either.
- Concerning the rank of factorization, the chosen value is always 100 (the highest one). It is not very surprising because intuitively the smaller the rank, the less complex (the more factorized) the model. If there is enough data to learn a more complex model (which seems to be the case), then there is no reason to further limit the complexity of the model than a rank set to 100 (which is already quite limited). However, the benefit of choosing a high value for this parameter seems to be low (*i.e.* the difference in the Pearson coefficient values is low) and the higher the value of the rank of factorization, the slower the learning of the model.
- The choice of the regression parameter is much more variable. The standard deviation across the different parameter values is even smaller than the one for the rank of factorization. None of the values really seem to stand out from the others.

Note that the means and standard deviations have been computed on the values by rounding to the 4th decimal place, which explains why we sometimes have equal values.
From these results, we conclude that several combinations of these two parameter values give very similar performances.

**Performance.** FIGURES 22 and 23 show a summary of the performance of ComboFM for the 11 scenarios, in terms of $R^2$ and $RMSE$ respectively. As a reminder, the range of values predicted by ComboFM is between -100 to 100. It has to be kept in mind while analyzing the $RMSE$. The performance were assessed only on the test set, meaning the entries of the dose-response matrices of interest that were kept apart. The means and standard deviations across the 5 different seeds have been computed, see FIGURES 35 and 36 in Appendix for detailed results. The scenarios have

| Scenario | Rank = 25 | Rank = 50 | Rank = 75 | Rank = 100 | Std |
|---|---|---|---|---|---|
| S1 | 0.9446 ± 2.86e-04 | 0.9572 ± 1.00e-04 | 0.9623 ± 1.30e-04 | **0.9651** ± 1.51e-04 | 0.0091 |
| S2 | 0.9423 ± 1.87e-4 | 0.9538 ± 5.47e-5 | 0.9585 ± 7.07e-5 | **0.9612** ± 1.58e-4 | 0.0083 |
| S3 | 0.9412 ± 1.48e-4 | 0.9535 ± 1.51e-4 | 0.9573 ± 2.28e-4 | **0.9692** ± 1.41e-4 | 0.0115 |
| S4 | 0.9438 ± 2.28e-4 | 0.9559 ± 1.09e-4 | 0.9608 ± 1.12e-4 | **0.9635** ± 5.47e-5 | 0.0087 |
| S5 | 0.9493 ± 2.88e-4 | 0.9598 ± 1.48e-4 | 0.9640 ± 3.53e-4 | **0.9667** ± 4.47e-5 | 0.0076 |
| S6 | 0.9334 ± 1.94e-4 | 0.9471 ± 2.07e-4 | 0.9529 ± 2.16e-4 | **0.9560** ± 1.30e-4 | 0.0100 |
| S7 | 0.9511 ± 2.73e-4 | 0.9625 ± 4.47e-5 | 0.9672 ± 1.93e-4 | **0.9697** ± 1.30e-4 | 0.0082 |
| S8 | 0.9389 ± 4.08e-4 | 0.9509 ± 2.58e-4 | 0.9559 ± 1.34e-4 | **0.9587** ± 1.09e-4 | 0.0087 |
| S9 | 0.9483 ± 1.18e-4 | 0.9577 ± 1.87e-4 | 0.9617 ± 2.58e-4 | **0.9641** ± 1.22e-4 | 0.0070 |
| S10 | 0.9402 ± 2.33e-4 | 0.9530 ± 2.94e-4 | 0.9581 ± 1.34e-4 | **0.9609** ± 1.30e-4 | 0.0093 |
| S11 | 0.9418 ± 1.48e-4 | 0.9547 ± 0.0022 | 0.9585 ± 2.25e-4 | **0.9615** ± 3.89e-4 | 0.0087 |

Table 14: Results of the cross-validation for the rank parameter. The means and standard deviations across the different running using different seeds are presented. The standard deviations across the different values of the parameters are also presented. The highest value for mean scenario is highlighted in bold font.

| Scenario | Reg = $10^2$ | Reg = $10^3$ | Reg = $10^4$ | Reg = $10^5$ | Std |
|---|---|---|---|---|---|
| S1 | 0.9561 ± 0.0024 | 0.9571 ± 1.09e-4 | **0.9572** ± 7.01e-5 | 0.9571 ± 1.00e-4 | 5.18e-4 |
| S2 | **0.9537** ± 8.36e-5 | **0.9537** ± 8.36e-5 | **0.9537** ± 8.36e-5 | 0.9536 ± 1.51e-4 | 5.0e-5 |
| S3 | **0.9525** ± 2.23e-4 | **0.9525** ± 1.78e-4 | **0.9525** ± 1.87e-4 | **0.952**5 ± 2.12e-4 | 0 |
| S4 | 0.9558 ± 1.64e-4 | **0.9559** ± 1.67e-4 | **0.9559** ± 1.34e-4 | 0.9557 ± 8.36e-5 | 9.57e-5 |
| S5 | **0.9598** ± 1.14e-4 | **0.9598** ± 1.78e-4 | **0.959**8 ± 1.64e-3 | 0.9597 ± 1.41e-4 | 5e-5 |
| S6 | 0.9444 ± 0.0061 | 0.9471 ± 2.30e-4 | 0.9501 ± 0.0046 | **0.9587** ± 0.004 | 0.0062 |
| S7 | **0.9625** ± 1.14e-4 | **0.9625** ± 8.36e-5 | **0.9625** ± 8.36e-5 | 0.9624 ± 4.47e-5 | 5e-5 |
| S8 | **0.9510** ± 5.47e-5 | 0.9509 ± 2.58e-4 | 0.9509 ± 1.30e-4 | 0.9509 ± 1.30e-4 | 5e-9 |
| S9 | **0.9577** ± 1.78e-4 | 0.9576 ± 1.30e-4 | 0.9576 ± 1.30e-4 | 0.9576 ± 1.78e-4 | 5e-5 |
| S10 | 0.9529 ± 2.47e-4 | **0.9530** ± 2.5e-4 | 0.9529 ± 3.03e-4 | 0.9529 ± 1.14-e4 | 5e-5 |
| S11 | 0.9517 ± 0.0045 | **0.9537** ± 1.30e-4 | 0.9536 ± 1.34e-4 | 0.9535 ± 1.41e-4 | 9.43e-4 |

Table 15: Results of the cross-validation for the regularization parameter. The means and standard deviations across the different running using different seeds are presented. The standard deviations across the different values of the parameters are also presented. The highest mean for each scenario is highlighted in bold font.

been ordered and grouped by color depending on the number of predicted entries in the matrices of interest.
Since the scores have been computed only on the predicted values, the bars of different colors can not really be compared to each other. It is however interesting to notice that the $RMSE$ varies even between bars of the same colour. This suggests that some entries of the dose-response matrices contain more information that other and/or are easier to predict (even if the total number of predicted value is the same).

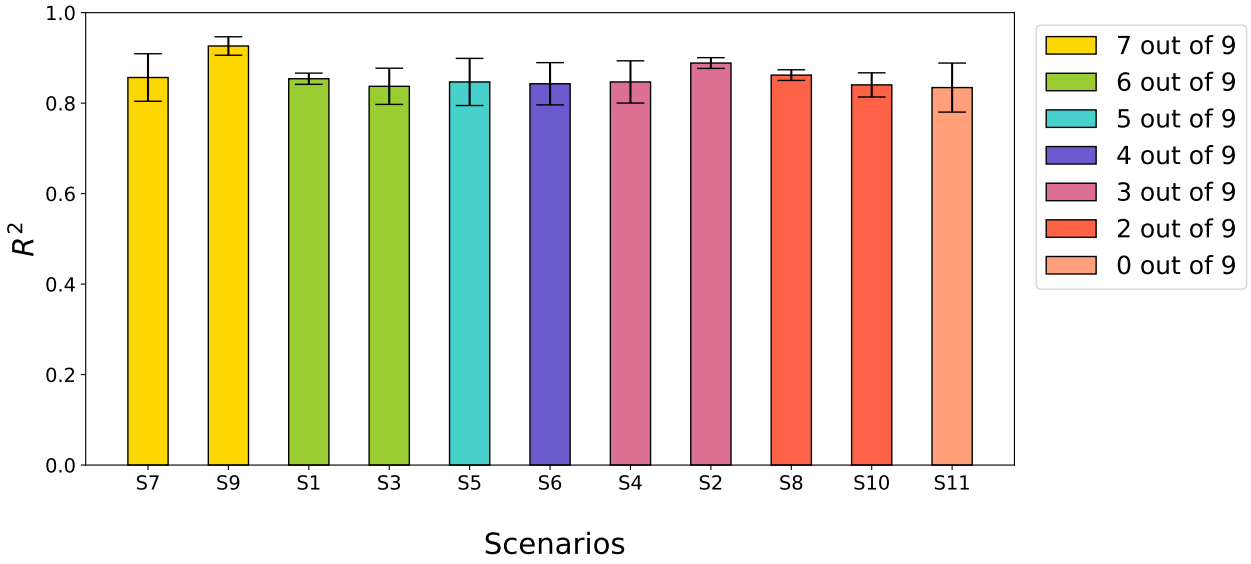



Figure 22: Performance of ComboFM in terms of $R^2$, for the 11 scenarios. The mean and standard deviation across the results corresponding to the 5 different seeds of ComboFM have been computed. Each color of bar corresponds to a number of entries in the training set.

There is another way to evaluate the performance of ComboFM. Indeed, we can compute the $R^2$ and $RMSE$ on the whole dose-response matrices of interest (*i.e.* $3 * 9 = 27$ data points, regardless of the size of the test set. For the entries that were not predicted, the real values were considered. Figures 24 and 25 show those new scores.
This second way of computing the score gives another information about the predictions. This allows to evaluate how good are the *entire* matrices. On can see that the more points predicted among the 27, the greater the $RMSE$ and the lower the $R^2$. This was totally expected and it seems very logical. Indeed, if many points of the matrix are predicted, we accumulate more error due to the predictions. In total, the entire matrix is further away from the original matrix.

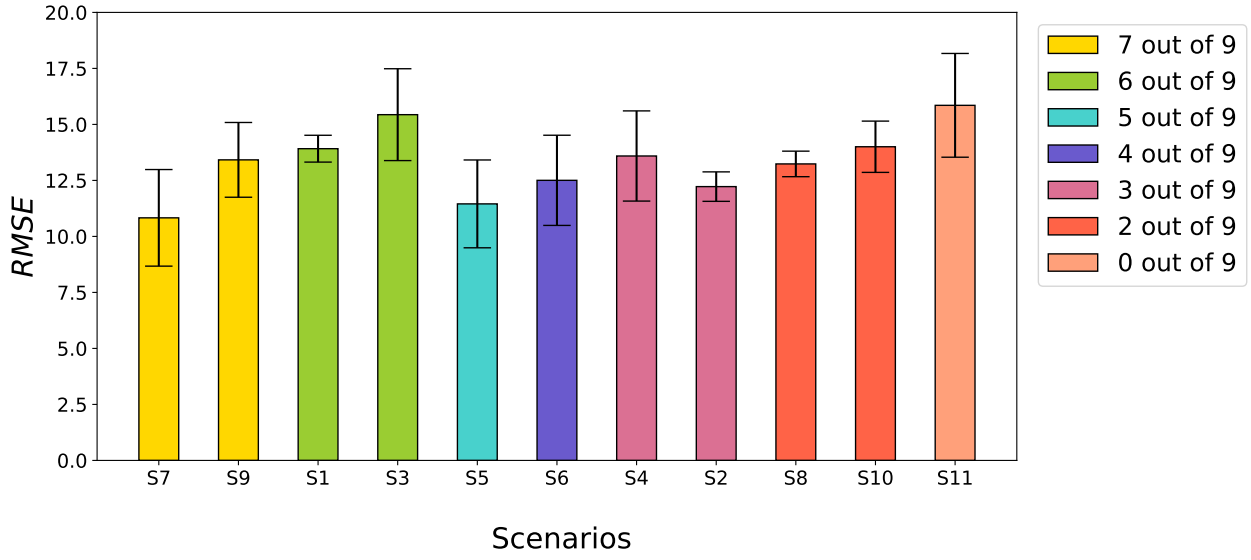



Figure 23: Performance of ComboFM in terms of $RMSE$, for the 11 scenarios. The mean and standard deviation across the results corresponding to the 5 different seeds of ComboFM have been computed. Each color of bar corresponds to a number of entries in the training set.

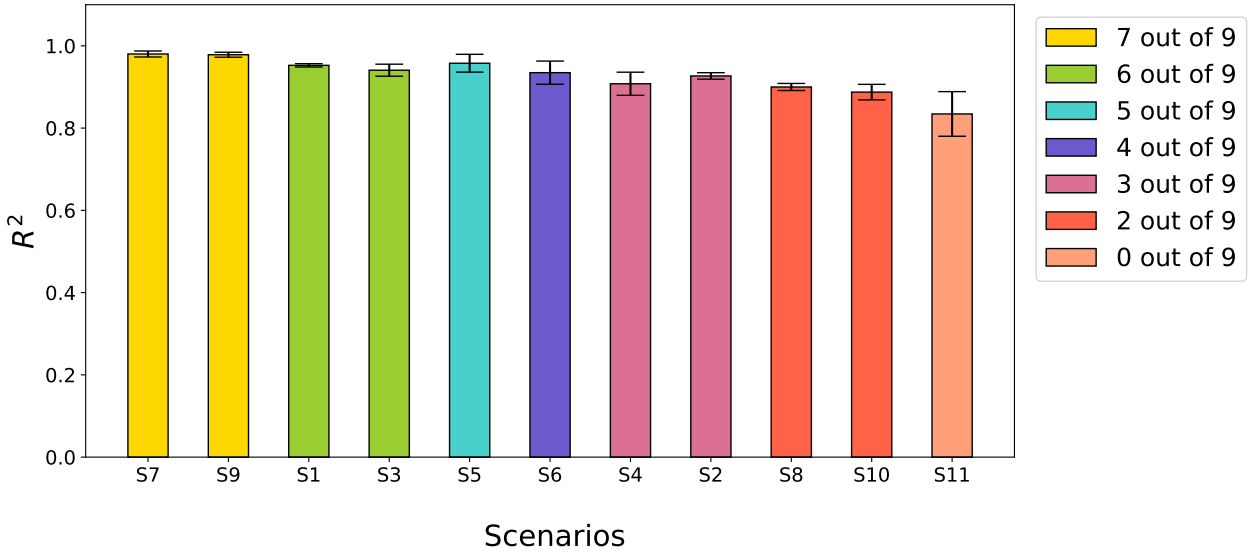



Figure 24: Performance of ComboFM in terms of $R^2$, for the 11 scenarios. The mean and standard deviation across the results corresponding to the 5 different seeds of ComboFM have been computed. Each color of bar corresponds to a number of entries in the training set. The score is computed on the entire dose-response matrices of interest (27 data points in total, regardless of the size of the test set).

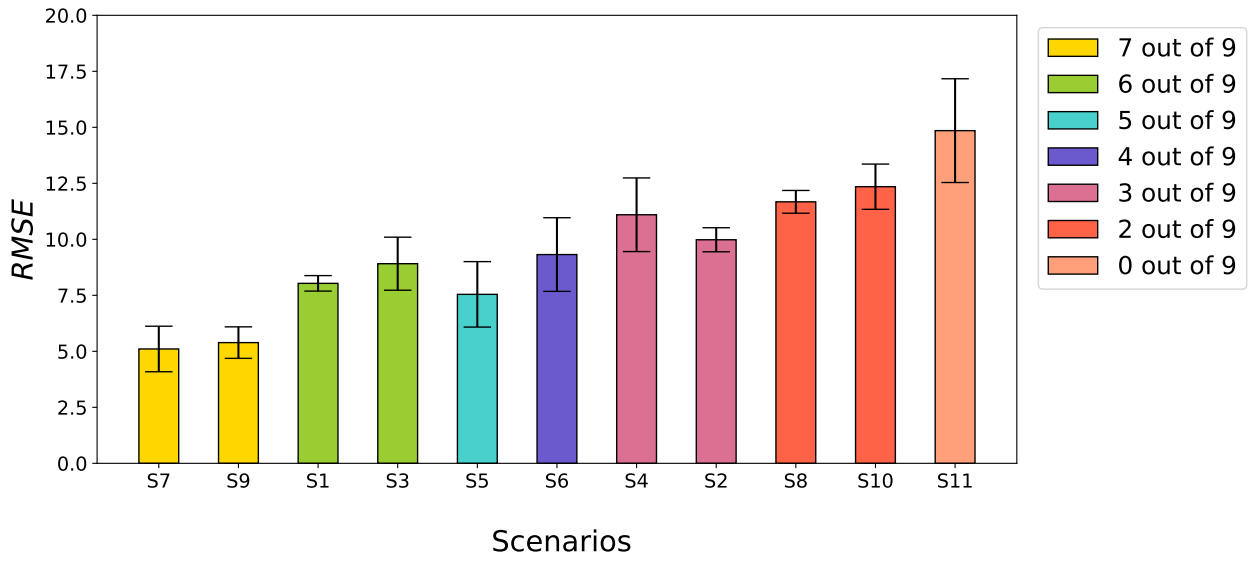



Figure 25: Performance of ComboFM in terms of $RMSE$, for the 11 scenarios. The mean and standard deviation across the results corresponding to the 5 different seeds of ComboFM have been computed. Each color of bar corresponds to a number of entries in the training set. The score is computed on the entire dose-response matrices of interest (27 data points in total, regardless of the size of the test set).

### Performance of the Dose model

Let us now look at the performance of the Dose model, in terms of $RMSE$ and $R^2$ as well. Those results are presented in FIGURES 26 and 27, in the same way as the performance of ComboFM was presented.
It appears obvious that the $R^2$ is much more sensitive to the scenario and the random seed.
The $RMSE$ is less sensitive than the $R^2$ but still more than in ComboFM alone. This suggests that a small change in the predictions of ComboFM, even if on average the predictions are equally good, has quite a big impact on the predictions of the Dose model.

Let us now compare the results between the different scenarios, in order to identify the most interesting one(s). The scenarios will also be compared two by two, each scenario with its mirroring one (except scenario 11).

- **Scenarios 1 and 2:** the first scenario gives better results than the second one. This seems very logical because of the number of entries in the test set. In the first scenario, only 3 entries are predicted while 6 entries are predicted in the second scenario. However, the second scenario gives satisfying results. Indeed, the $R^2$ and $RMSE$ are close to the ones obtained without using ComboFM at all.
  This means that Scenario 2 could be an interesting way to decrease the number of needed experiments.

- **Scenarios 3 and 4:** Scenario 3 gives better results than Scenario 4. This is easily explained in the same way as for scenarios 1 and 2. In Scenario 3, the same number of entries is predicted as in scenario 1. The gain in terms of the number of measurements that have to be made is thus the same in Scenarios 3 and 1. As a consequence, Scenario 1 is more interesting because

it gives better results.
Concerning Scenario 4, it is highly unstable, especially in terms of $R^2$. That means that the entries "Min-Min", "Max-Max" and the center one are tricky entries to rely on. Scenario 4 is thus a risky choice and one can not rely on it in order to predict higher order responses of drug combinations.

- **Scenarios 5 and 6:** both scenarios are quite stable regarding the different random seeds. Scenario 5 is slightly better than Scenario 6, which is once again explainable by the number of predicted entries. Scenario 6 is more interesting in terms of the needed number of measurements. This scenario seems to be a good choice in order to obtain satisfying results and measure only 4 of the 9 entries.

- **Scenarios 7 and 8:** Scenario 7 is very good. It is not surprising because only 2 out of the 9 entries of the matrices are predicted. It is interesting to note that Scenario 7 is a subset of Scenario 1. The only difference between them is the presence of the center entry of the matrices in the training set in Scenario 7. The similarity in performance between these 2 scenarios indicates that the center entry of the matrices is not the most important one when we want to predict the "Min-Min" and "Max-Max" entries. Therefore, Scenario 7 is not very interesting since the goal is to decrease the number of measurements to do.
  It is interesting to also compare Scenarios 2 and 8. They also differ by the presence of the center entry in the training set. The observation is not the same as the one made when comparing Scenarios 1 and 7. Indeed, Scenario 2 gives better results than Scenario 8. This means that having the center entry in the training set (in addition to the "Min-Min" and "Max-Max" entries) is important to predict the rest of the matrices. To choose between Scenarios 2 and 8, one has to decide what is more important between reducing the number of measurements and getting the best predictions.

- **Scenarios 9 and 10:** Scenario 9 gives satisfying results but, as Scenario 7, it is not very interesting because it only allows to reduce the number of needed measurements by 2.
  Scenario 10 is highly unstable. In the light of the performance of Scenario 4, this was expected. Indeed, Scenario 10 is a subset of Scenario 4. Once again, the only difference between them is the presence of the center entry in the training set. As Scenario 4, Scenario 10 seems to be a very risky choice.

- **Scenario 11:** As it was expected, this scenario is the one giving the less satisfying results. In this scenario, the predictions rely only on other dose-response matrices of the pair of drugs (*i.e* in other cell lines than the one of interest). It thus seems logical to see that this scenario is the most difficult one.
  The variance between the different seeds is quite high but it is not the worst scenario regarding the stability, if we compare with scenario 4 for example.

Finally, Figures 26 and 27 do not tell the whole story. Indeed, since only the mean is presented, one can not see that sometimes the Dose model does not manage to make its prediction. This means that, for some dose combinations of the triplet, the Dose model did not manage to find the effective dose and thus did not provide any prediction (the value is $NaN$ as there is no default value provided in the Dose model). However, ignoring those values and compute the scores without them may be over/under optimistic. As a consequence, the sets of predictions in which there is a $NaN$ value are ignored. This implies that the means and standard deviations are computed on the predictions among less than the 5 sets of predictions. Table 16 shows on how many sets of predictions the means and standard deviations have been computed. Note that in the two cases where one set of predictions have been ignored, only 1 of the 280 predictions was $NaN$.

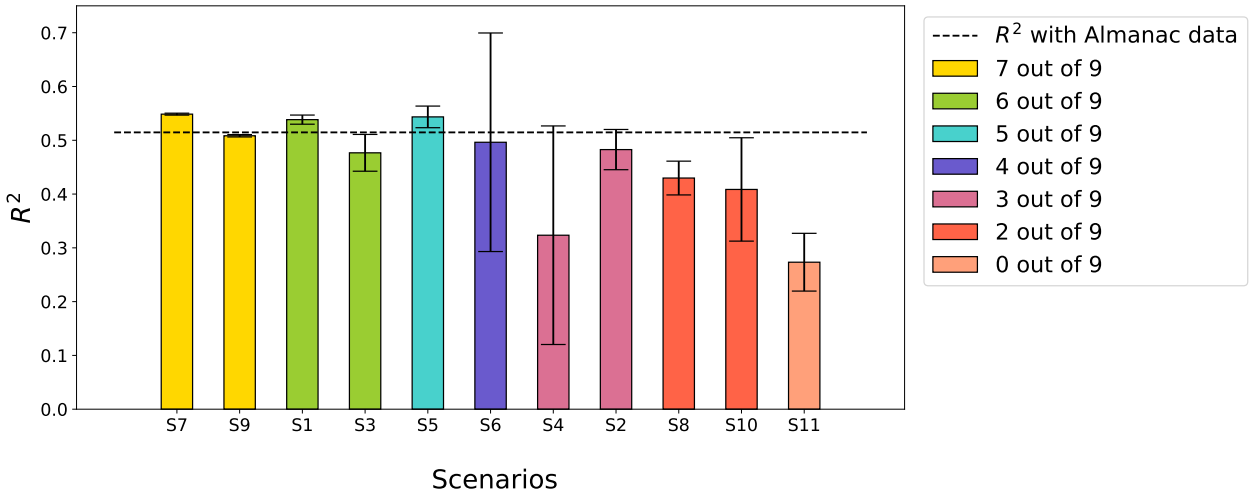



Figure 26: Performance of the Dose model (+ComboFM) in terms of $R^2$, for the 11 scenarios. The mean and standard deviation across the results corresponding to the 5 different seeds of ComboFM have been computed. A missing bar means that the Dose model did not converge.

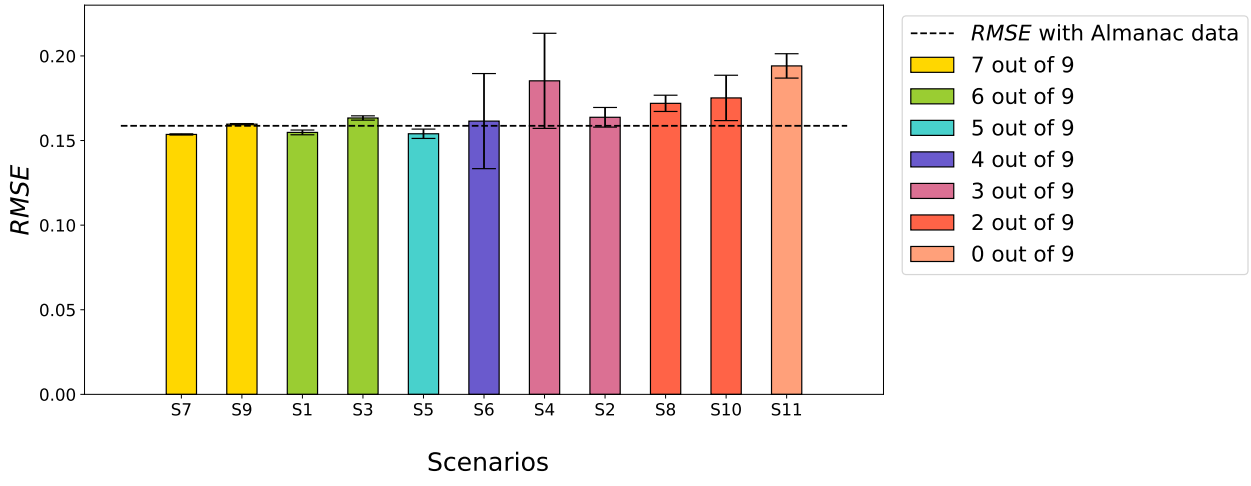



Figure 27: Performance of the Dose model (+ComboFM) in terms of $RMSE$, for the 11 scenarios. The mean and standard deviation across the results corresponding to the 5 different seeds of ComboFM have been computed.

It is nevertheless interesting to see if considering only 4 of the 5 sets of predictions has a big impact or not. In order to do so, the mean of $R^2$ has been computed for Scenarios 5 and 11 by removing the $NaN$ values and considering 279 triplets predictions for this set of predictions and 280 for the 4 others (see the results in FIGURE 39 in Appendix). The difference between the two ways of computing the means is not significant and, more importantly, it does not change the comparison between the different scenarios.

| | | | |
|---|---|---|---|
| S1 | 5 out of 5 | S7 | 5 out of 5 |
| S2 | 5 out of 5 | S8 | 5 out of 5 |
| S3 | 5 out of 5 | S9 | 5 out of 5 |
| S4 | 5 out of 5 | S10 | 5 out of 5 |
| S5 | **4** out of 5 | S11 | **4** out of 5 |
| S6 | 5 out of 5 | | |

Table 16: Summary of how many times the Dose model (+ComboFM) *completely* converged for each scenario.

### 6.5.3 Conclusion

Those experiments allow to highlight several interesting observations:

- One can, by being careful, use the predictions of ComboFM as an input to the Dose model in order to reduce the number of required measurements.
- Even if ComboFM gives good predictions in all the tested scenarios, this is not the case for the Dose model. This highlights the fact that the Dose model is quite sensitive to the variation of the pairwise responses.
- All the investigated scenarios are not equivalent. Even though this was expected, it is interesting to notice the importance of some entries of the dose-response matrices. Indeed, it seems like the entries "Min-Min" and "Max-Max" are more critical than the others. As a consequence, one has to carefully choose how one wants to combine ComboFM with the Dose model. In addition to the number of entries we want to predict with ComboFM, it is necessary to carefully choose which ones.

Note that there is not guarantee that those conclusions are valid for any combination of drugs of any order. We can conclude that, in any case, the chosen scenario has an impact (some of them give better results that others) but the best scenario may not be always the same. In order to define the best scenario in most cases, we would need much more validation data to run the experiment.

## 6.6 Experiments part 3: Sensitivity of the models

In this part, the sensitivity of the Dose model will be investigated. This evaluation of the sensitivity of the models is made on the model performance. Note that we mainly consider Scenario 11 (the entire matrices of interest are predicted by ComboFM) in what follows.

### 6.6.1 Gaussian noise

This experiment is done by adding a Gaussian noise on top of the predictions of ComboFM, in order to evaluate the impact of such variation on the predictions of the Dose model. The Gaussian noise

is characterized by a zero mean, while several values for the standard deviation have been tested. The different chosen standard deviations are: 1, 2, 3, 4, 5, 6, 10, 15 and 20. As a reminder, the values predicted by ComboFM are bounded between -100 and 100. Note that, in the first instance, the noise has been added to one of the predictions of ComboFM (as a reminder, it has been done 5 times, with 5 different random seeds), corresponding to the first random seed.

For each value of the standard deviation of the Gaussian noise, the noise has been added 50 times on the predictions of ComboFM. As a reminder, the entire matrices of interest have been predicted by ComboFM in Scenario 11. As a consequence, noise has been added on 27 input values of the Dose model. The noisy predictions have then been used in the Dose model (in addition to the mono-therapy data points that are left unchanged). In order to evaluate the impact of the noise, the standard deviation of the performance of the Dose model (in terms of $R^2$ and $RMSE$) has been computed. FIGURE 28 shows the results.
The standard deviations of the performance (both in terms of $R^2$ and $RMSE$) of the Dose model due to the 5 runs of ComboFM (the 5 different random seeds) are also displayed as dashed lines. This allows to read the results in the light of the variation of the Dose model due to the variability of ComboFM predictions.

Firstly, one can observe that the higher the standard deviation of the Gaussian noise, the higher the standard deviations of the scores. This was the expected result. When the input of the Dose model is more disturbed, the output is less stable. It is interesting to notice that the standard deviation of the performance when the standard deviation of the noise is set to 3 corresponds more or less to the actual variation of ComboFM (see the dashed lines). It allows to evaluate the cost in terms of performance of the Dose model due to the variation of ComboFM. This variation seems to be reasonable, meaning that the Dose model is relatively robust to random variations of ComboFM. Furthermore, those plots allow to quantify the margin of progression that can be made in order to stabilize the performance. The dashed lines show the fluctuation of the performance due to a change of the random seed in ComboFM. Therefore, one may want to optimize the other ComboFM parameters in order to reduce this fluctuation.
Another interesting point to notice is the fact that the $RMSE$ score seems to be much less impacted by the noise than the $R^2$ score. This is highlighted by FIGURE 28.

### 6.6.2 Stability of ComboFM

Based on the results of the previous experiment, the stability of ComboFM could be improved. There are several possibilities to try to do that. In this experiment, one possibility will be investigated. Once again, it is only done for the last scenario (number 11, corresponding to the entire matrices of interest in the test set). The predictions have been redone by setting the number of epochs in the training of ComboFM to 1000 (instead of 200). It has been done 5 times, for the 5 different random seeds. Note that the cross-validation in ComboFM has not been redone in this experiment, since the number of epochs is modified for the final training of the model. However, it could have been interesting to see if increasing the number of epochs in the cross-validation (which is set to 100 as a reminder) makes the results of the cross-validation more variable (in the sense that, as shown in the second part of the experiments, all the tested values of the two parameters gave nearly the same result).

FIGURE 29 shows the comparison of the performance of the Dose model while using ComboFM with 200 and 1000 epochs in the training step. The means and standard deviations across the results with the 5 different seeds of ComboFM are presented, see FIGURES 40 and 41 in Appendix for detailed

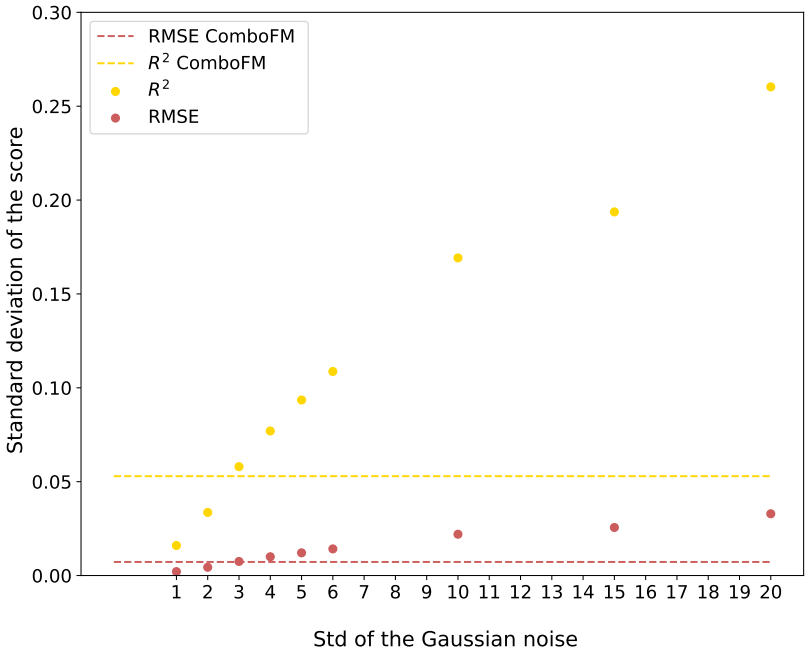
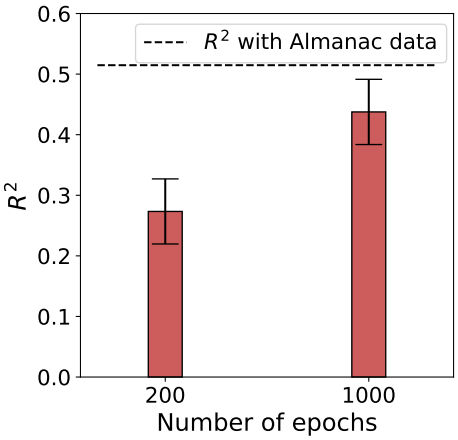



Figure 28: Evolution of the standard deviation of the performance of the Dose model (+ComboFM) when Gaussian noise is added to the predictions of ComboFM. The standard deviation of the $R^2$ is shown in yellow, while the standard deviation of the $RMSE$ is in red. The dashed lines represent the standard deviations of the performance of ComboFM due to the change of the random seed of ComboFM.

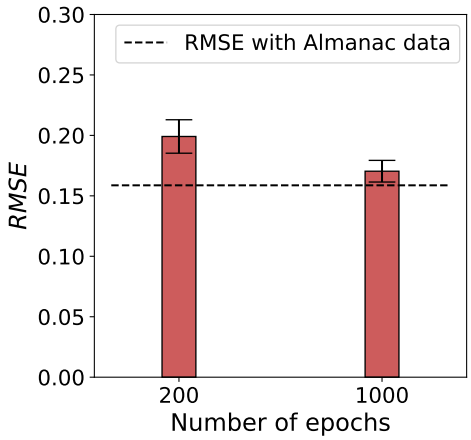
results. Firstly, we can clearly see that the number of epochs in ComboFM has a non-negligible impact on the predictions of the Dose model. The performance of the Dose model while using 1000 epochs in ComboFM is much closer to the performances without using ComboFM. Indeed, the mean of the $R^2$ between the 5 different seeds goes from 0.2733 to 0.42858. This means that Scenario 11 can give satisfying results, contrary to what it seemed in Section 6.5.2.

Concerning the stability of the predictions, we can firstly notice that the Dose model completely converged the 5 times when ComboFM was trained with 1000 epochs, while it converged 4 times with 200 epochs. The standard deviation of the $R^2$ between the 5 different seeds changed from 0.0537 to 0.0546. This change is not significant, which suggests that the stability of ComboFM is not really improved when the number of epochs is increased.

It is interesting to have a look at the predictions of ComboFM itself. Indeed, it will allow to understand where this change in the performance comes from. The statistics regarding the predictions of ComboFM are presented in Table 17. The two correlation coefficients are quite high, meaning that the two vectors of data are consistent with each other. However, the mean and standard deviation of the difference between the vectors of predictions are not negligible, and there are significant differences. This means that the predictions of ComboFM can be improved by increasing the number of epochs in the training step.
The improvement of the performance of the Dose model with the increase of the number of epochs in the training of ComboFM is quite important in this 11th scenario. However, it is important to remember that it was the most challenging scenario for ComboFM to predict. As a consequence, it may be the most impacted scenario by the number of epochs. In order to verify this hypothesis, the same experiment has been made on scenario number 8. As a reminder, scenario 8 predicts 7




(a) Comparison of the performance (in terms of $R^2$) of the Dose model combined with ComboFM (scenario 11), depending on the number of epochs used in the training of ComboFM.




(b) Comparison of the performance (in terms of $RMSE$) of the Dose model combined with ComboFM (scenario 11), depending on the number of epochs used in the training of ComboFM.

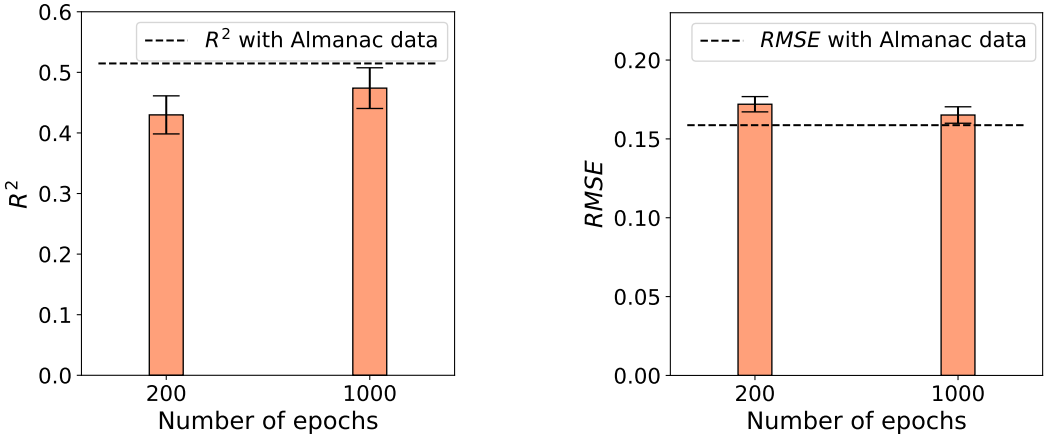
Figure 29: Comparison of the performance of the Dose (+ComboFM) model between 200 and 1000 epochs in training step of ComboFM, for Scenario 11.

| | Seed # 1 | Seed # 2 | Seed # 3 | Seed # 4 | Seed # 5 |
|---|---|---|---|---|---|
| Pearson | 0.9639 | 0.9806 | 0.9912 | 0.9816 | 0.9815 |
| Spearman | 0.9786 | 0.9749 | 0.9877 | 0.9780 | 0.9700 |
| Mean of difference | 10.49 | 6.498 | 5.988 | 6.631 | 6.284 |
| Std of difference | 9.799 | 4.113 | 3.8214 | 3.412 | 4.537 |

Table 17: Comparison of the predictions of ComboFM (scenario 11) based on the number of epochs in the training step: 200 epochs versus 1000 epochs.

out of the 9 entries of the dose-response matrices, the two entries being 'Min-Max' and 'Max-Min'. This scenario has been chosen because it is not the easiest one but it already gave satisfying results with 200 epochs. The results are presented on FIGURE 30 (as for Scenario 11, the means and standard deviations are presented, see FIGURES 42 and 43 in Appendix for the details) and TABLE 18.

One can see that increasing the number of epochs in the training of ComboFM also improves the predictions of the Dose model. However, the improvement is less than the one with Scenario 11. This observation was totally expected because the performance should not go beyond the performance obtained without using ComboFM. As a consequence, the margin of progress of Scenario 8 is lower than the margin of progress of Scenario 11.
Concerning the stability, there is not much change between the performance with 200 and 1000 epochs. This is the same observation as the one made concerning Scenario 11.




(a) Comparison of the performance (in terms of $R^2$) of the Dose model combined with ComboFM (scenario 8), depending on the number of epochs used in the training of ComboFM.

(b) Comparison of the performance (in terms of $RMSE$) of the Dose model combined with ComboFM (scenario 8), depending on the number of epochs used in the training of ComboFM.

Figure 30: Comparison of the performance of the Dose model (+ComboFM) between 200 and 1000 epochs in training step of ComboFM, for Scenario 8.

| | Seed # 1 | Seed # 2 | Seed # 3 | Seed # 4 | Seed # 5 |
|---|---|---|---|---|---|
| Pearson | 0.9845 | 0.9893 | 0.9869 | 0.9863 | 0.9889 |
| Spearman | 0.9870 | 0.9714 | 0.9844 | 0.9662 | 0.9649 |
| Mean of difference | 4.295 | 4.726 | 4.814 | 5.777 | 5.4921 |
| Std of difference | 5.588 | 3.354 | 3.666 | 3.5714 | 4.1869 |

Table 18: Comparison of the predictions of ComboFM (scenario 8) based on the number of epochs in the training step: 200 epochs versus 1000 epochs.

Based on those results, it is interesting to investigate the impact of the number of epochs a little bit further. This has been done by doing the same experiments with 500 and 1500 epochs in the training of ComboFM. The idea is to check if 1000 is the optimal value or if a lower value gives already the best results. On the contrary, it is also possible that 1000 epochs is still not enough. FIGURES 31 and 32 show the evolution of the performance (in terms of $R^2$ and $RMSE$) of the Dose model in function of the number of epochs used in the training of ComboFM.

The first observation is that Scenario 11 is more impacted by the number of epochs than Scenario 8. This was expected based on the results with 200 and 1000 epochs. That highlights the fact that more difficult scenarios need more epochs in order to give satisfying results. Scenario 8 is not really impacted by the number of epochs while Scenario 11 improves when the number of epochs increases.

A second interesting point is the standard deviation of both $R^2$ and $RMSE$ of Scenario 11 with 500 epochs. This standard deviation is very high, which is surprising. This very high value is quite easily explainable. Indeed, with 200 epochs, the predictions of ComboFM made with the second random seed did not allow the Dose model to completely converge. However, with 500 epochs the predictions of ComboFM made with the second random seed allowed the Dose model to completely converge but it gave very bad results: $R^2 = 0.1681$ and $RMSE = 0.2078$. This alone makes the standard deviation across the 5 seeds very high. If we do not take those values into account, the standard deviation goes from 0.113 to 0.041 for $R^2$ and from 0.0149 to 0.006 for $RMSE$. This suggests that increasing the number of epochs allows to be less dependent on the chosen random seed.

This last observation is confirmed by the evolution of the standard deviation of Scenario 11. It is decreasing when the number of epochs increases, which means that the predictions are less dependent on the random seed. FIGURE 33 allows to evaluate the gain in terms of standard deviation when the number of epochs is set to 1500, for Scenario 11. As is was observed earlier, the standard deviation of the performance coming from the change of the random seed in ComboFM when we have 200 epochs, corresponds more or less to the standard deviation of the performance when Gaussian noise with a standard deviation equals to 3 is added to the predictions of ComboFM. If we now consider the predictions of ComboFM when we have 1500 epochs in the training set, the standard deviation corresponds to the one when a Gaussian noise having a standard deviation equals to 2 is added.

Note that in 29 and FIGURES 30 it seemed that the standard deviations of the $R^2$ and $RMSE$ were not impacted by the change of epochs. The same observation can be done based on FIGURES 31 and 32 if we compare the values corresponding to 200 and 1000 epochs. However, the trend is downward if we look at the values with 200, 500, 1000 and 1500 epochs. It highlights the importance to make several experiments in order to have a global view.

Those results show that ComboFM should be modified/optimized in order to be reliable in this context. Depending on the considered scenario, one must choose carefully the parameters of ComboFM. Indeed, it is important to keep in mind that increasing the number of epochs has a non negligible drawback: it slows down the training step. As a consequence, one must find a good trade-off between the speed and the performance. For Scenario 8 for example, the balance is not really in favor of the increase of the number of epochs: it is much slower and the gain in performance is not significant. On the contrary for Scenario 11, the gain in performance is much bigger and it is thus much more interesting to increase the number of epochs in this scenario.

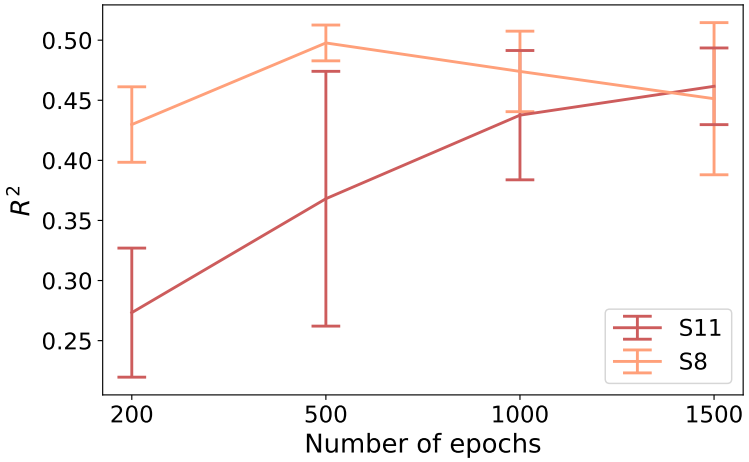



Figure 31: Evolution of the mean and the standard deviation of the $R^2$ (Dose model) across the 5 random seeds when the number of epochs in the training step of ComboFM increases.

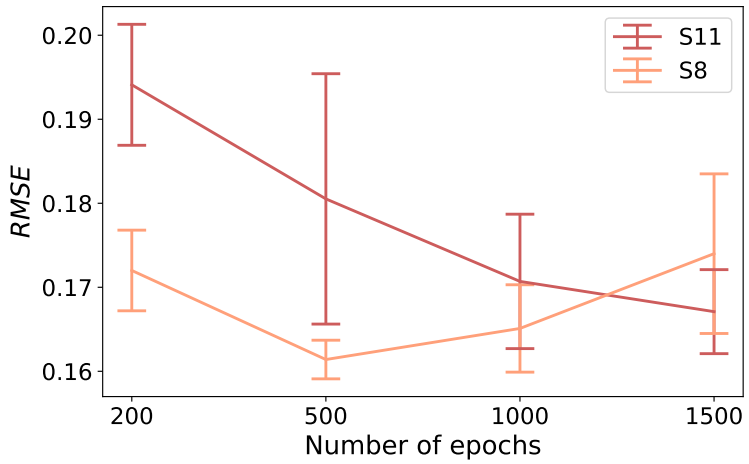



Figure 32: Evolution of the mean and the standard deviation of the $RMSE$ of the Dose model (+ComboFM) across the 5 random seeds when the number of epochs in the training step of ComboFM increases.

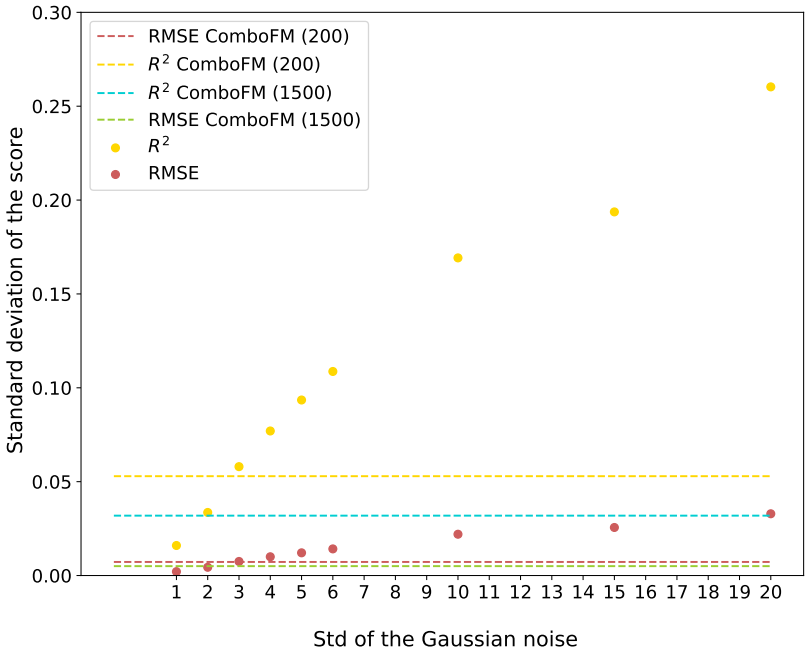



Figure 33: Evolution of the standard deviation of the performance of the Dose model (+ComboFM) when Gaussian noise is added to the predictions of ComboFM. In addition, the dashed lines represent the standard deviations of the performance of the Dose model only due to a change of the random seed in ComboFM. These results are for Scenario 11.

### 6.6.3 Conclusion

Those experiments allowed to observe different points:

- The Dose model is quite sensitive to variations of the predictions of ComboFM. The standard deviation in $R^2$ becomes quickly quite high. Keeping in mind that $R^2$ is bounded between -1 and 1, and predictions of ComboFM are bounded between -100 and 100, one can evaluate the acceptable variability of the predictions of ComboFM.
- The variability of the Dose model due to small changes in the predictions of ComboFM is not very convenient, especially when it has been shown that the predictions of ComboFM are not very stable (Section 6.5.2). As a consequence, one has to stabilize ComboFM in order to make the predictions of the combined model (ComboFM and the Dose model) reliable.
- The number of epochs used in the training of ComboFM has its importance, even if the predictions of ComboFM itself are already good with 200 epochs. However, the importance of this parameter depends on the considered scenario.

## 6.7 Experiments part 4 : Synergy scores

As it has already been mentioned, the final goal of making predictions of responses of drug combinations is to identify synergistic (and thus highly interesting) combinations. This fourth part of the experiments aims at answering the following question: What is the impact of using ComboFM in combination with the Dose model on synergy predictions? Different synergy scores (as explained in Section 2.1.5) will be computed and compared. This will be done on the predictions using only the NCI-ALMANAC dataset and on the predictions using ComboFM. Note that the ComboScore has been designed for the data of the NCI-ALMANAC dataset. Indeed, the formula of the ComboScore

takes into account negative values but the Dose model predicts only positive values. If we only have positive values, the ComboScore is equivalent to the sum of the Bliss scores.
It is interesting to evaluate the robustness of the dose-response predictions regarding the synergy predictions. If a relatively high change in the dose-response prediction does not induce a change in the synergy prediction, making an error on the dose-response prediction is not actually very critical.

Note that in view of the formulas to compute the different synergy scores, one can only make this experiment on the doses available in the NCI-ALMANAC dataset. The process is thus different from the one when using the Dose model. As a reminder, all the score formulas required to have the corresponding dose-responses of each pair in the combination. Let us illustrate this as follows: let us consider a triplet A-B-C for which we have the measured responses at the dose $D_A$ for drug A, $D_B$ for drug B and $D_C$ for drug C. The score formulas presented require to have the pairwise responses at the corresponding doses, meaning $(D_AD_B)$, $(D_BD_C)$ and $(D_AD_C)$.
This is challenging because the doses at which we have the measured responses of the triplet do not correspond to the doses at which we have the pairwise responses. As a consequence, the synergy scores of the triplet have been computed on all the combinations of the available in the NCI-ALMANAC doses of the three drugs. Those doses thus correspond to the ones of the three pairs of interest (see TABLE 10 for a reminder). This gives $3 * 3 * 3 = 27$ data points for the triplet.

Another challenge raised by the use of those other formulas is the mono-therapy data. As it has been explained in the general set-up of the experiments (see SECTION 6 for a reminder), there are several mono-therapy responses for the same dose of a drug in the NCI-ALMANAC dataset. In the Dose model, this is not a real problem since those mono-therapy data points are not used as such. They are 'only' used to fit a Hill function. As a consequence, using several points of response for a single dose is not a real problem and we let the Dose model fit those points at best (see FIGURE 34 for an example).
However, this is not valid for the use of the other formulas. Indeed, the mono-therapy data points appear explicitly in the formulas. It is thus not possible to have more than one point. The chosen solution is to take the mean of those different mono-therapy data points, when it is needed.

The synergy scores will be compared at two levels: the pairwise combinations and the triplets. It is interesting to see if those predictions are consistent. For example, it would be really surprising to obtain a synergistic triplets of drugs while the three pairs of drugs are antagonistic. Furthermore, the impact of using ComboFM previously to the Dose model will also be investigated, as well as the impact of noise.
Note that the formulas of synergy scores require data to be bounded between 0 and 1. Thus the percentage without time zero has been used and the data has been truncated by 1. It means that each time a value was higher than 1, it has been replaced by 1.

### 6.7.1 Synergy prediction without using ComboFM

In this first part of the experiment, only data from the NCI-ALMANAC dataset will be considered (meaning ComboFM will not be used to predict pairwise responses). For each combination of doses, there is one synergy score. Here, a summary of the results is presented, see FIGURES 44, 45, 46 and TABLE 32 in Appendix for the details. HSA and Bliss scores are compared in TABLE 19. The table presents, for each dose-response matrix of interest (there are three of them in total) and for the triplet, the number of entries that are predicted as synergistic (meaning that the entry has a score higher than 0). The goal here is not to obtain high scores but to compare the two synergy scores and, more importantly, to assess the impact of using ComboFM on those synergy scores.

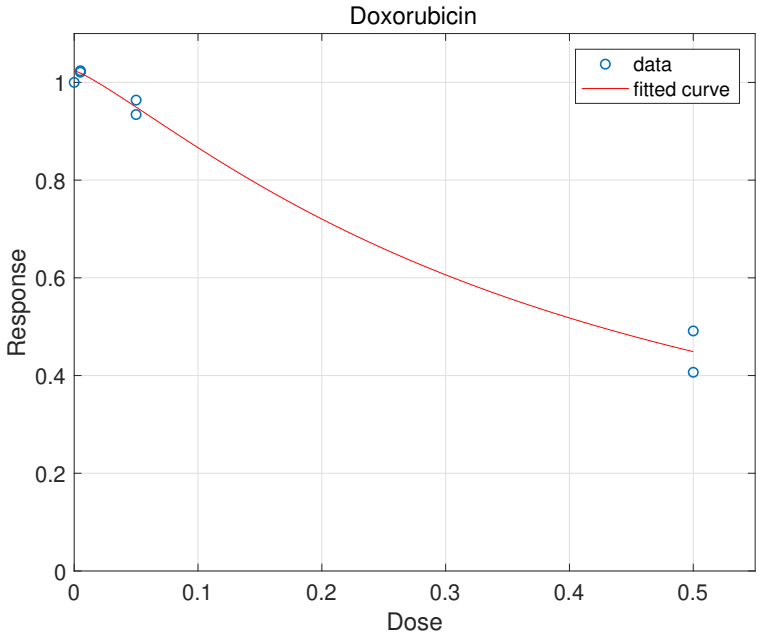



Figure 34: Example of the fitted Hill function in the Dose model where there are several responses for a given dose.

One can observe that the two scores generally agree concerning the pairs of drugs but disagree regarding the triplet. However, no conclusion can be drawn based only on the number of predicted synergistic entries. Indeed, if we have a look at the mean score, they are both negative and not so far from each other. This means that when the synergy score is higher than 0 (and thus the entry is considered as synergistic), the score is very low and so the entry is just barely synergistic. This highlights the importance to take a margin on those scores to predict synergism (*i.e.* we consider an entry to be synergistic only if the score is higher than 0.1 for example). Based on that, the two scores are not really in total disagreement.

### 6.7.2 Synergy prediction using ComboFM

As it has been explained, the purpose is to evaluate the impact of using ComboFM on synergy predictions. The exact same experiment has thus been made but using the predictions of ComboFM. Once again, only Scenario 11 has been considered here. Particularly, the predictions made using the first random seed have been considered. The summarized results are presented on Table 20 (see Figures 47, 48, 49 and Table 33 in Appendix for the details). There is not much difference between Tables 19 and 20, except for the triplet. Once again, looking only at the number of entries that are predicted as synergistic is not enough. The mean of the scores is much more interesting. For the Bliss score, we go from 0 to 15 synergistic entries while the mean of the score stays negative. This suggests that, once again, the entries that are predicted as synergistic have a score barely above 0, which is not significant.
Concerning the HSA score, the mean score becomes positive but the value is quite low. This once again suggests that the disagreement between the score with and without using ComboFM is not very important.
Furthermore, the HSA score is always the most optimistic (*i.e.* the values are higher). This is quite logical in view of the formulas (see Section 2.1.5).
Finally, we can have a look at the ComboScore. As a reminder, it does correspond to the sum of

| Pair Doxorubicin-Paclitaxel | | | |
|---|---|---|---|
| | Bliss | HSA | ComboScore |
| Synergy prediction | 2 out of 9 | 3 out of 9 | -1.3919 |
| Mean score | -0.154 | -0.1128 | |
| Pair Doxorubicin-Cisplatin | | | |
| | Bliss | HSA | ComboScore |
| Synergy prediction | 5 out of 9 | 6 out of 9 | 0.0136 |
| Mean score | 1.46e-3 | 0.036 | |
| Pair Cisplatin-Paclitaxel | | | |
| | Bliss | HSA | ComboScore |
| Synergy prediction | 7 out of 9 | 9 out of 9 | 0.09467 |
| Mean score | 0.105 | 0.1404 | |
| Triplet | | | |
| | Bliss | HSA | ComboScore |
| Synergy prediction | 0 out of 27 | 15 out of 27 | -2.9341 |
| Mean score | -0.107 | -0.026 | |

Table 19: Comparison of the Bliss, HSA and ComboScore synergy scores on the NCI-ALMANAC dataset, for the three dose-response matrices of interest and the corresponding triplet.

the values of the Bliss score across the pairwise dose-response matrices. It is thus logical to observe that if the mean of the Bliss score is positive, the ComboScore is also positive and vice versa. If we firstly focus on the ComboScore computed on the NCI-ALMANAC dataset, we have an antagonist triplet made from two barely synergistic pairs and one antagonist pair. However, the synergistic pairs have low scores (close to 0, especially for doxorubicin-cisplatin) and thus it does not seem inconsistent to have an antagonist triplet. However, it is a little bit more surprising concerning the score computed by using the pairwise dose-response matrices predicted by ComboFM. Indeed, from two more strongly synergistic pairs and one antagonist pair, we have an antagonist triplet. However the ComboScore is a sum over the values of the matrices, the score is not bounded by 1 (as it is the case for the Bliss and HSA scores), it is more difficult to interpret.

| Pair Doxorubicin-Paclitaxel | | | |
|---|---|---|---|
| | Bliss | HSA | ComboScore |
| Synergy prediction | 0 out of 9 | 3 out of 9 | -0.6002 |
| Mean score | -0.0666 | -0.0265 | |
| Pair Doxorubicin-Cisplatin | | | |
| | Bliss | HSA | ComboScore |
| Synergy prediction | 8 out of 9 | 8 out of 9 | 0.5371 |
| Mean score | 0.0596 | 0.0942 | |
| Pair Cisplatin-Paclitaxel | | | |
| | Bliss | HSA | ComboScore |
| Synergy prediction | 9 out of 9 | 9 out of 9 | 1.9988 |
| Mean score | 0.222 | 0.2577 | |
| Triplet | | | |
| | Bliss | HSA | ComboScore |
| Synergy prediction | 13 out of 27 | 20 out of 27 | -0.10465 |
| Mean score | -0.0038 | 0.0616 | |

Table 20: Comparison of the Bliss, HSA and ComboScore synergy scores on the predictions made by ComboFM and the Dose model (+ComboFM), for the three dose-response matrices of interest and the corresponding triplet.

### 6.7.3 Conclusion

This small experiment highlights the fact that even if the triplet predictions made using ComboFM *and* the Dose model are not perfect, they still can be reliable if the final goal is to find new synergistic combinations of drugs. However, this experiment has only been done on a single triplet of drugs and thus no general conclusion can be drawn regarding even higher order combinations.
Note that this experiment can be done on any combination of any drugs. The only condition is that all the dose-response matrices of all the pairs of drugs present in the combination should be available in the dataset.

## 6.8 Experiments part 5: Comparison with other formulas

This part of the experiments will compare the Dose model to the other analytical formulas presented in Section 2.2.1, in order to address the following question: How does the Dose model compare to others formulas? This experiment was only performed for Scenario 11.

Note that we face the same challenge as for the comparison of synergy scores. Indeed, all the other formulas required the mono-therapy and pairwise responses at the dose at which the considered drug is present in the triplet. As a consequence, the validation data is not useful in this experiment. The computation of the responses of the triplet using other formulas is thus made on the same doses as the previous experiment. As a reminder, all the combinations of the doses presented in Table 10 are considered. This gives $3 * 3 * 3 = 27$ data points for the triplet.
Furthermore, the same solution regarding the mono-therapy data points has been adopted (the mean has been taken when it is necessary). As for the last experiment, this can be done on any combination we want.

Table 21 shows the comparison between the predictions using the Dose model and the other formulas, using only the NCI-ALMANAC data (ComboFM has not been used). Table 22 shows the comparison between the prediction using ComboFM combined with the Dose model, and the other formulas. The mean of the comparisons across the 5 different predictions of ComboFM (with the 5 different random seeds) has been considered. See Tables 34, 35, 36, 37, 38 and 39 for the detailed results.

Firstly, one can observe that the closest formula to the Dose model is the Bliss formula. This is not surprising since both models are based on the same formula. The only difference remains in the integration of effective doses in the Dose model. Having Bliss predictions close to the Dose model predictions means that the effective doses are not very far from the real doses. Note that this does not mean that the performance (regarding validation) of the Bliss model is as good as the one of the Dose model. To illustrate this, let us have a look at the comparison between the Bliss model and Dose model predictions on the doses of the triplet for which we have the validation data. This comparison is shown in Tables 23 and 24. This shows that even if the predictions of the Bliss model are highly correlated to the predictions of the Dose model, the performance are very different. This suggests that the difference made by the Dose model has a positive impact.
It is also interesting to notice that using ComboFM decreases the performance *while increasing* the correlation between the Dose and Bliss predictions.
This confirms that the Dose model is more appropriate than the Bliss model.

Secondly, the regression formula seems to be the furthest from the Dose model. If we look at the formulas (Table 2), we can see that the Regression model is the only formula containing a quotient of responses, while the others only consider products/additions.

Let us now have a look at the impact of using ComboFM on those predictions:

- On the Pair model: using ComboFM makes the predictions of the pair model closer to the predictions of the Dose model
- On the Iserliss formula: using ComboFM makes the predictions of the Iserliss formula further from the predictions of the Dose model if we look at the $R^2$, although more correlated.
- On the Regression formula: using ComboFM makes the predictions closer to the ones of the Dose model.

| | Pair model | Iserliss | Regression | Bliss |
|---|---|---|---|---|
| Pearson | 0.8087 | 0.8512 | 0.6892 | 0.9106 |
| Spearman | 0.8602 | 0.8321 | 0.6398 | 0.9342 |
| RMSE | 0.2191 | 0.1790 | 0.2664 | 0.1626 |
| MAE | 0.1391 | 0.1323 | 0.1739 | 0.1085 |
| $R^2$ | 0.4089 | 0.6025 | 0.1816 | 0.6764 |

Table 21: Comparison between the predictions of the other formulas and the predictions of the Dose model. In this case, only the data from the NCI-ALMANAC dataset has been used (ComboFM has not been used).

| | Pair model | Iserliss | Regression | Bliss |
|---|---|---|---|---|
| Pearson | 0.9106 | 0.8797 | 0.8491 | 0.9670 |
| Spearman | 0.9100 | 0.8750 | 0.8863 | 0.9738 |
| RMSE | 0.1628 | 0.1972 | 0.2264 | 0.0786 |
| MAE | 0.1136 | 0.1498 | 0.1698 | 0.0467 |
| $R^2$ | 0.6586 | 0.3960 | 0.3202 | 0.9318 |

Table 22: Comparison between the predictions of the other formulas and the predictions of the Dose model (+ComboFM). In this case, the means of the different quantities across the 5 predictions of ComboFM (the 5 different random seeds) have been considered.

| Performance of the models | | |
|---|---|---|
| | Dose model | Bliss model |
| $R^2$ | 0.5147 | -0.1183 |
| $RMSE$ | 0.1587 | 0.2409 |
| Comparison of the predictions | | |
| $R^2$ | 0.3353 | |
| $RMSE$ | 0.2366 | |
| Spearman | 0.9004 | |
| Pearson | 0.8390 | |

Table 23: Comparison of the Bliss and Dose model on the validation data. The Dose model uses only the NCI-ALMANAC dataset (no use of ComboFM).

| Performance of the models | | |
|---|---|---|
| | Dose model | Bliss model |
| $R^2$ | 0.2257 | -0.1182 |
| $RMSE$ | 0.2005 | 0.2409 |
| Comparison of the predictions | | |
| $R^2$ | 0.8913 | |
| $RMSE$ | 0.0956 | |
| Spearman | 0.9574 | |
| Pearson | 0.9600 | |

Table 24: Comparison of the Bliss and Dose model. The Dose model uses the prediction of ComboFM, following scenario number 11.

### 6.8.1 Conclusion

It is very difficult to draw any conclusion based on those small experiments. Indeed, without having any validation data to assess the performance of the other formulas, one can not really conclude anything. It is clear that even if the predictions are not so far from each other, this does not mean that the performance are close. However, the Dose model seems appropriate when compared with the Bliss model.

## 6.9 Experiments part 6: Increase of the order of the combination

The last two parts of experiments have been made without using measured values as validation data. The analysis of the results have been made by comparing the predictions made using only the Dose model (on the NCI-ALMANAC dataset) with the predictions made using ComboFM prior to the use of the Dose model. This implies that similar experiments could be done for even higher order combinations of drugs. This will once again be done on Scenario 11 of ComboFM (the whole dose-response matrices are predicted). The goal is to evaluate the consistency of the Dose model when the order of the combinations is increased. Additionally to Scenario 11, the experiments are also conducted on Scenario 2. Scenario 2 has been chosen because the results of the triplet predictions using ComboFM were very close to the results of the triplet predictions using only the NCI-ALMANAC data, while having 6 out 9 entries predicted by ComboFM. This means that, for a triplet, Scenario 2 is suitable to the combination of ComboFM with the Dose model. As a consequence, it is interesting to investigate the same scenario for higher order combinations of drugs. Indeed, even if the combination of ComboFM and the Dose model seems to give satisfying results for the predictions of a triplet of drugs, one can not simply extend the affirmation to any order of cancer drug combinations.

Note that those two experiments do *not* allow to evaluate the validity of the Dose model for combinations of order 4 and 5. One would need measured data to compare with. However, those experiments allow to evaluate the validity of the combination of ComboFM and the Dose model compared to the single use of the Dose model.

### 6.9.1 Quadruplet of drugs

For the sake of simplicity of the experiments, the same three drugs and cell line as before are considered. One drug is added to the triplet (which is Doxorubicin, Paclitaxel and Cisplatin as a reminder). The fourth drug must be chosen such as the pairwise responses with all the already chosen three drugs in the chosen cell line (A549) are available in the NCI-ALMANAC dataset. There are 85 different drugs that fulfill this condition. In order to once again facilitate further experiments, two of those drugs will be chosen such that the pairwise responses between them is also available. Doing that allows to easily make the same experiment on a quintuplet of drugs. The last point that influenced the choice of the drugs is the screener who made the measurements in the NCI-ALMANAC dataset. In order to ensure consistency, the drugs have been chosen such as all the pairwise matrices of interest have been measured by the same screener (in this case SRI-international named *FF* in the dataset). The two arbitrarily chosen drugs are Dasatinib and Ifosfamide.

The considered quadruplet is thus: Doxorubicin, Paclitaxel, Cisplatin and Dasatinib.
In order to predict the dose-responses of this quadruplet, one thus needs the mono-therapy data points of the four drugs, as well as the 6 dose-responses matrices (there are 6 different pairs of drugs in a 4-order combination), one for each pair of drugs among the four drugs.
The mono-therapy data-points have been selected in the same way as for the triplets of drugs, by considering the study in which the dose-responses matrices were evaluated. All the pairwise matrices

of interest have not been measured in the same study.

The quadruplets of drug has been evaluated in all the combinations of doses of the pairwise dose-response matrices. In total, we thus have $3 * 3 * 3 * 3 = 81$ dose combinations and thus the same number of predictions. In order to ensure interpretability, the experiments have been made 5 times, using the 5 predictions of ComboFM obtained with the different random seeds.

Table 25 compares the predictions made on the triplets and quadruplets, by computing different metrics comparing the predictions made with and without using the predictions of ComboFM as an input in the Dose model. The values are the means across the results with the 5 different random seeds of ComboFM. See Tables 40, 41, 42, 43 for the details.
The first important thing to notice is the percentage of convergence. As it has been explained, it happens sometimes that the Dose model does not manage to predict some of the combinations of doses. As a consequence, the metrics are computed on the predictions, meaning that if the percentage of convergence is not 100%, the metrics are not computed on all the dose combinations. This can introduce a bias in the metrics.
Keeping that in mind, one can analyze the results. Firstly, the impact of going from a triplet to a quadruplet is always bigger in Scenario 11 than in Scenario 2, excecpt for the $RMSE$. This was expected because it has been shown that Scenario 11 is more difficult than Scenario 2.
The impact of using ComboFM previously to the Dose model on the predictions in terms of correlations and errors seems quite limited when we go from a triplet to a quadruplet.
That suggests that, if the Dose model is efficient for quadruplets of cancer drugs, it should be nearly as efficient if we previously use ComboFM. One has to keep in mind that we do not know if the Dose model is actually effective (even using only the NCI-ALMANAC dataset) for combinations of order higher than 3, since we do not have access to any measurements.
Therefore, we have no possibility to verify this intuition. Indeed, even though the predictions are strongly correlated, this does not imply that the performance ($R^2$ and $RMSE$ on validation data) will be equally good. Let us recall that strongly correlated variables could lead to large deviation in terms of $R^2$. For example, on the validation data despite a strong correlation between predictions of the Dose model only and the combination ComboFM/Dose model, the $R^2$ changes from 0.5147 to 0.2952 for Scenario 11. Concerning Scenario 2, it decreases to 0.4826. This highlights the fact that the predictions have to be very close to each other in order to give equally good results. There is thus a risk that, for Scenario 11 at least, the predictions on the quadruplets are not reliable enough.

### 6.9.2 Quintuplet of drugs

The experimental set-up is the same as the one for the quadruplet, except that one more drug is added: Ifosfamide.

As for the quadruplet, the predictions have been made on the combinations of doses of the pairwise dose-response matrices (we have 10 pairwise dose-response matrices in case of a quintuplet). In total, we thus have $3*3*3*3*3 = 243$ dose combinations and thus the same number of predictions. The experiments done are exactly the same as the ones for the quadruplet of drugs.

Table 26 compares the predictions on the triplet and on the quintuplet, in the same way as for the quadruplet. See Tables 40, 41, 44, 45 for the details.
In this case, Scenario 11 seems to be less impacted by the change from triplet to quintuplet than Scenario 2. Interestingly for Scenario 11, the predictions with and without using ComboFM are much closer to each other in the case of the quadruplet than in the case of the quintuplet.

| | Triplet S11 | Quadruplet S11 | Triplet S2 | Quadruplet S2 |
|---|---|---|---|---|
| Pearson | 0.9312 | 0.9178 (−1.439%) | 0.9964 | 0.9792 (−1.726%) |
| Spearman | 0.9547 | 0.9227 (−3.3518%) | 0.9953 | 0.9810 (−1.436%) |
| $R^2$ | 0.7635 | 0.8148 (+6.71%) | 0.9884 | 0.9514 (−3.743%) |
| RMSE | 0.1537 | 0.1198 (−22.05%) | 0.0338 | 0.0685 (+102.66%) |
| Percentage of convergence | 99.92 | 99.75 | 99.18 | 100 |

Table 25: Comparison between the predictions (mean score over five runs) of the Dose model with and without using ComboFM. The comparisons are made on two different scenarios and for the triplet and the quadruplets of drugs. The relative change (in %, with respect to the triplet) between triplet and quadruplet for a scenario is given. A green (respectively, red) arrow means that predictions with and without ComboFM are more similar for the quadruplet than the triplet (resp., for the triplet than the quadruplet).

This suggests that the chosen scenario will also impact the performance for the predictions of responses of higher order drug combinations.

### 6.9.3 Conclusion

This experiment does not allow us to know whether the Dose model gives good performance of combinations of order higher than 3 or not. Without having measured data, we can not assess our predictions. However, the goal of this work is to see if we can use the Dose model in combination with ComboFM (assuming thus that the Dose model is efficient). This experiment can give the beginning of an answer to this.
Depending on the chosen scenario, the impact of using ComboFM in addition to the Dose model is not the same. This suggests that, in order to use ComboFM with the Dose model for drug combinations of order higher than 3, one must carefully choose its set-up. It is important to note that the predictions of ComboFM have been made using 200 epochs in the training set. However, it has been shown in this work that those predictions can be improved. This improvement could decrease the impact of going from a triplet to a quadruplet or a quintuplet or even higher orders of combinations.

## 6.10 Conclusion of the experiments

The different experiments allowed to understand how ComboFM can be associated with the Dose model in order to predict the responses of higher order combinations of cancer drugs. However, those two models can not be seen as two black boxes that can simply be connected together. The two models must be adapted. Especially, the Dose model is not perfectly suitable for the NCI-ALMANAC dataset (even if we consider the data without time zero measurements). Concerning ComboFM, the pairwise predictions must be good enough in order not to induce errors in the Dose model. As a consequence, the parameters of ComboFM must be optimized. By optimizing the two models, it seems like all the different tested scenarios could give satisfying results. This is interesting because

| | Triplet S11 | Quintuplet S11 | Triplet S2 | Quintuplet S2 |
|---|---|---|---|---|
| Pearson | 0.9312 | 0.9261 (−0.5476%) | 0.9964 | 0.9524 (−4.41%) |
| Spearman | 0.9547 | 0.9275 (−2.849%) | 0.9953 | 0.9508 (−4.47%) |
| $R^2$ | 0.7635 | 0.7123 (−6.70%) | 0.9884 | 0.8705 (−11.928%) |
| RMSE | 0.1537 | 0.0823 (−46.45%) | 0.0338 | 0.0924 (+173.37%) |
| Percentage of convergence | 99.92 | 99.18 | 100 | 100 |

Table 26: Comparison between the predictions (mean score over five runs) of the Dose model with and without using ComboFM. The comparisons are made on two different scenarios and for the triplet and the quadruplets of drugs. The relative change (in %, , with respect to the triplet) between triplet and quintuplet for a scenario is given. A green (respectively, red) arrow means that predictions with and without ComboFM are more similar for the quintuplet than the triplet (resp., for the triplet than the quintuplet)

it means that the chosen scenario could be the one corresponding to the available data, without needing to perform additional measurements.
Nevertheless, all those results and analyses should be read with caution, knowing that they have been done with validation data on only one triplet of drugs. This prevents us from drawing more general conclusions about those models and the new one that can be built from them.

# Chapter 7 Conclusion and future work

The goal of this work was to investigate the possibility to predict the responses of high order cancer drug combinations, using ComboFM. Following the first work hypothesis stating that pairwise responses are sufficient data to rely on in order to predict responses of drug combinations of any order, one can combine ComboFM with the Dose model. The combination of the two models is simple, they are used in series such that the output of ComboFM (*i.e.* the pairwise predictions) is used as (part of) the input of the Dose model. This allows to decrease the amount of required experimental data and, consequently, to save time.

The main challenge in the field of predicting responses of (cancer) drugs in the lack of experimental data, in order to assess the performance of the models. This work has been built around the NCI-ALMANAC dataset, which contains more than 3 millions responses of *pair* of cancer drugs, in different cell lines, and on one additional triplet of drugs (whose pairs are present in the NCI-ALMANAC), in one particular cell line (making a total of 280 data points). This permitted to evaluate the performance of the triplets predictions but results should be analyzed cautiously. Let us also note that we did not have any validation data regarding combinations of cancer drugs of order higher than 3. In the absence of proper validation data, we can not extrapolate the promising results in predicting triplets to higher orders of combination. However, a comparison with the Dose model only tends to suggest that predictions of the Dose model and the combination of ComboFM and the Dose model are relatively similar.

The conducted experiments allowed to better understand how the two models combine together and what it is important to pay attention to. ComboFM can be used in many ways (*i.e.* different prediction scenarios can be defined) and this choice highly impacts the performance. Each scenario allows to reduce the number of needed experimental measurements. Indeed, if ComboFM is not used, one needs to measure all the dose-response matrices of all the pairs in the combination. Using ComboFM allows to only measure the responses that are part of the training set. The less entries there are in the training set, the more complicated is the scenario. ComboFM's predictions are thus slightly less good and, since the Dose model is quite sensitive, the predictions on the triplet are less accurate.

Even though the different experiments allowed to understand how the two models can be combined together and to see that it is possible to obtain satisfying predictions with this, those models were used in series and not as a new model in itself. It would thus be necessary to make different changes in order to have one model composed of the series of ComboFM followed by the Dose model.
Firstly, the cross-validation and parameters optimization must be adapted. Indeed, once again the two models are separated and they must be seen as only one model. As a consequence, the optimization of the parameters must be done on the final predictions. Let us review some parameters that we observed to have an impact. Concerning ComboFM, instead of choosing the parameters that maximizes the Pearson correlation coefficient on the *pairwise* predictions, one should choose the ones maximizing the Pearson coefficient on the *higher order* predictions. In the light of the experiments, other parameters should be optimized in the cross-validation. Indeed, the parameter $t$ of the Hill function of the Dose model (see Equation 19), as well as the number of epochs in the training step of ComboFM could be optimized in order to obtain the best predictions.
In addition, the cross-validation should be made each time a new order of combinations is predicted. Indeed, a major challenge of this work is the lack of validation data for combination of order higher than 3. As a consequence, we do not know if the best combination of parameters would be the same for a triplet of drugs as for a quadruplet or quintuplet for example.
Secondly and for a more practical point of view, the initial implementations of both models were not

designed to combined them, which led to practical issues. Furthermore, it has been explained that the predictions of ComboFM must be converted in order to be used in the Dose model. In total, one thus have to run three different codes in order to obtain the final predictions. This is of course not user-friendly.

Another point to remember is the data on which the models are used. The NCI-ALMANAC dataset may be very convenient for ComboFM, it is not completely adapted to the Dose model, for several reasons. Firstly, the NCI-ALMANAC dataset contains 3×3 dose-response matrices. This gives 9 points to the Dose model, in order to fit the surface of response. This may not be sufficient.
Secondly and more importantly, the mono-therapy data points that are available in the NCI-ALMANAC dataset is not really suitable for the Dose model. As it has been explained, the same measurements (regarding mono-therapy measurements) have been made several time, depending on the studies in which the drug is involved in a pair of drugs. As a consequence, there is a *duplication* and *heterogeneity* of mono-therapy responses. In order to overcome this, choices regarding the mono-therapy data must be made. While making choices do not prevent the Dose model to converge, it is not ideal in order to have an unambiguous and general model. Note also that the choices impact the number of doses at which we have mono-therapy responses. In this work, we had 3 or 6 points (corresponding to only 3 different doses) which may not be enough in order to correctly fit a curve.

The main goal of this work was to investigate the feasibility of predicting responses of higher order cancer drug combinations, while using ComboFM. Once this has been done, one can look more closely to the accuracy of those predictions (*i.e.* the performance of the model) to push the analysis one step further. As it has been mentioned, an experiment suggested that increasing the number of epochs used in the training of ComboFM could help getting better triplet predictions. However, all the other experiments have been made using the default number of epochs, 200. It would thus be interesting to do those experiments again but with an optimized number of epochs. However, this could quickly become very costly in terms of running time.

It must be kept in mind that this whole work has been built around the available validation data. With more validation data and of higher order, the experiments would have been different and would allow to really assess the performance of this new model, combining ComboFM and the Dose model. Future works would definitely be to obtain or measure such validation data.

# Chapter 8 Appendices

## 8.1 Part 2 of experiments: scenario evaluation

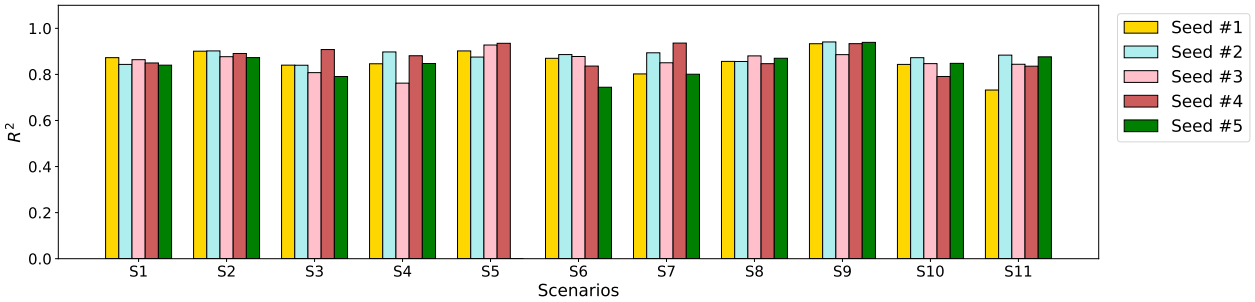



Figure 35: Performance of ComboFM in terms of $R^2$, for the 11 scenarios and the 5 different random seeds.

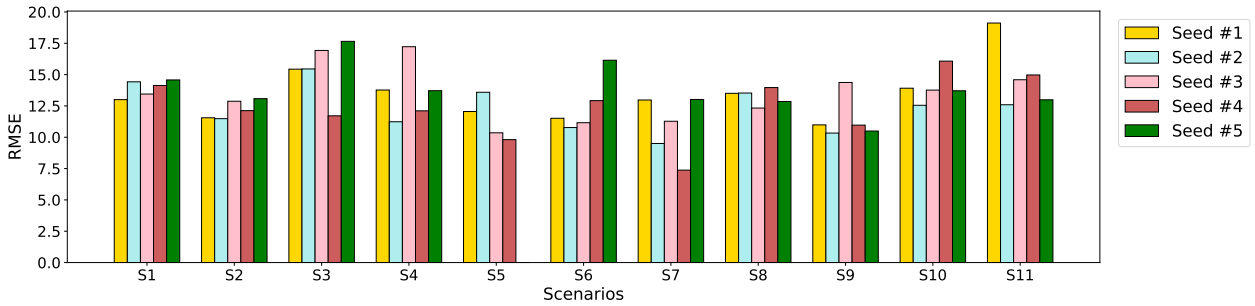



Figure 36: Performance of ComboFM in terms of $RMSE$, for the 11 scenarios and the 5 different random seeds.

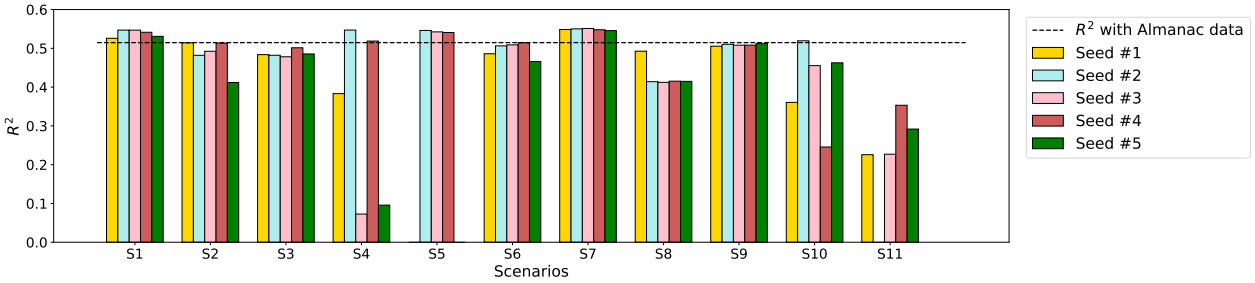



Figure 37: Performance of the Dose model (+ComboFM) in terms of $R^2$, for the 11 scenarios and the 5 different random seeds in ComboFM. A missing bar means that the Dose model did not converge.

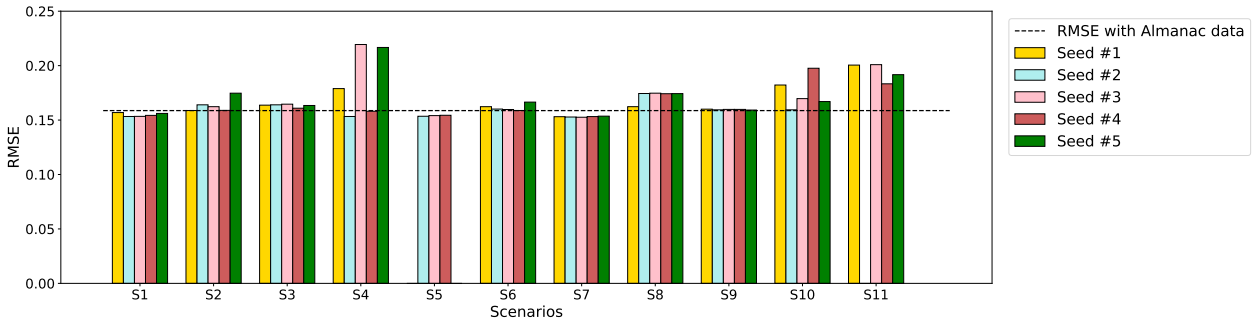



Figure 38: Performance of the Dose model (+ComboFM) in terms of $RMSE$, for the 11 scenarios and the 5 different random seeds in ComboFM.

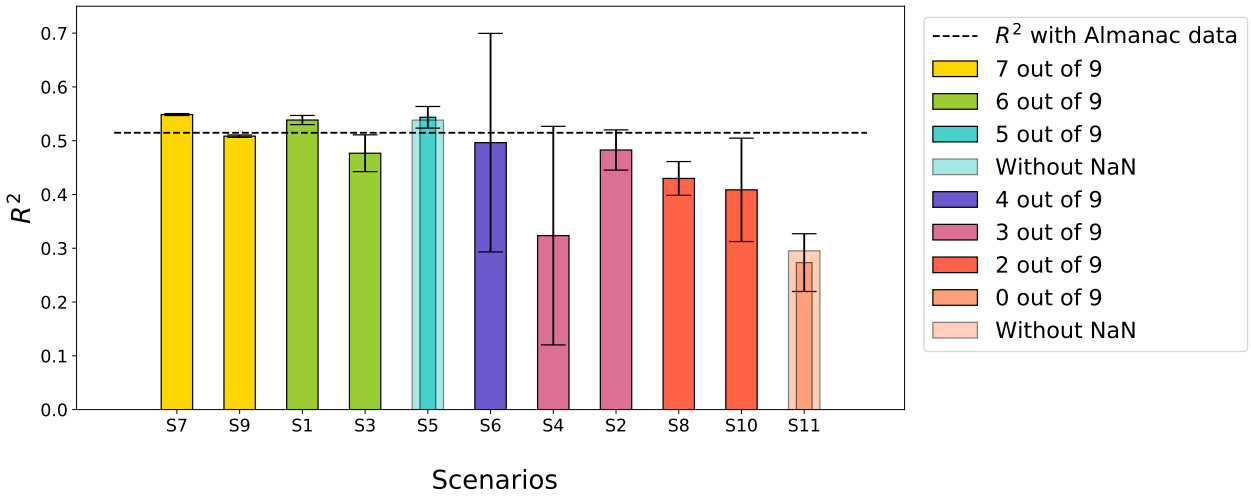



Figure 39: Performance of the Dose model (+ComboFM) in term of $R^2$. The value of the mean across the 5 random seeds by considering all the seeds and removing the NaN values are showed in transparency.

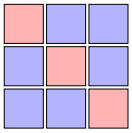
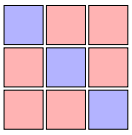
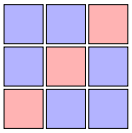
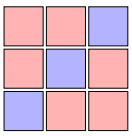
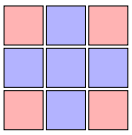
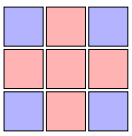
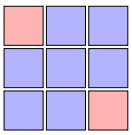
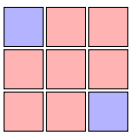
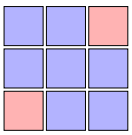
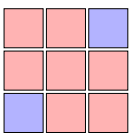
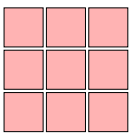
| Name | Schema | CV results | ComboFM | ComboFM + Dose model |
|---|---|---|---|---|
| Scenario 1 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.955<br>Spearman = 0.9333<br>RMSE = 13.0024 | RMSE = 0.1569<br>$R^2$ = 0.5259 |
| Scenario 2 |  | Rank = 100<br>Reg = 100 000 | Pearson = 0.9591<br>Spearman = 0.8906<br>RMSE = 11.554 | RMSE = 0.1588<br>$R^2$ = 0.5141 |
| Scenario 3 |  | Rank = 100<br>Reg = 100 000 | Pearson = 0.9078<br>Spearman = 0.8499<br>RMSE = 15.4343 | RMSE = 0.1637<br>$R^2$ = 0.4838 |
| Scenario 4 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9238<br>Spearman = 0.9153<br>RMSE = 13.7685 | RMSE = 0.1789<br>$R^2$ = 0.3832 |
| Scenario 5 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9549<br>Spearman = 0.9090<br>RMSE = 12.054 | RMSE = /<br>$R^2$ = / |
| Scenario 6 |  | Rank = 100<br>Reg = 1000 | Pearson = 0.9348<br>Spearman = 0.8821<br>RMSE = 11.515 | RMSE = 0.1623<br>$R^2$ = 0.486 |
| Scenario 7 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9200<br>Spearman = 0.9428<br>RMSE = 12.9725 | RMSE = 0.1531<br>$R^2$ = 0.5485 |
| Scenario 8 |  | Rank = 100<br>Reg = 1000 | Pearson = 0.9471<br>Spearman = 0.8779<br>RMSE = 13.5068 | RMSE = 0.1623<br>$R^2$ = 0.4926 |
| Scenario 9 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9665<br>Spearman = 0.7714<br>RMSE = 10.9876 | RMSE = 0.1623<br>$R^2$ = 0.5055 |
| Scenario 10 |  | Rank = 100<br>Reg = 1000 | Pearson = 0.9356<br>Spearman = 0.9103<br>RMSE = 13.914 | RMSE = 0.1822<br>$R^2$ = 0.3606 |
| Scenario 11 |  | Rank = 1000<br>Reg = 10 000 | Pearson = 0.9226<br>Spearman = 0.9346<br>RMSE = 19.114 | RMSE = 0.2005<br>$R^2$ = 0.2257 |

Table 27: Detailed results of the first experiment for the first random seed and the 11 scenarios.

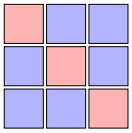
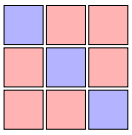
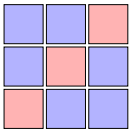
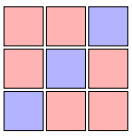
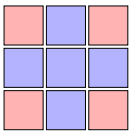
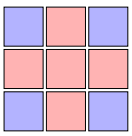
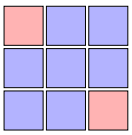
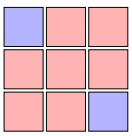
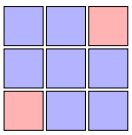
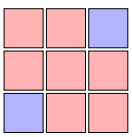
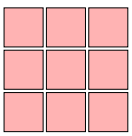
| Name | Schema | CV results | ComboFM | ComboFM + Dose model |
|---|---|---|---|---|
| Scenario 1 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9415<br>Spearman = 0.9333<br>RMSE = 14.424 | RMSE = 0.1533<br>$R^2$ = 0.5470 |
| Scenario 2 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9531<br>Spearman = 0.8885<br>RMSE = 11.4839 | RMSE = 0.164<br>$R^2$ = 0.482 |
| Scenario 3 |  | Rank = 100<br>Reg = 100 000 | Pearson = 0.9404<br>Spearman = 0.9667<br>RMSE = 15.4511 | RMSE = 0.1640<br>$R^2$ = 0.4821 |
| Scenario 4 |  | Rank = 100<br>Reg = 1000 | Pearson = 0.9238<br>Spearman = 0.9153<br>RMSE = 13.7685 | RMSE = 0.1533<br>$R^2$ = 0.5471 |
| Scenario 5 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9394<br>Spearman = 0.9510<br>RMSE = 13.59 | RMSE = 0.1535<br>$R^2$ = 0.5461 |
| Scenario 6 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9478<br>Spearman = 0.8964<br>RMSE = 10.7744 | RMSE = 0.1601<br>$R^2$ = 0.5063 |
| Scenario 7 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9373<br>Spearman = 1<br>RMSE = 9.5036 | RMSE = 0.1528<br>$R^2$ = 0.5501 |
| Scenario 8 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9411<br>Spearman = 0.9376<br>RMSE = 13.5285 | RMSE = 0.1744<br>$R^2$ = 0.4140 |
| Scenario 9 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9797<br>Spearman = 0.7714<br>RMSE = 10.334 | RMSE = 0.1594<br>$R^2$ = 0.5104 |
| Scenario 10 |  | Rank = 100<br>Reg = 100 000 | Pearson = 0.9454<br>Spearman = 0.9246<br>RMSE = 12.5516 | RMSE = 0.1594<br>$R^2$ = 0.5194 |
| Scenario 11 |  | Rank = 1000<br>Reg = 10 000 | Pearson = 0.9479<br>Spearman = 0.9175<br>RMSE = 12.5938 | RMSE = /<br>$R^2$ = / |

Table 28: Detailed results of the first experiment for the second random seed and the 11 scenarios.

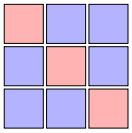
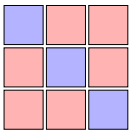
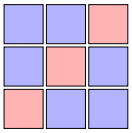
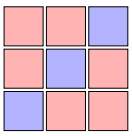
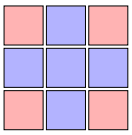
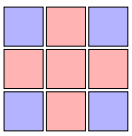
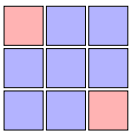
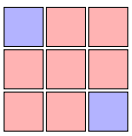
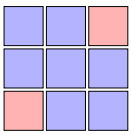
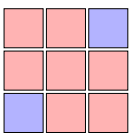
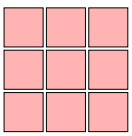
| Name | Schema | CV results | ComboFM | ComboFM + Dose model |
|---|---|---|---|---|
| Scenario 1 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9494<br>Spearman = 0.9500<br>RMSE = 13.447 | RMSE = 0.1534<br>$R^2$ = 0.5468 |
| Scenario 2 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9514<br>Spearman = 0.9256<br>RMSE = 12.8732 | RMSE = 0.1623<br>$R^2$ = 0.4924 |
| Scenario 3 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9371<br>Spearman = 0.9500<br>RMSE = 16.9276 | RMSE = 0.1646<br>$R^2$ = 0.4782 |
| Scenario 4 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.882<br>Spearman = 0.8679<br>RMSE = 17.122 | RMSE = 0.2194<br>$R^2$ = 0.0727 |
| Scenario 5 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9632<br>Spearman = 0.9650<br>RMSE = 10.3547 | RMSE = 0.1541<br>$R^2$ = 0.5424 |
| Scenario 6 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9421<br>Spearman = 0.8678<br>RMSE = 11.1603 | RMSE = 0.1597<br>$R^2$ = 0.5089 |
| Scenario 7 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9670<br>Spearman = 0.9428<br>RMSE = 11.2764 | RMSE = 0.1526<br>$R^2$ = 0.5512 |
| Scenario 8 |  | Rank = 100<br>Reg = 1000 | Pearson = 0.9459<br>Spearman = 0.9090<br>RMSE = 12.3308 | RMSE = 0.1747<br>$R^2$ = 0.4123 |
| Scenario 9 |  | Rank = 100<br>Reg = 10 00 | Pearson = 0.9549<br>Spearman = 0.7714<br>RMSE = 14.3724 | RMSE = 0.1598<br>$R^2$ = 0.5081 |
| Scenario 10 |  | Rank = 100<br>Reg = 100 000 | Pearson = 0.9234<br>Spearman = 0.8844<br>RMSE = 13.7629 | RMSE = 0.1697<br>$R^2$ = 0.4553 |
| Scenario 11 |  | Rank = 1000<br>Reg = 100 | Pearson = 0.9532<br>Spearman = 0.9285<br>RMSE = 14.5895 | RMSE = 0.2009<br>$R^2$ = 0.2227 |

Table 29: Detailed results of the first experiment for the third random seed and the 11 scenarios.

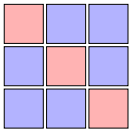
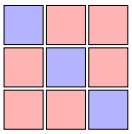
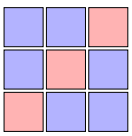
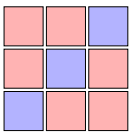
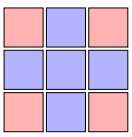
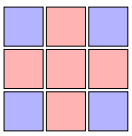
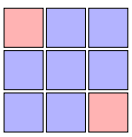
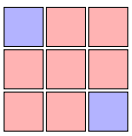
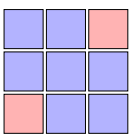
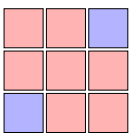
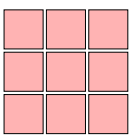
| Name | Schema | CV results | ComboFM | ComboFM + Dose model |
|---|---|---|---|---|
| Scenario 1 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9594<br>Spearman = 0.9333<br>RMSE = 14.1316 | RMSE = 0.1543<br>$R^2$ = 0.5414 |
| Scenario 2 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9629<br>Spearman = 0.9318<br>RMSE = 12.1244 | RMSE = 0.1589<br>$R^2$ = 0.5133 |
| Scenario 3 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9639<br>Spearman = 0.9500<br>RMSE = 11.711 | RMSE = 0.1603<br>$R^2$ = 0.5014 |
| Scenario 4 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9441<br>Spearman = 0.9195<br>RMSE = 12.1086 | RMSE = 0.1581<br>$R^2$ = 0.5187 |
| Scenario 5 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9664<br>Spearman = 0.9720<br>RMSE = 9.9656 | RMSE = 0.1544<br>$R^2$ = 0.5408 |
| Scenario 6 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9216<br>Spearman = 0.8571<br>RMSE = 12.9228 | RMSE = 0.1588<br>$R^2$ = 0.5145 |
| Scenario 7 |  | Rank = 100<br>Reg = 1000 | Pearson = 0.9746<br>Spearman = 1<br>RMSE = 7.377 | RMSE = 0.1532<br>$R^2$ = 0.5479 |
| Scenario 8 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9478<br>Spearman = 0.9051<br>RMSE = 13.9658 | RMSE = 0.1742<br>$R^2$ = 0.4153 |
| Scenario 9 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9815<br>Spearman = 0.8857<br>RMSE = 10.968 | RMSE = 0.1598<br>$R^2$ = 0.5082 |
| Scenario 10 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9278<br>Spearman = 0.8948<br>RMSE = 16.075 | RMSE = 0.1976<br>$R^2$ = 0.2456 |
| Scenario 11 |  | Rank = 1000<br>Reg = 1000 | Pearson = 0.9350<br>Spearman = 0.9041<br>RMSE = 14.9715 | RMSE = 0.1833<br>$R^2$ = 0.3531 |

Table 30: Detailed results of the first experiment for the fourth random seed and the 11 scenarios.

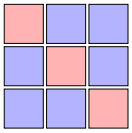
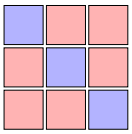
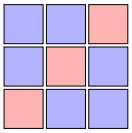
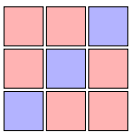
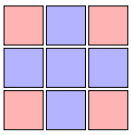
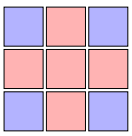
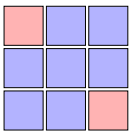
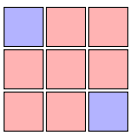
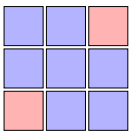
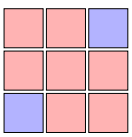
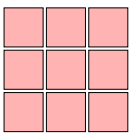
| Name | Schema | CV results | ComboFM | ComboFM + Dose model |
|---|---|---|---|---|
| Scenario 1 |  | Rank = 100<br>Reg = 10 000 | Pearson = 0.9365<br>Spearman = 0.9333<br>RMSE = 14.5752 | RMSE = 0.1561<br>$R^2$ = 0.5309 |
| Scenario 2 |  | Rank = 100<br>Reg = 1000 | Pearson = 0.9598<br>Spearman = 0.9298<br>RMSE = 13.081 | RMSE = 0.1747<br>$R^2$ = 0.4120 |
| Scenario 3 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9201<br>Spearman = 0.8166<br>RMSE = 17.6542 | RMSE = 0.1634<br>$R^2$ = 0.4855 |
| Scenario 4 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9409<br>Spearman = 0.8968<br>RMSE = 13.715 | RMSE = 0.2167<br>$R^2$ = 0.0957 |
| Scenario 5 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9713<br>Spearman = 0.9230<br>RMSE = 10.6091 | RMSE = 0.1545<br>$R^2$ = 0.5399 |
| Scenario 6 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9384<br>Spearman = 0.8652<br>RMSE = 16.4608 | RMSE = 0.1665<br>$R^2$ = 0.4659 |
| Scenario 7 |  | Rank = 100<br>Reg = 1000 | Pearson = 0.9563<br>Spearman = 0.9428<br>RMSE = 13.0158 | RMSE = 0.1536<br>$R^2$ = 0.5457 |
| Scenario 8 |  | Rank = 100<br>Reg = 100 | Pearson = 0.9421<br>Spearman = 0.8900<br>RMSE = 11.8518 | RMSE = 0.1743<br>$R^2$ = 0.4148 |
| Scenario 9 |  | Rank = 100<br>Reg = 100 000 | Pearson = 0.9714<br>Spearman = 0.9428<br>RMSE = 10.4970 | RMSE = 0.1592<br>$R^2$ = 0.5120 |
| Scenario 10 |  | Rank = 100<br>Reg = 1000 | Pearson = 0.9281<br>Spearman = 0.8831<br>RMSE = 13.709 | RMSE = 0.1670<br>$R^2$ = 0.4626 |
| Scenario 11 |  | Rank = 1000<br>Reg = 100 | Pearson = 0.9444<br>Spearman = 0.9279<br>RMSE = 12.989 | RMSE = 0.1917<br>$R^2$ = 0.2918 |

Table 31: Detailed results of the first experiment for the fifth random seed and the 11 scenarios.

## 8.2 Part 3 of experiments : sensitivity of the models

### 8.2.1 Scenario 11

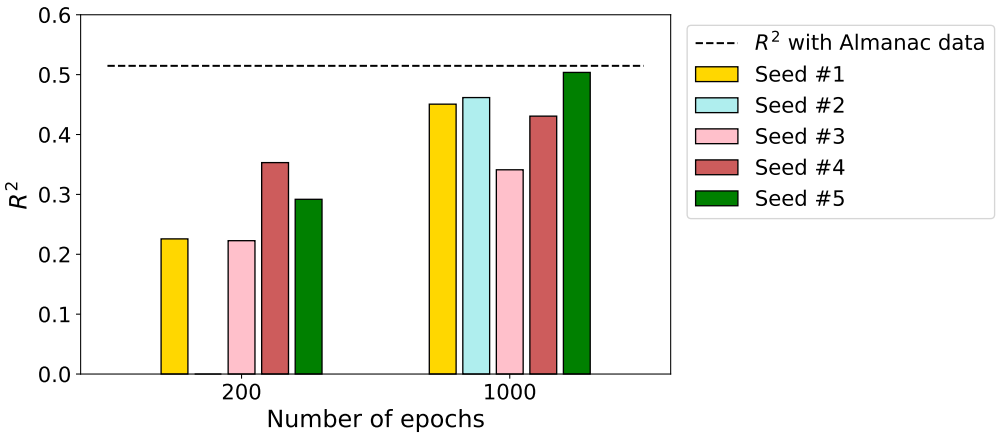



Figure 40: Comparison of the performance (in terms of $R^2$) of the Dose model combined with ComboFM (Scenario 11), depending on the number of epochs used in the training of ComboFM.

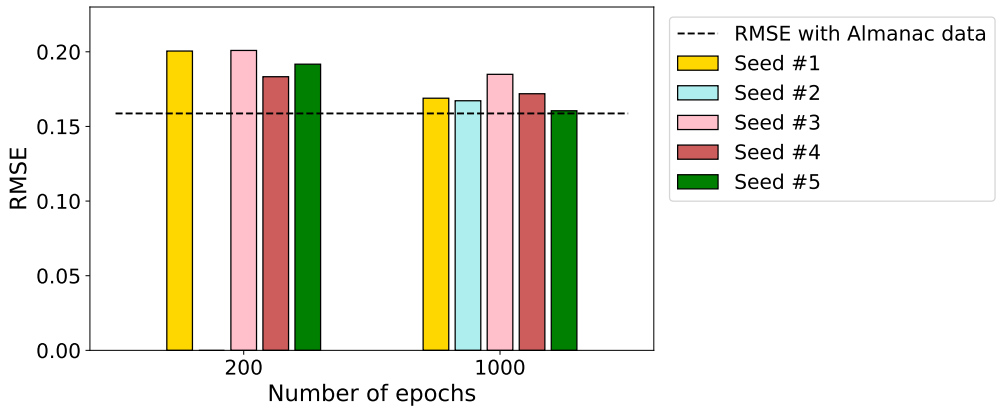



Figure 41: Comparison of the performance (in terms of $RMSE$) of the Dose model combined with ComboFM (Scenario. 11), depending on the number of epochs used in the training of ComboFM.

### 8.2.2 Scenario 8

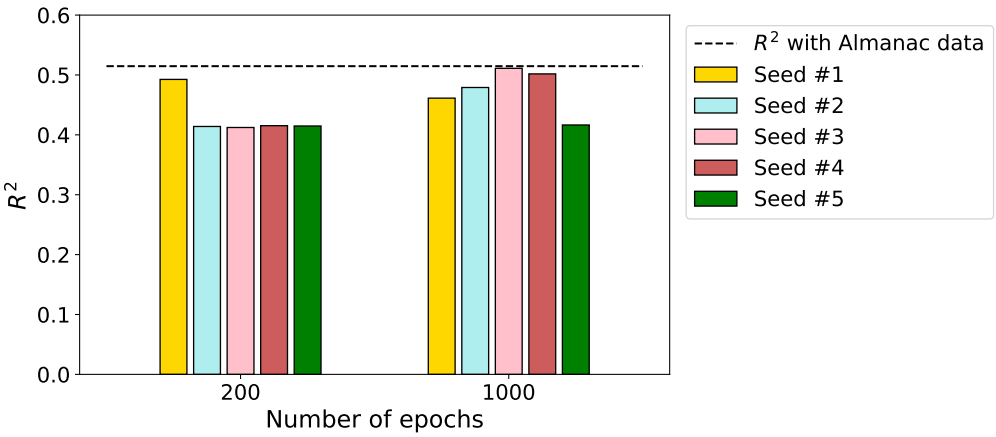



Figure 42: Comparison of the performance (in terms of $R^2$) of the Dose model combined with ComboFM (Scenario 8), depending on the number of epochs used in the training of ComboFM.

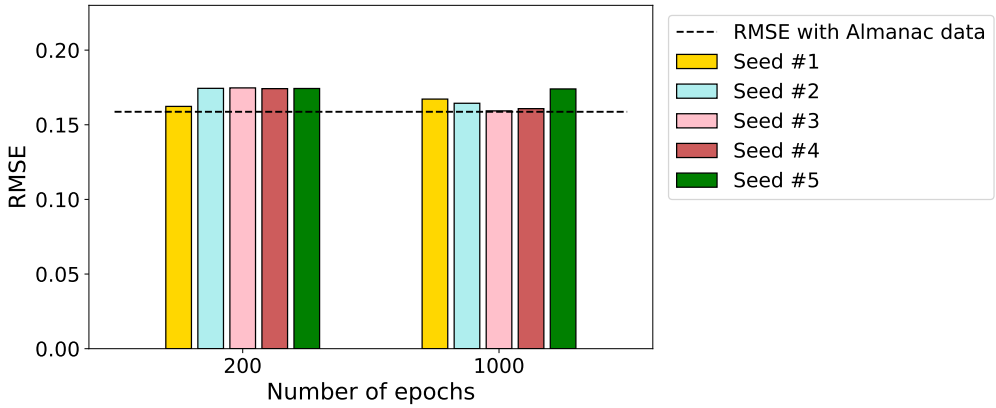



Figure 43: Comparison of the performance (in terms of $RMSE$) of the Dose model combined with ComboFM (Scenario 8), depending on the number of epochs used in the training of ComboFM.

## 8.3 Part 4 of experiments : synergy score

### 8.3.1 On validation data

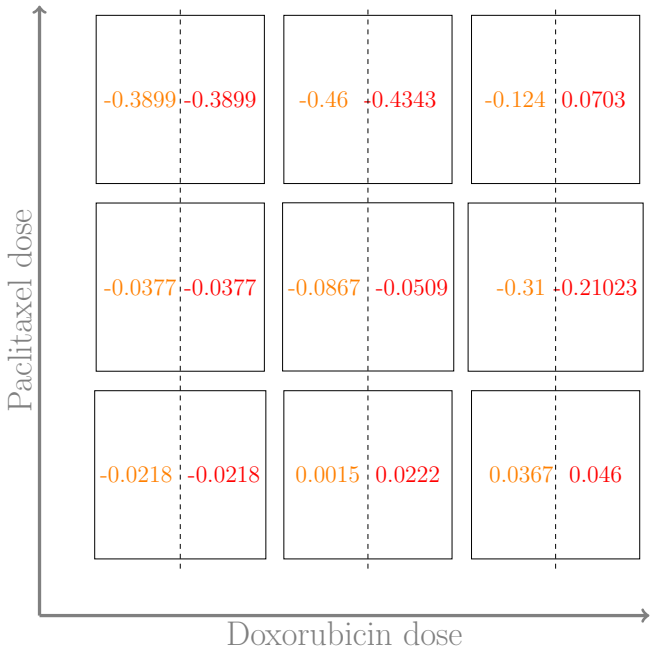



Figure 44: Details of synergy scores for the pair Doxorubicin-Paclitaxel on validation data. Each entry of the dose-response matrix has a score. Color representation: orange=Bliss score; red=HSA score.

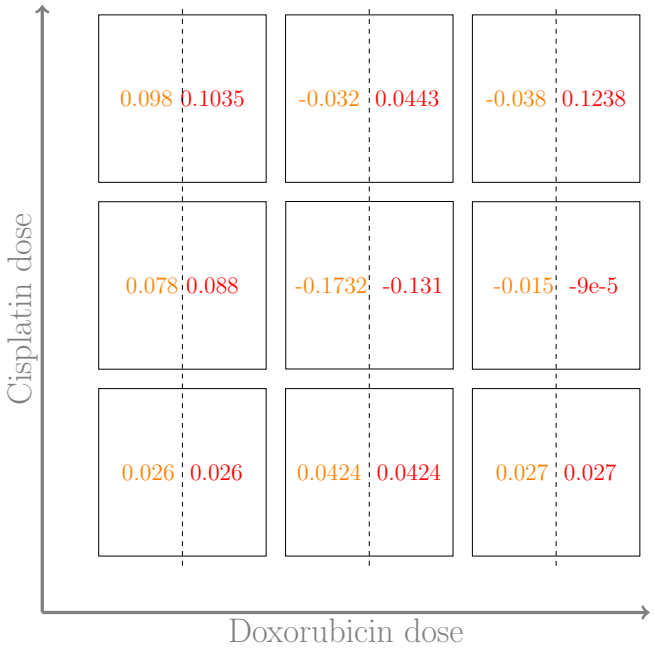



Figure 45: Details of synergy scores for the pair Doxorubicin-Cisplatin on validation data. Each entry of the dose-response matrix has a score. Color representation: orange=Bliss score; red=HSA score.

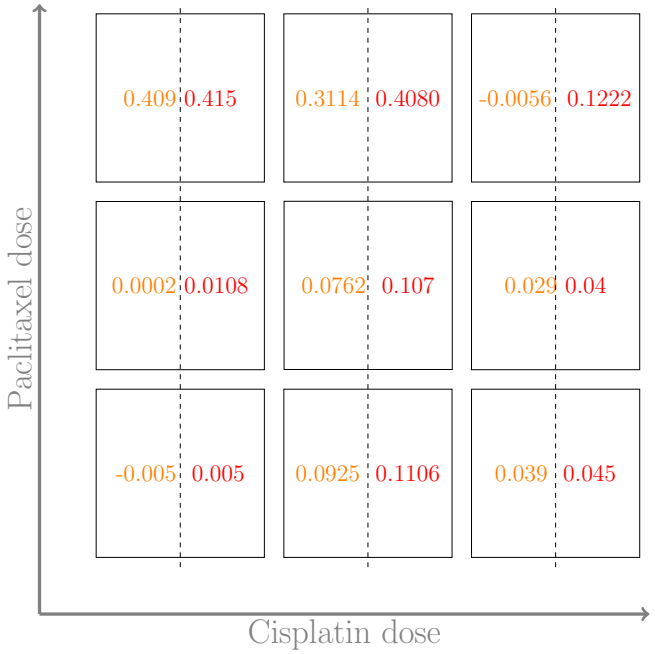



Figure 46: Details of synergy scores for the pair Cisplatin-Paclitaxel on validation data. Each entry of the dose-response matrix has a score. Color representation: orange=Bliss score; red=HSA score.

| Bliss | HSA | Bliss | HSA |
|---|---|---|---|
| -0.0326 | -0.021 | -0.046 | 0.038 |
| -0.0479 | -0.037 | -0.0519 | 0.038 |
| -0.398 | -0.392 | -0.199 | 0.039 |
| -0.018 | 0.012 | -0.0165 | -0.01 |
| -0.032 | 0.012 | -0.0187 | -0.0076 |
| -0.4023 | -0.36 | -0.1078 | 0.02 |
| -0.011 | 0.0036 | -0.0473 | -0.0262 |
| -0.018 | 0.0036 | -0.0515 | -0.0260 |
| -0.1939 | 0.0038 | -0.159 | -0.022 |
| -0.0385 | -0.02 | -0.083 | 0.081 |
| -0.0486 | -0.0017 | -0.0853 | 0.081 |
| -0.332 | 0.0236 | -0.1375 | 0.082 |
| -0.0442 | 0.015 | | |
| -0.0564 | 0.015 | | |
| -0.365 | -0.244 | | |

Table 32: Details of synergy scores (Bliss and HSA score) for the triplet Doxorubicin-Cisplatin-Paclitaxel on validation data.

### 8.3.2 On ComboFM seed # 1

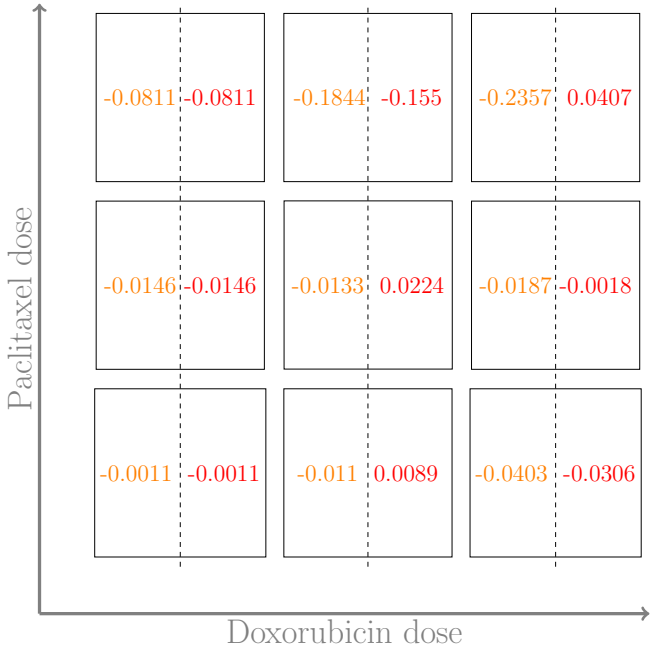



Figure 47: Details of synergy scores for the pair Doxorubicin-Paclitaxel on the predictions made by ComboFM (first random seed). Each entry of the dose-response matrix has a score. Color representation: orange=Bliss score; red=HSA score.

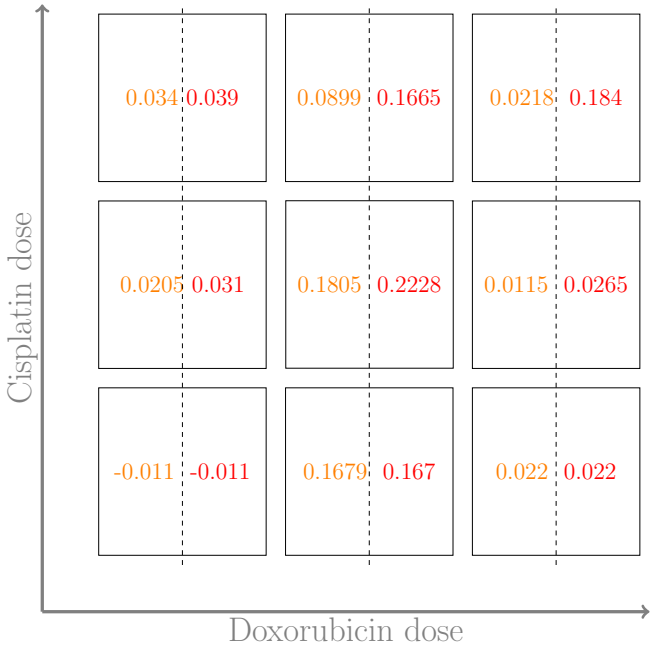



Figure 48: Details of synergy scores for the pair Doxorubicin-Cisplatin on the predictions made by ComboFM (first random seed). Each entry of the dose-response matrix has a score. Color representation: orange=Bliss score; red=HSA score.

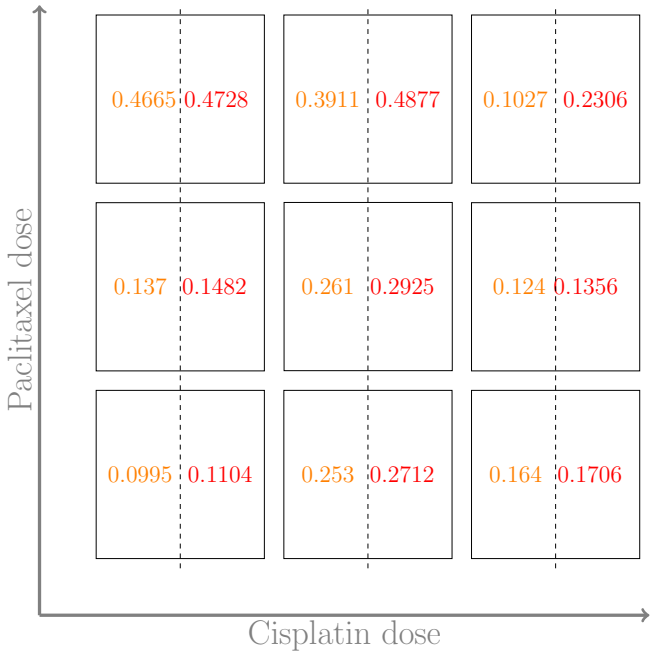



Figure 49: Details of synergy scores for the pair Cisplatin-Paclitaxel on the predictions made by ComboFM (first random seed). Each entry of the dose-response matrix has a score. Color representation: orange=Bliss score; red=HSA score.

| Bliss | HSA | Bliss | HSA |
|---|---|---|---|
| -0.0321 | -0.021 | -0.141 | -0.021 |
| -0.0155 | -0.0048 | 0.061 | 0.1458 |
| -0.0479 | -0.041 | 0.056 | 0.1474 |
| -0.0125 | 0.018 | -0.069 | 0.1687 |
| -0.0158 | 0.03 | -0.0066 | -0.0002 |
| -0.218 | -0.1836 | 0.01 | 0.0217 |
| 0.003 | 0.0177 | 0.047 | 0.175 |
| -0.002 | 0.0189 | 0.018 | 0.0398 |
| 0.158 | 0.039 | 0.022 | 0.0484 |
| -0.03 | -0.012 | -0.007 | 0.129 |
| -0.008 | 0.022 | 0.06 | 0.227 |
| -0.009 | 0.106 | 0.06 | 0.227 |
| -0.0006 | 0.0587 | | |
| 0.00035 | 0.0723 | | |
| 0.015 | 0.235 | | |

Table 33: Details of synergy scores (Bliss and HSA score) for the triplet Doxorubicin-Cisplatin-Paclitaxel on the combination of ComboFM (seed # 1) and the Dose model.

## 8.4 Part 5 of experiments : comparison with analytical formulas

### 8.4.1 Without ComboFM

| Entry | Dose model | Pair model | Iserliss | Regression | Bliss |
|---|---|---|---|---|---|
| 1 | 1.0019 | 0.9902 | 0.9937 | 0.9914 | 0.9890 |
| 2 | 0.9995 | 0.9617 | 0.9517 | 0.9505 | 0.9729 |
| 3 | 0.9583 | 0.3714 | 0.4991 | 0.2412 | 0.5720 |
| 4 | 0.9362 | 0.8806 | 0.8451 | 0.8448 | 0.9180 |
| 5 | 0.9361 | 0.9124 | 0.9306 | 0.9219 | 0.9301 |
| 6 | 0.9332 | 0.3591 | 0.5561 | 0.2429 | 0.5309 |
| 7 | 0.4454 | 0.3678 | 0.3042 | 0.3115 | 0.4343 |
| 8 | 0.4454 | 0.4956 | 0.6441 | 0.5749 | 0.4273 |
| 9 | 0.4452 | 0.14 | 0.1349 | 0.0781 | 0.2512 |
| 10 | 0.8497 | 0.7692 | 0.7136 | 0.7135 | 0.8293 |
| 11 | 0.8466 | 0.7471 | 0.7004 | 0.7026 | 0.8158 |
| 12 | 0.8019 | 0.3442 | 0.4394 | 0.2470 | 0.4796 |
| 13 | 0.8140 | 0.7997 | 0.8502 | 0.8308 | 0.7698 |
| 14 | 0.8137 | 0.8397 | 0.9380 | 0.9311 | 0.7572 |
| 15 | 0.8102 | 0.389 | 0.6318 | 0.34 | 0.4452 |
| 16 | 0.4102 | 0.3421 | 0.3239 | 0.3214 | 0.3642 |
| 17 | 0.4102 | 0.4672 | 0.6166 | 0.6092 | 0.3583 |
| 18 | 0.41 | 0.1554 | 0.1925 | 0.1147 | 0.2106 |
| 19 | 0.3046 | 0.2638 | 0.2349 | 0.2363 | 0.2945 |
| 20 | 0.3022 | 0.2618 | 0.2347 | 0.2366 | 0.2897 |
| 21 | 0.2744 | 0.2096 | 0.2679 | 0.2579 | 0.1703 |
| 22 | 0.3208 | 0.2607 | 0.2506 | 0.2487 | 0.2734 |
| 23 | 0.3205 | 0.2761 | 0.2866 | 0.2834 | 0.2689 |
| 24 | 0.3175 | 0.2253 | 0.3084 | 0.3209 | 0.1581 |
| 25 | 0.2126 | 0.1308 | 0.1386 | 0.1323 | 0.1293 |
| 26 | 0.2126 | 0.1801 | 0.2440 | 0.2549 | 0.1271 |
| 27 | 0.2123 | 0.1055 | 0.1358 | 0.1488 | 0.0748 |

Table 34: Details results of the triplets predictions with other analytical formulas. The predictions have been made using only the Dose model.

### 8.4.2 With ComboFM

**Seed #1**

| Entry | Dose model | Pair model | Iserliss | Regression | Bliss |
|---|---|---|---|---|---|
| 1 | 0.9995 | 0.9417 | 0.8984 | 0.8967 | 0.989 |
| 2 | 0.9672 | 0.9110 | 0.8566 | 0.8529 | 0.9729 |
| 3 | 0.6074 | 0.2503 | 0.1810 | 0.1096 | 0.5720 |
| 4 | 0.9305 | 0.8653 | 0.8149 | 0.8156 | 0.9180 |
| 5 | 0.9189 | 0.8320 | 0.7659 | 0.7666 | 0.9031 |
| 6 | 0.7494 | 0.2479 | 0.2588 | 0.1158 | 0.5309 |
| 7 | 0.4313 | 0.4130 | 0.3962 | 0.3928 | 0.4343 |
| 8 | 0.4300 | 0.3879 | 0.3512 | 0.3521 | 0.4279 |
| 9 | 0.4097 | 0.1365 | 0.2555 | 0.0742 | 0.2512 |
| 10 | 0.8419 | 0.6011 | 0.3707 | 0.4357 | 0.8293 |
| 11 | 0.8064 | 0.5888 | 0.3634 | 0.4249 | 0.8158 |
| 12 | 0.4594 | 0.1826 | 0.0309 | 0.0695 | 0.4796 |
| 13 | 0.7707 | 0.5639 | 0.3625 | 0.4131 | 0.7698 |
| 14 | 0.7569 | 0.5941 | 0.3466 | 0.3982 | 0.75752 |
| 15 | 0.5870 | 0.1846 | 0.1247 | 0.0766 | 0.4452 |
| 16 | 0.3032 | 0.2749 | 0.1961 | 0.2075 | 0.3642 |
| 17 | 0.3016 | 0.2614 | 0.17 | 0.1907 | 0.3583 |
| 18 | 0.2806 | 0.1038 | 0.1796 | 0.0512 | 0.2106 |
| 19 | 0.2948 | 0.1817 | 0.0923 | 0.1121 | 0.2945 |
| 20 | 0.2727 | 0.2056 | 0.1326 | 0.1458 | 0.2897 |
| 21 | 0.1188 | 0.1061 | 0.0692 | 0.0661 | 0.1703 |
| 22 | 0.2546 | 0.1766 | 0.1097 | 0.1141 | 0.2734 |
| 23 | 0.2461 | 0.1986 | 0.1436 | 0.1467 | 0.2689 |
| 24 | 0.1653 | 0.1111 | 0.1084 | 0.0781 | 0.1581 |
| 25 | 0.0674 | 0.081 | 0.0462 | 0.0507 | 0.1293 |
| 26 | 0.0668 | 0.0889 | 0.0559 | 0.0622 | 0.1271 |
| 27 | 0.0594 | 0.0588 | 0.0858 | 0.0462 | 0.0748 |

Table 35: Details results of the triplets predictions with other analytical formulas. The predictions have been made using ComboFM (first random seed) and the Dose model.

**Seed #2**

| Entry | Dose model | Pair model | Iserliss | Regression | Bliss |
|---|---|---|---|---|---|
| 1 | 1.0004 | 0.9726 | 0.9560 | 0.9564 | 0.9890 |
| 2 | 0.9840 | 0.9918 | 1.0123 | 1.0110 | 0.9729 |
| 3 | 0.7559 | 0.4643 | 0.4794 | 0.3768 | 0.5720 |
| 4 | 0.9311 | 0.9164 | 0.9150 | 0.9148 | 0.9180 |
| 5 | 0.9291 | 0.9206 | 0.9391 | 0.9345 | 0.9031 |
| 6 | 0.8947 | 0.4560 | 0.5279 | 0.3916 | 0.5309 |
| 7 | 0.4313 | 0.3882 | 0.3685 | 0.3470 | 0.4343 |
| 8 | 0.4312 | 0.4041 | 0.4124 | 0.3812 | 0.4273 |
| 9 | 0.4286 | 0.2195 | 0.3600 | 0.1918 | 0.2512 |
| 10 | 0.8428 | 0.7265 | 0.6227 | 0.6364 | 0.8293 |
| 11 | 0.8194 | 0.7598 | 0.7103 | 0.7076 | 0.8158 |
| 12 | 0.5335 | 0.3416 | 0.3147 | 0.2433 | 0.4796 |
| 13 | 0.7712 | 0.6972 | 0.6255 | 0.6315 | 0.7698 |
| 14 | 0.7679 | 0.7184 | 0.6829 | 0.6815 | 0.7572 |
| 15 | 0.7196 | 0.3417 | 0.3757 | 0.2623 | 0.4452 |
| 16 | 0.3032 | 0.2907 | 0.2449 | 0.2320 | 0.3642 |
| 17 | 0.3028 | 0.3103 | 0.3077 | 0.2687 | 0.3583 |
| 18 | 0.2980 | 0.1619 | 0.2719 | 0.1244 | 0.2106 |
| 19 | 0.2928 | 0.2169 | 0.1387 | 0.1598 | 0.2945 |
| 20 | 0.2475 | 0.2472 | 0.2097 | 0.2109 | 0.2897 |
| 21 | 0.0017 | 0.1310 | 0.1173 | 0.1007 | 0.1703 |
| 22 | 0.2547 | 0.2247 | 0.1801 | 0.1847 | 0.2734 |
| 23 | 0.2475 | 0.2523 | 0.2370 | 0.2366 | 0.2689 |
| 24 | 0.1629 | 0.1414 | 0.1586 | 0.1265 | 0.1581 |
| 25 | 0.0674 | 0.0924 | 0.0644 | 0.0661 | 01293 |
| 26 | 0.0667 | 0.1075 | 0.1040 | 0.0909 | 0.1271 |
| 27 | 0.0581 | 0.0661 | 0.1069 | 0.0584 | 0.0748 |

Table 36: Details results of the triplets predictions with other analytical formulas. The predictions have been made using ComboFM (second random seed) and the Dose model.

**Seed #3**

| Entry | Dose model | Pair model | Iserliss | Regression | Bliss |
|---|---|---|---|---|---|
| 1 | 0.9995 | 0.9429 | 0.8969 | 0.8989 | 0.9890 |
| 2 | 0.9671 | 0.9214 | 0.8734 | 0.8726 | 0.9729 |
| 3 | 0.6067 | 0.3440 | 0.2943 | 0.2069 | 0.5720 |
| 4 | 0.9305 | 0.8503 | 0.7813 | 0.7875 | 0.9180 |
| 5 | 0.9188 | 0.8265 | 0.7504 | 0.7564 | 0.9031 |
| 6 | 0.7476 | 0.3207 | 0.2873 | 0.1937 | 0.5309 |
| 7 | 0.4313 | 0.3523 | 0.2738 | 0.2859 | 0.4343 |
| 8 | 0.4300 | 0.3853 | 0.3728 | 0.3474 | 0.4273 |
| 9 | 0.4093 | 0.1592 | 0.2209 | 0.1009 | 0.2512 |
| 10 | 0.8419 | 0.6678 | 0.4992 | 0.5377 | 0.8293 |
| 11 | 0.8063 | 0.6469 | 0.4859 | 0.5129 | 0.8158 |
| 12 | 0.4588 | 0.2200 | 0.1377 | 0.1009 | 0.4796 |
| 13 | 0.7704 | 0.6141 | 0.4509 | 0.4900 | 0.7698 |
| 14 | 0.7567 | 0.5918 | 0.4304 | 0.4625 | 0.7572 |
| 15 | 0.5853 | 0.2092 | 0.1573 | 0.0983 | 0.4452 |
| 16 | 0.3031 | 0.2466 | 0.1256 | 0.1670 | 0.3642 |
| 17 | 0.3014 | 0.2673 | 0.2113 | 0.1994 | 0.3583 |
| 18 | 0.2798 | 0.1006 | 0.1414 | 0.0480 | 0.2106 |
| 19 | 0.2498 | 0.2246 | 0.1688 | 0.1712 | 0.2945 |
| 20 | 0.2727 | 0.2224 | 0.1772 | 0.1707 | 0.2897 |
| 21 | 0.1185 | 0.1368 | 0.1312 | 0.1099 | 0.1703 |
| 22 | 0.2546 | 0.2144 | 0.1734 | 0.1681 | 0.2734 |
| 23 | 0.2460 | 0.2112 | 0.1780 | 0.1658 | 0.2689 |
| 24 | 0.1646 | 0.1351 | 0.1444 | 0.1154 | 0.1581 |
| 25 | 0.0674 | 0.0811 | 0.0302 | 0.0508 | 0.1293 |
| 26 | 0.0667 | 0.0898 | 0.0666 | 0.0634 | 0.1272 |
| 27 | 0.0592 | 0.0612 | 0.0802 | 0.0500 | 0.0748 |

Table 37: Details results of the triplets predictions with other analytical formulas. The predictions have been made using ComboFM (random seed number three) and the Dose model.

**Seed #4**

| Entry | Dose model | Pair model | Iserliss | Regression | Bliss |
|---|---|---|---|---|---|
| 1 | 0.9999 | 1.0166 | 1.0467 | 1.0449 | 0.9890 |
| 2 | 0.9672 | 0.9878 | 1.0054 | 1.0029 | 0.9729 |
| 3 | 0.6048 | 0.3534 | 0.3214 | 0.2183 | 0.5720 |
| 4 | 0.9339 | 0.9380 | 0.9596 | 0.9584 | 0.9180 |
| 5 | 0.9215 | 0.9282 | 0.9546 | 0.9540 | 0.9031 |
| 6 | 0.7433 | 0.3386 | 0.3414 | 0.2159 | 0.5309 |
| 7 | 0.4418 | 0.4330 | 0.4317 | 0.4316 | 0.4343 |
| 8 | 0.4405 | 0.4462 | 0.4677 | 0.4660 | 0.4273 |
| 9 | 0.4183 | 0.1902 | 0.3110 | 0.1449 | 0.2512 |
| 10 | 0.8468 | 0.7113 | 0.5959 | 0.6101 | 0.8293 |
| 11 | 0.8120 | 0.6977 | 0.5850 | 0.5967 | 0.8158 |
| 12 | 0.4727 | 0.2305 | 0.1664 | 0.1108 | 0.4796 |
| 13 | 0.8021 | 0.6700 | 0.5750 | 0.5832 | 0.7698 |
| 14 | 0.7893 | 0.6692 | 0.5922 | 0.5915 | 0.7572 |
| 15 | 0.6269 | 0.2255 | 0.2089 | 0.1142 | 0.4452 |
| 16 | 0.3811 | 0.2977 | 0.2307 | 0.2434 | 0.3642 |
| 17 | 0.3795 | 0.3097 | 0.2714 | 0.2677 | 0.3583 |
| 18 | 0.3567 | 0.1223 | 0.2111 | 0.0710 | 0.2106 |
| 19 | 0.3085 | 0.2152 | 0.1558 | 0.1573 | 0.2945 |
| 20 | 0.2910 | 0.2048 | 0.1455 | 0.1447 | 0.2897 |
| 21 | 0.1600 | 0.1265 | 0.1132 | 0.0940 | 0.1703 |
| 22 | 0.3119 | 0.2264 | 0.2239 | 0.1876 | 0.2734 |
| 23 | 0.3060 | 0.2194 | 0.2228 | 0.1790 | 0.2689 |
| 24 | 0.2453 | 0.1382 | 0.1658 | 0.1208 | 0.1581 |
| 25 | 0.1304 | 0.0795 | 0.0305 | 0.0489 | 0.1293 |
| 26 | 0.1293 | 0.0802 | 0.0425 | 0.0506 | 0.1272 |
| 27 | 0.1168 | 0.0592 | 0.0831 | 0.0469 | 0.0748 |

Table 38: Details results of the triplets predictions with other analytical formulas. The predictions have been made using ComboFM (random seed number four) and the Dose model.

**Seed #5**

| Entry | Dose model | Pair model | Iserliss | Regression | Bliss |
|---|---|---|---|---|---|
| 1 | 0.9996 | 0.9137 | 0.8370 | 0.8441 | 0.9890 |
| 2 | 0.9700 | 0.9622 | 0.9577 | 0.9517 | 0.9729 |
| 3 | 0.6283 | 0.3359 | 0.2492 | 0.1973 | 0.5720 |
| 4 | 0.9307 | 0.8624 | 0.8069 | 0.8101 | 0.9180 |
| 5 | 0.9223 | 0.8989 | 0.9012 | 0.8948 | 0.9031 |
| 6 | 0.7928 | 0.3307 | 0.3103 | 0.2060 | 0.5309 |
| 7 | 0.4313 | 0.3687 | 0.3177 | 0.3130 | 0.4343 |
| 8 | 0.4305 | 0.4026 | 0.4149 | 0.3793 | 0.4273 |
| 9 | 0.4173 | 0.1734 | 0.2998 | 0.1198 | 0.2512 |
| 10 | 0.8421 | 0.6866 | 0.5403 | 0.5685 | 0.8293 |
| 11 | 0.8094 | 0.7400 | 0.6745 | 0.6712 | 0.8158 |
| 12 | 0.4786 | 0.2375 | 0.1381 | 0.1176 | 0.4796 |
| 13 | 0.7707 | 0.6554 | 0.5412 | 0.5580 | 0.7698 |
| 14 | 0.7607 | 0.6990 | 0.6506 | 0.6453 | 0.7572 |
| 15 | 0.6290 | 0.2364 | 0.2029 | 0.1255 | 0.4452 |
| 16 | 0.3032 | 0.2843 | 0.2145 | 0.2219 | 0.3642 |
| 17 | 0.3021 | 0.3176 | 0.3101 | 0.2815 | 0.3583 |
| 18 | 0.2878 | 0.1258 | 0.2313 | 0.0751 | 0.2106 |
| 19 | 0.2949 | 0.3300 | 0.3690 | 0.3698 | 0.2945 |
| 20 | 0.2744 | 0.3424 | 0.3947 | 0.4047 | 0.2897 |
| 21 | 0.1257 | 0.1573 | 0.1721 | 0.1453 | 0.1703 |
| 22 | 0.2548 | 0.3185 | 0.3671 | 0.3711 | 0.2734 |
| 23 | 0.2483 | 0.3271 | 0.3853 | 0.3978 | 0.2689 |
| 24 | 0.1832 | 0.1583 | 0.1923 | 0.1586 | 0.1581 |
| 25 | 0.0674 | 0.1273 | 0.1271 | 0.1253 | 0.1293 |
| 26 | 0.0673 | 0.1369 | 0.1519 | 0.1573 | 0.1272 |
| 27 | 0.0618 | 0.0776 | 0.1207 | 0.0805 | 0.0748 |

Table 39: Details results of the triplets predictions with other analytical formulas. The predictions have been made using ComboFM (random seed number five) and the Dose model.

## 8.5 Part 6 of experiments : quadruplets and quintuplets

### 8.5.1 Triplet

**Scenario 11**

| | Seed #1 | Seed #2 | Seed #3 | Seed #4 | Seed #5 | Mean |
|---|---|---|---|---|---|---|
| Pearson | 0.9170 | 0.9686 | 0.9163 | 0.9207 | 0.9336 | 0.9312 |
| Spearman | 0.9477 | 0.9699 | 0.9474 | 0.9511 | 0.9574 | 0.9457 |
| $R^2$ | 0.6887 | 0.8966 | 0.7463 | 0.7284 | 0.7575 | 0.7635 |
| RMSE | 0.1730 | 0.1091 | 0.1737 | 0.1576 | 0.1553 | 0.1537 |
| Convergence | 279/280 | 280/280 | 280/280 | 280/280 | 280/280 | 99.92% |

Table 40: Details results for the triplet of drugs with ComboFM in Scenario 11.

**Scenario 2**

| | Seed #1 | Seed #2 | Seed #3 | Seed #4 | Seed #5 | Mean |
|---|---|---|---|---|---|---|
| Pearson | 0.9994 | 0.9958 | 0.9968 | 0.9983 | 0.9919 | 0.9964 |
| Spearman | 0.9997 | 0.9929 | 0.9958 | 0.9978 | 0.9901 | 0.9953 |
| $R^2$ | 0.9986 | 0.9871 | 0.9899 | 0.9952 | 0.9714 | 0.9884 |
| RMSE | 0.0128 | 0.0392 | 0.0347 | 0.0240 | 0.0584 | 0.0338 |
| Convergence | 280/280 | 280/280 | 280/280 | 280/280 | 280/280 | 100% |

Table 41: Details results for the triplet of drugs with ComboFM in Scenario 2.

### 8.5.2 Quadruplet

**Scenario 11**

| | Seed #1 | Seed #2 | Seed #3 | Seed #4 | Seed #5 | Mean |
|---|---|---|---|---|---|---|
| Pearson | 0.8938 | 0.9869 | 0.9431 | 0.8777 | 0.8873 | 0.9178 |
| Spearman | 0.9063 | 0.9849 | 0.9498 | 0.8823 | 0.8902 | 0.9227 |
| $R^2$ | 0.7788 | 0.9693 | 0.8445 | 0.7326 | 0.7490 | 0.8148 |
| RMSE | 0.1362 | 0.0540 | 0.1122 | 0.1501 | 0.1434 | 0.1192 |
| Convergence | 81/81 | 81/81 | 80/81 | 81/81 | 81/81 | 99.75 % |

Table 42: Details results for the quadruplet of drugs with ComboFM in Scenario 11.

**Scenario 2**

| | Seed #1 | Seed #2 | Seed #3 | Seed #4 | Seed #5 | Mean |
|---|---|---|---|---|---|---|
| Pearson | 9834 | 0.9697 | 0.9857 | 0.9834 | 0.9737 | 0.9702 |
| Spearman | 0.9839 | 9?9722 | 0.9850 | 0.9874 | 0.9765 | 0.9810 |
| $R^2$ | 0.9605 | 0.9393 | 0.9641 | 0.9466 | 0.9464 | 0.9514 |
| RMSE | 0.0613 | 0.0798 | 0.0599 | 0.0664 | 0.0749 | 0.0685 |
| Convergence | 81/81 | 81/81 | 80/81 | 81/81 | 81/81 | 99.75 % |

Table 43: Details results for the quadruplet of drugs with ComboFM in Scenario 2.

### 8.5.3 Quintuplet

**Scenario 11**

| | Seed #1 | Seed #2 | Seed #3 | Seed #4 | Seed #5 | Mean |
|---|---|---|---|---|---|---|
| Pearson | 0.9537 | 0.9006 | 0.8596 | 0.9986 | 0.9181 | 0.9261 |
| Spearman | 0.9418 | 0.8855 | 0.9023 | 0.9988 | 0.9092 | 0.9275 |
| $R^2$ | 0.9048 | 0.7992 | 0.7436 | 0.9959 | 0.8301 | 0.7123 |
| RMSE | 0.0819 | 0.1219 | 0.1439 | 0.0175 | 0.1102 | 0.0823 |
| Convergence | 238/243 | 243/234 | 241/243 | 243/243 | 240/243 | 99.18% |

Table 44: Details results for the quintuplet of drugs with ComboFM in Scenario 11.

**Scenario 8**

| | Seed #1 | Seed #2 | Seed #3 | Seed #4 | Seed #5 | Mean |
|---|---|---|---|---|---|---|
| Pearson | 0.9672 | 0.9938 | 0.9356 | 0.9340 | 0.9316 | 0.9524 |
| Spearman | 0.9683 | 0.9938 | 0.9315 | 0.9292 | 0.9310 | 0.9508 |
| $R^2$ | 0.8646 | 0.9829 | 0.8409 | 0.8450 | 0.8193 | 0.8705 |
| RMSE | 0.1040 | 0.0366 | 0.1051 | 0.1042 | 0.1120 | 0.0924 |
| Convergence | 243/243 | 243/234 | 243/243 | 243/243 | 243/243 | 100% |

Table 45: Details results for the quintuplet of drugs with ComboFM in Scenario 11.

# References

[1] Nci-almanac. URL https://wiki.nci.nih.gov/display/NCIDTPdata/NCI-ALMANAC.

[2] Qu'est-ce que le cancer? URL https://www.cancer.be/le-cancer/quest-ce-que-le-cancer. (accessed: 12.04.2021).

[3] Global cancer statistics 2020. URL https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21660. (accessed: 20.05.2021).

[4] Cancer-causes. URL https://ressourcessante.salutbonjour.ca/condition/getcondition/cancer. (accessed: 20.05.2021).

[5] Combinatorial explosion. URL https://en.wikipedia.org/wiki/Combinatorial_explosion. (accessed: 04.05.2021).

[6] Factorization machines. URL https://www.jefkine.com/recsys/2017/03/27/factorization-machines/. (accessed: 10.04.2021).

[7] Factorization machine. URL http://ethen8181.github.io/machine-learning/recsys/factorization_machine/factorization_machine.html. (accessed: 31.03.2021).

[8] Loewe additivity model. URL https://horizondiscovery.com/en/screening/cell-panel-screening/chalice-bioinformatics-software/chalice-analyzer/chalice-combination-models/data/loewe-additivity.

[9] R squared. URL http://www.wardsystems.com/manuals/neuroshell2/index.html?rsquared.htm. (accessed: 31.03.2021).

[10] Les enjeux de santé publique. URL https://solidarites-sante.gouv.fr/IMG/pdf/enjeuxsantepub.pdf. (accessed: 12.04.2021).

[11] URL https://synergyfinder.fimm.fi/synergy/synfin_docs/. (accessed: 14.04.2021).

[12] Why is cancer so hard to treat? URL https://gicancer.org.au/news/faq/cancer-hard-treat/.

[13] Hyperparamètre. URL https://fr.wikipedia.org/wiki/Hyperparamètre. (accessed: 25.05.2021).

[14] Nested versus non-nested cross-validation. URL https://scikit-learn.org/stable/auto_examples/model_selection/plot_nested_cross_validation_iris.html. (accessed: 25.05.2021).

[15] C. I. Bliss. The toxicity of poisons applied jointly. URL https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1744-7348.1939.tb06990.x. (accessed: 27.05.2021).

[16] M. Blondel, A. Fujino, N. Ueda, and M. Ishihata. Higher-order factorization machines. *Advances in Neural Information Processing Systems*, (Nips):3359–3367, 2016. ISSN 10495258.

[17] S. M. Bonanome, Andrea; Grundy. The New England Journal of Medicine Downloaded from nejm.org at MOUNT SINAI SCHOOL OF MEDICINE on October 26, 2014. For personal use only. No other uses without permission. From the NEJM Archive. Copyright © 2010 Massachusetts Medical Society. All rights. *The New England Journal of Medicine*, pages 1244–1248, 1988.

[18] K. C. Bulusu, R. Guha, D. J. Mason, R. P. Lewis, E. Muratov, Y. Kalantar Motamedi, M. Cokol, and A. Bender. Modelling of compound combination effects and applications to efficacy and toxicity: State-of-the-art, challenges and perspectives. *Drug Discovery Today*, 21(2):225–238, 2016. ISSN 18785832. doi: 10.1016/j.drudis.2015.09.003. URL `http://dx.doi.org/10.1016/j.drudis.2015.09.003`.

[19] D. C. Elton, Z. Boukouvalas, M. S. Butrico, M. D. Fuge, and P. W. Chung. Applying machine learning techniques to predict the properties of energetic materials. *Scientific Reports*, 8(1): 1–12, 2018. ISSN 20452322. doi: 10.1038/s41598-018-27344-x.

[20] R. P. Gale. Combination cancer therapy. URL `https://www.msdmanuals.com/home/cancer/prevention-and-treatment-of-cancer/combination-cancer-therapy`. (accessed: 04.05.2021).

[21] P. Geurts. An introduction to machine learning. URL `https://people.montefiore.uliege.be/lwh/AIA/intro-ml-2017-light.pdf`. (accessed: 31.03.2021).

[22] P. Geurts and L. Wehenkel. Bias/variance trade-off, model assessment and model selection - introduction to machine learning. URL `https://people.montefiore.uliege.be/lwh/AIA/IML___Model_evaluation.pdf`. (accessed: 31.03.2021).

[23] S. L. Holbeck, R. Camalier, J. A. Crowell, J. P. Govindharajulu, M. Hollingshead, L. W. Anderson, E. Polley, L. Rubinstein, A. Srivastava, D. Wilsker, J. M. Collins, and J. H. Doroshow. The National Cancer Institute ALMANAC: A comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Research*, 77(13):3564–3576, 2017. ISSN 15387445. doi: 10.1158/0008-5472.CAN-17-0489.

[24] A. Ianevski, A. K. Giri, P. Gautam, A. Kononov, S. Potdar, J. Saarela, K. Wennerberg, and T. Aittokallio. Prediction of drug combination effects with a minimal set of experiments. *Nature Machine Intelligence*, 1(12):568–577, 2019. ISSN 2522-5839. doi: 10.1038/s42256-019-0122-4. URL `http://dx.doi.org/10.1038/s42256-019-0122-4`.

[25] J. D. Janizek, S. Celik, and S. I. Lee. Explainable machine learning prediction of synergistic drug combinations for precision cancer medicine. *bioRxiv*, pages 0–4, 2018. doi: 10.1101/331769.

[26] Jefkine. Factorization machines. URL `https://www.jefkine.com/recsys/2017/03/27/factorization-machines/`. (accessed: 31.03.2021).

[27] H. Julkunen, A. Cichonska, P. Gautam, S. Szedmak, J. Douat, T. Pahikkala, T. Aittokallio, and J. Rousu. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nature Communications*, 11(1), 2020. ISSN 20411723. doi: 10.1038/s41467-020-19950-z. URL `http://dx.doi.org/10.1038/s41467-020-19950-z`.

[28] A. N. Mikhail Trofimov. tffm: Tensorflow implementation of an arbitrary order factorization machine. `https://github.com/geffy/tffm`, 2016.

[29] K. Preuer, R. P. Lewis, S. Hochreiter, A. Bender, K. C. Bulusu, and G. Klambauer. DeepSynergy: Predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics*, 34(9):1538–1546, 2018. ISSN 14602059. doi: 10.1093/bioinformatics/btx806.

[30] L. Rampášek, D. Hidru, P. Smirnov, B. Haibe-Kains, and A. Goldenberg. Dr.VAE: Improving drug response prediction via modeling of drug perturbation effects. *Bioinformatics*, 35(19): 3743–3751, 2019. ISSN 14602059. doi: 10.1093/bioinformatics/btz158.

[31] E. Tekin, C. White, T. M. Kang, N. Singh, M. Cruz-Loya, R. Damoiseaux, V. M. Savage, and P. J. Yeh. Prevalence and patterns of higher-order drug interactions in Escherichia coli. *npj Systems Biology and Applications*, 4(1), 2018. ISSN 20567189. doi: 10.1038/s41540-018-0069-9. URL `http://dx.doi.org/10.1038/s41540-018-0069-9`.

[32] P. M. Vanhoutte, E. A. Barnard, G. J. Cosmides, P. P. Humphrey, M. Spedding, and T. Godfraind. I. International Union of Pharmacology committee on receptor nomenclature and drug classification. *Pharmacological Reviews*, 46(2):111–116, 1994. ISSN 00316997. doi: 10.1124/pr.55.4.4.and.

[33] A. Zimmer, I. Katzir, E. Dekel, A. E. Mayo, and U. Alon. Prediction of multidimensional drug dose responses based on measurements of drug pairs. *Proceedings of the National Academy of Sciences of the United States of America*, 113(37):10442–10447, 2016. ISSN 10916490. doi: 10.1073/pnas.1606301113.

[34] A. Zimmer, A. Tendler, I. Katzir, A. Mayo, and U. Alon. Prediction of drug cocktail effects when the number of measurements is limited. *PLoS Biology*, 15(10):1–16, 2017. ISSN 15457885. doi: 10.1371/journal.pbio.2002518.