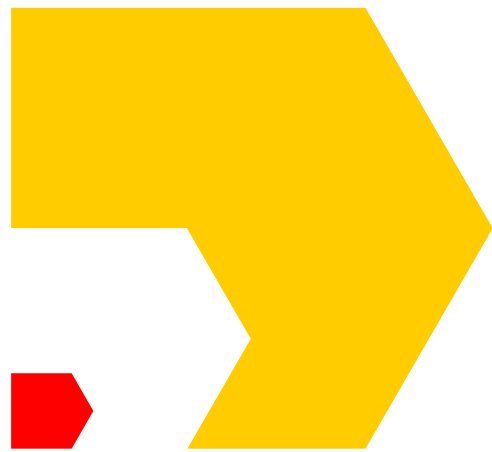


# Оптимизация расхода ферросплавов при производстве стали



Yandex  
Data Factory

# Некоторые понятия

**Сталь** – сплав на основе железа с заданным содержанием примесей. Это то, что надо получить.

**Чугун** - сплав на основе железа с известным содержанием примесей и с большим содержанием углерода. Это  $\frac{3}{4}$  того, из чего делают сталь.

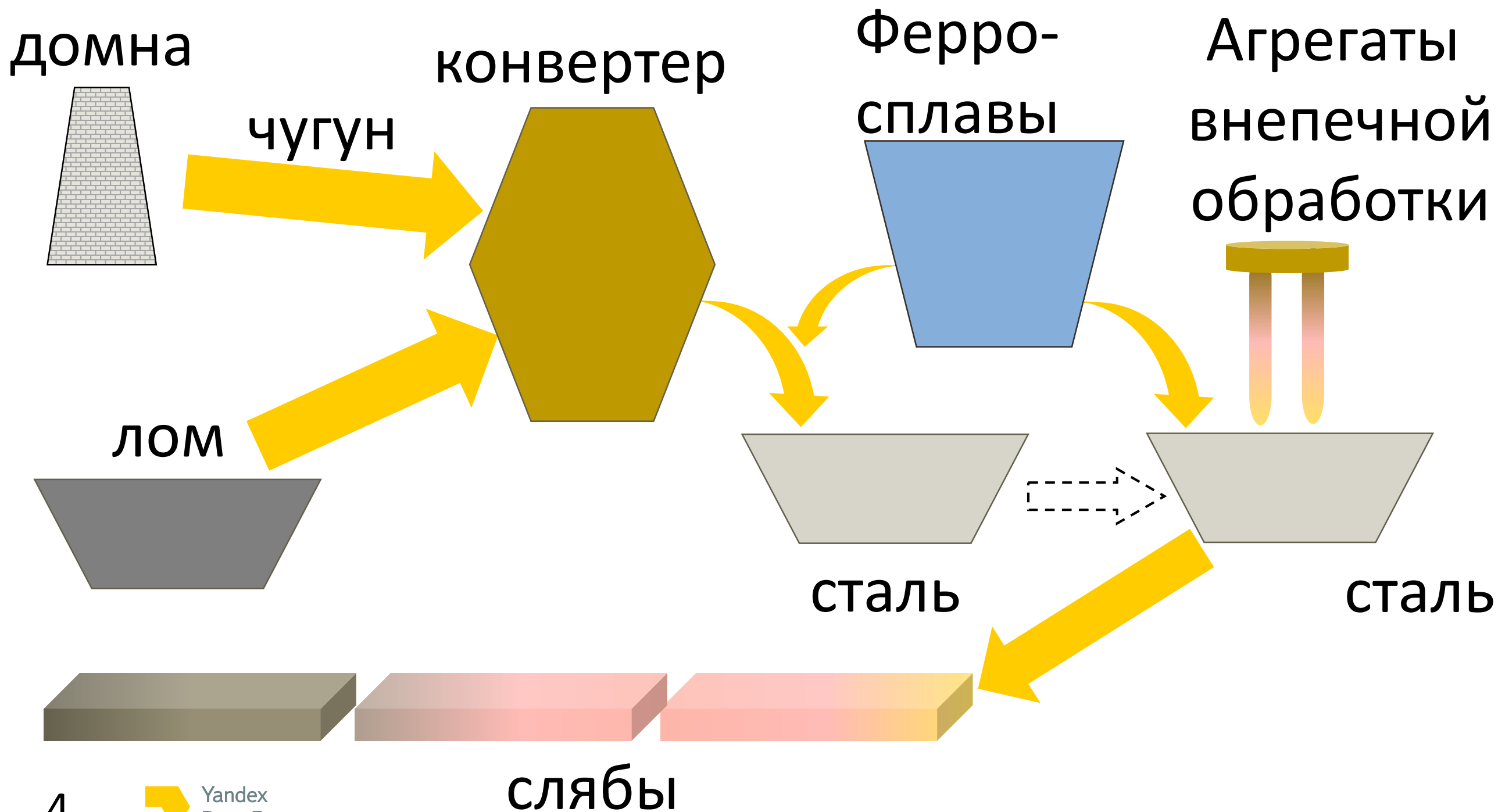
**Лом** – металлические отходы с плохо известным содержанием примесей. Это примерно  $\frac{1}{4}$  того, из чего делают сталь.

**Ферросплав** – сплав на основе железа с известным содержанием примесей. Это то, что добавляют в сталь для получения нужного состава.

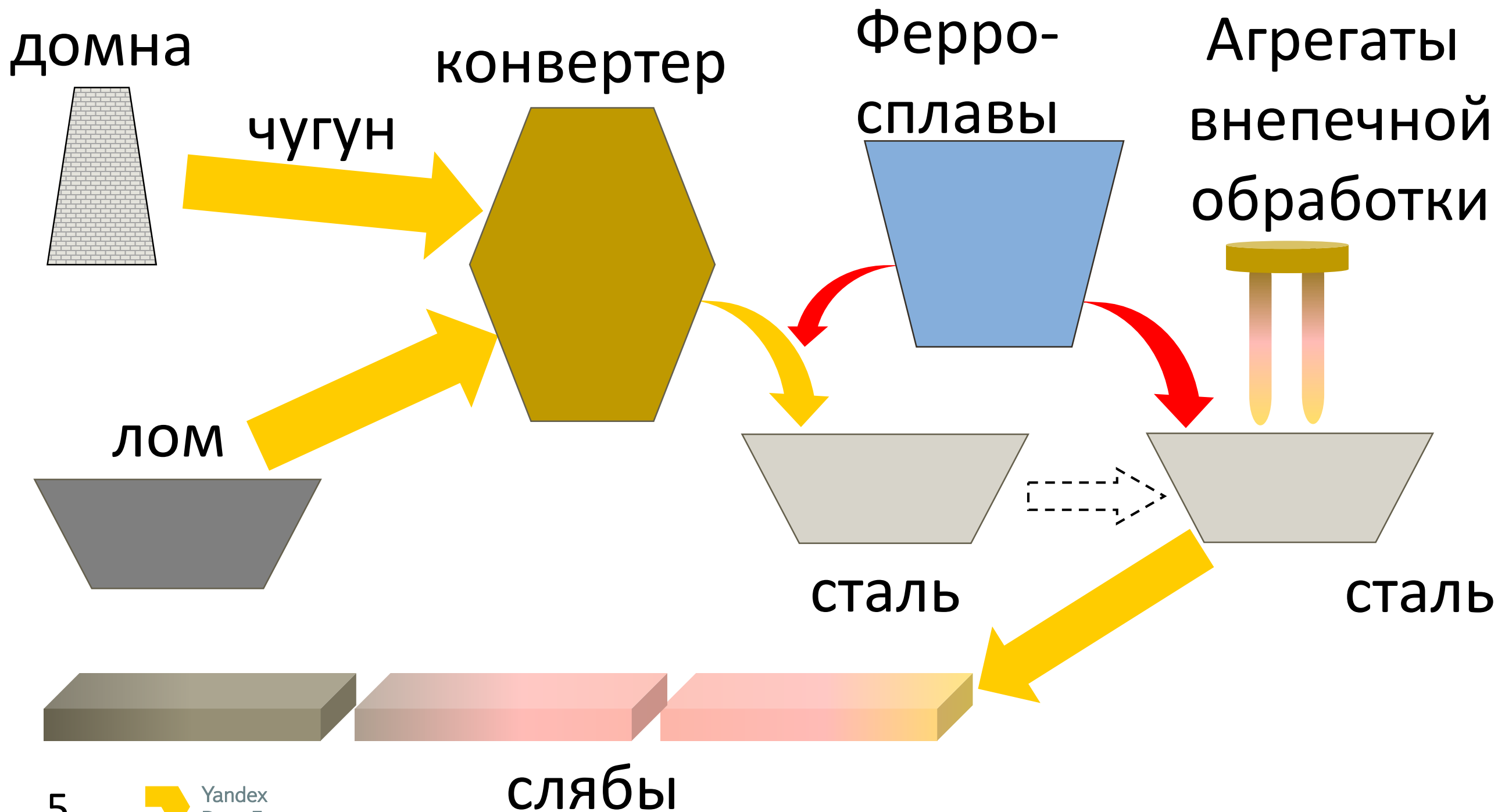
# Вот как это выглядит



# Процесс выплавки стали



# Процесс выплавки стали



# Цели и задачи проекта

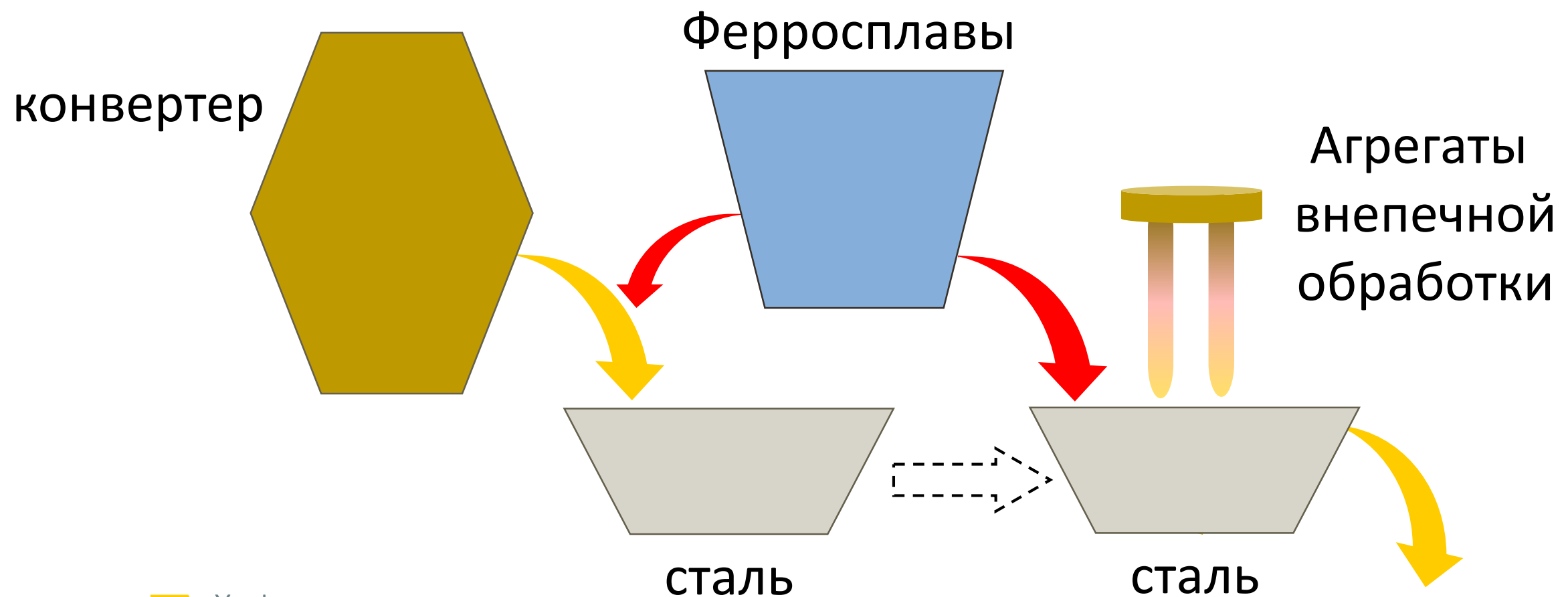
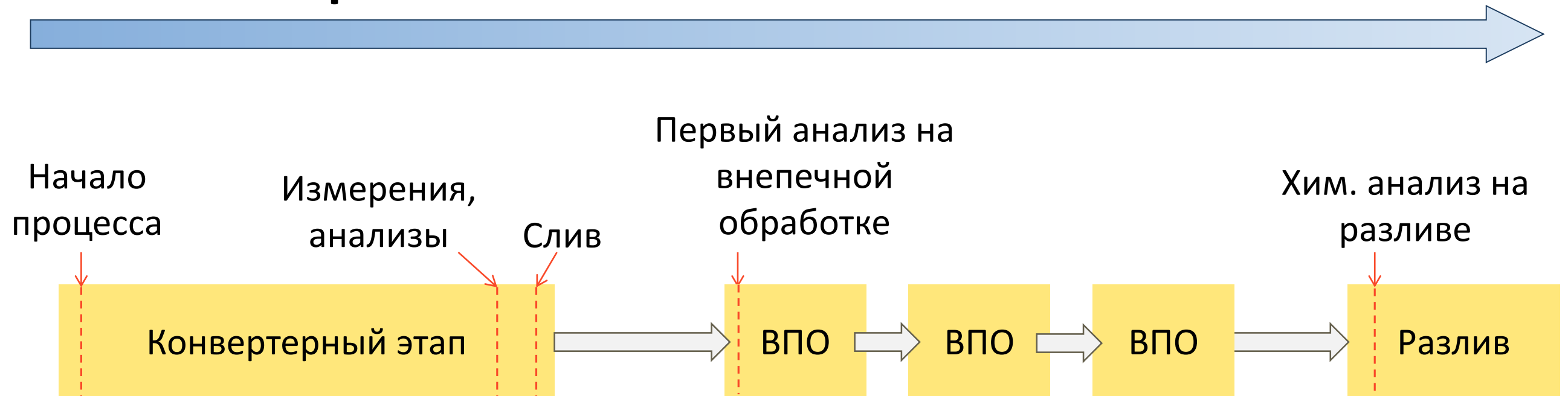
- › **Цель:** Оценить пригодность технологий и методик Яндекса в области Big Data и Machine Learning для решения задач ММК в части оптимизации использования добавок при производстве стали
- › **Задача:** Разработать сервис, который будет выдавать рекомендации по расходу использования ферросплавов при производстве стали на конвертерном этапе и этапе внепечной обработки



# Использование результатов

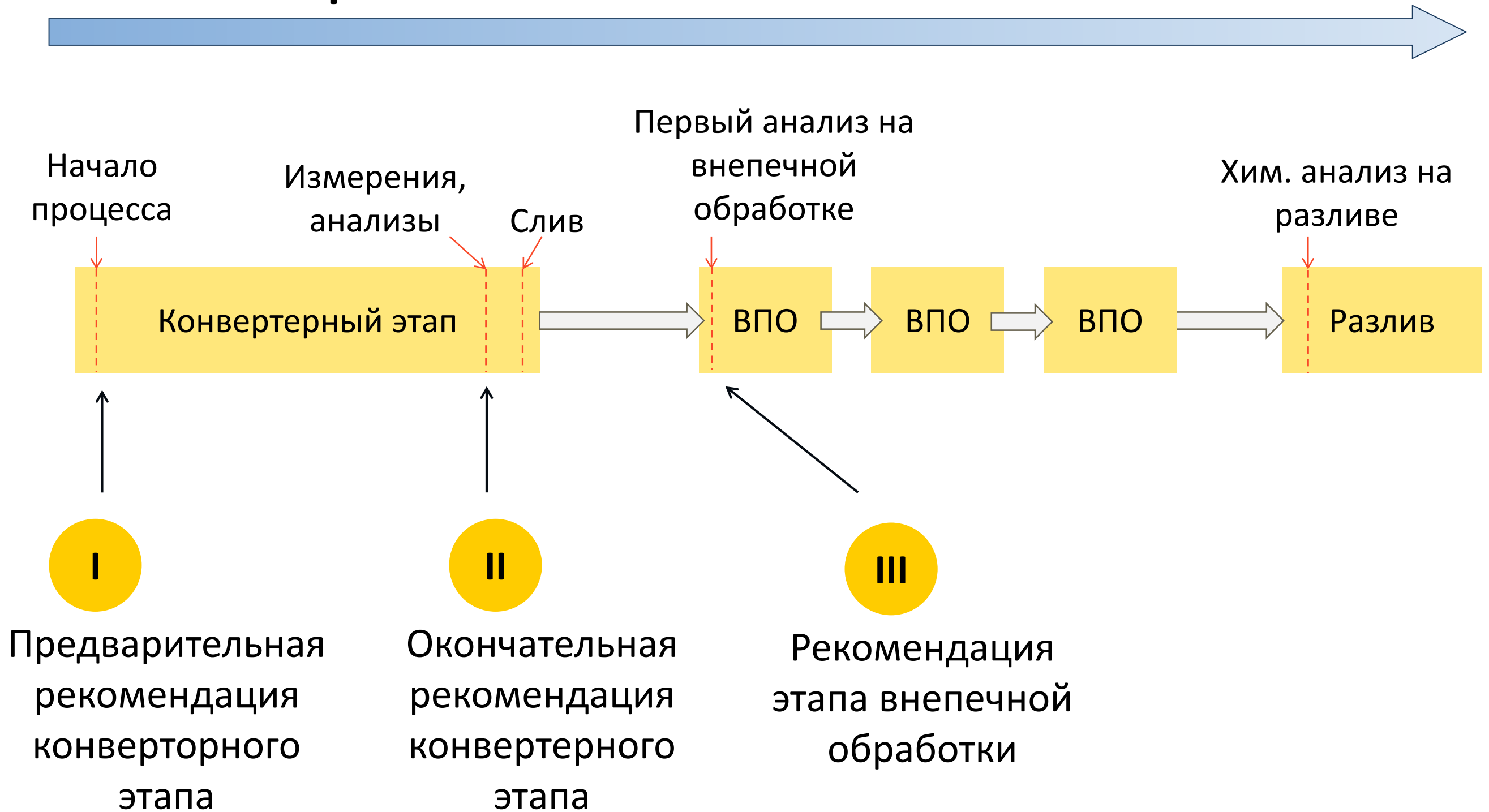


# Линия времени





# Линия времени



# О задаче

## Известны:

- Все измерения, совершенные к моменту запроса
- Набор ферросплавов
- Спецификации стали

## Надо определить:

- Количество каждого из ферросплавов

## Условия:

- Надо удовлетворить спецификациям
- Надо оптимизировать дополнительную цель (например, минимизировать цену)

# Определения

$\vec{x}$  - массы вносимых ферросплавов

$\vec{y}$  - химический состав стали

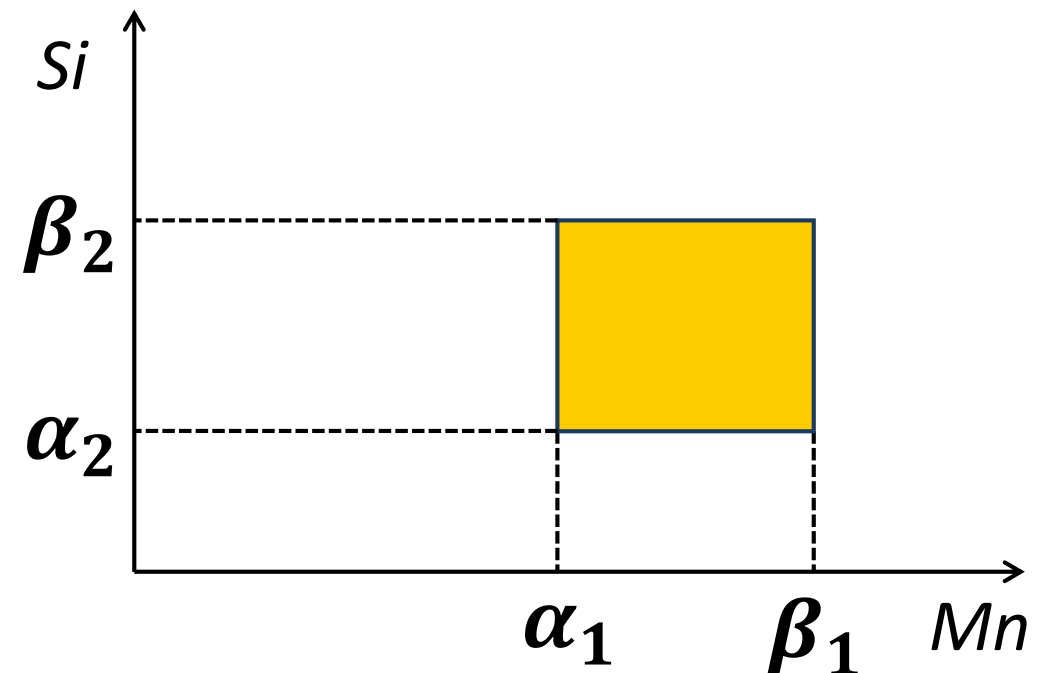
$\vec{a}$  - параметры плавки, не вошедшие в  $\vec{x}$

$\vec{y}(\vec{x}, \vec{a})$  – регрессионная зависимость химического состава стали от условий плавки

# Ограничения - Требования по химии

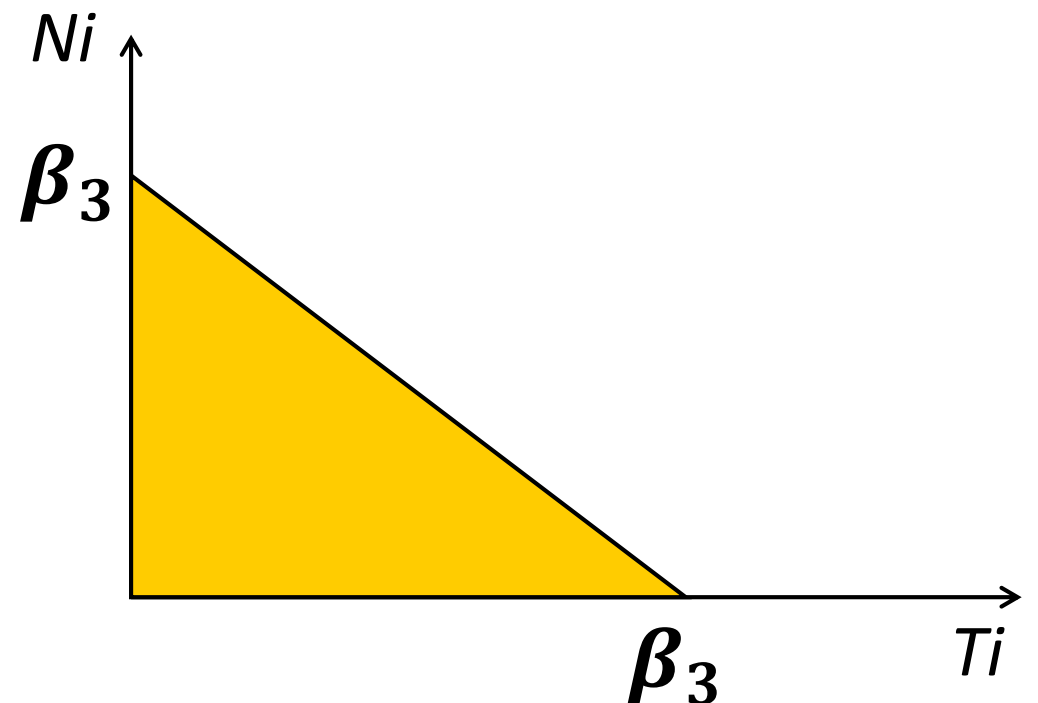
По каждому из нескольких химических элементов заданы допустимые границы содержания:

$$\alpha_i \leq y_i \leq \beta_i$$



Иногда ограничения задаются для комбинации элементов:

$$y_i + y_j \leq \beta_k$$



# Цели оптимизации

**Один из вариантов :**

- Минимальные затраты
- Максимальная близость к «идеальному» химическому составу



# Цели оптимизации

## ➤ Затраты:

$$\vec{x} \cdot \vec{p} \rightarrow \min$$

$\vec{x}$  - количества вносимых ферросплавов

$\vec{p}$  - стоимости ферросплавов

## ➤ Близость к заданному хим.составу

$$\|\vec{y}(\vec{x}, \vec{a}) - \vec{y}^*\| \rightarrow \min$$

$\vec{y}(\vec{x})$  – химический состав, полученный регрессией по количествам ферросплавов и другим параметрам  $\vec{a}$

$\vec{y}^*$  - целевой химический состав

# Постановка задачи. Особенности

- Высокая неопределенность важных факторов, например, состава лома



- Как следствие, при заданных параметрах и точно определенных количествах всех ферросплавов хим. состав стали не гарантирован



- Как следствие, можно говорить лишь о вероятностном предсказании

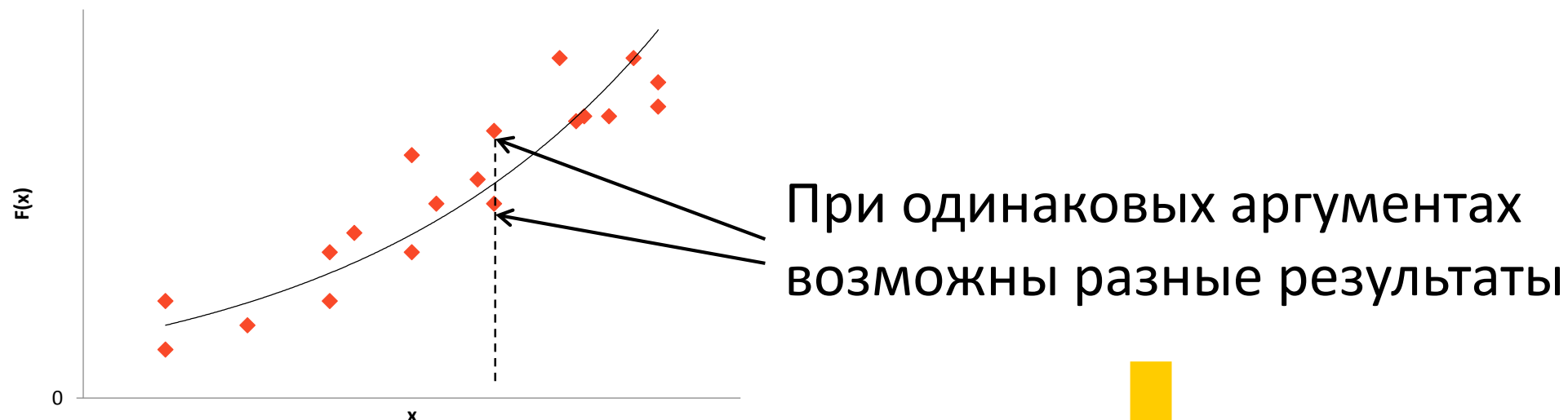
# Почему бы не построить регрессию $\vec{x}(\vec{y}, \vec{a})$ ? Это сразу дало бы ответ.

- Разные комбинации вносимых материалов могут приводить к одному и тому же химическому составу

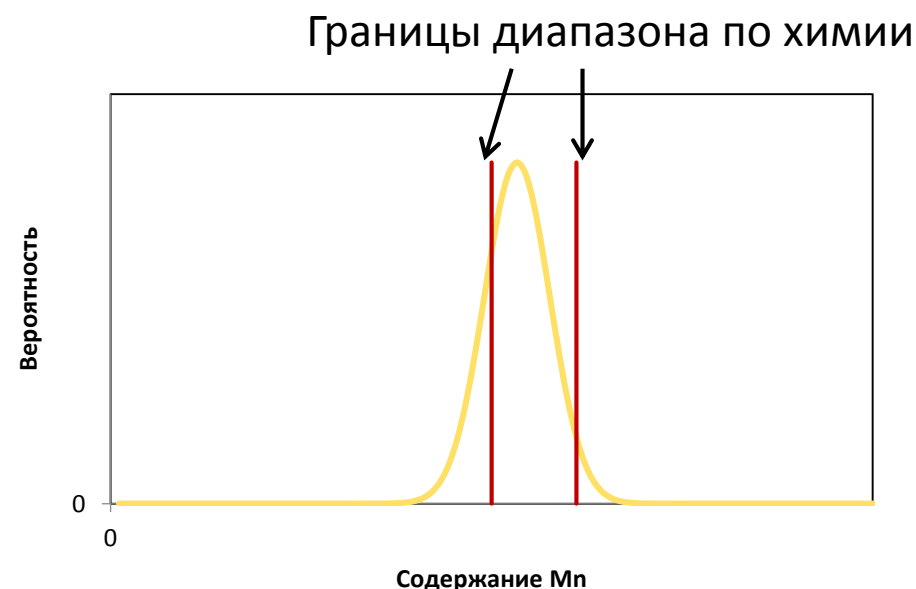


- Регрессионной зависимости  $\vec{x}(\vec{y}, \vec{a})$  не существует даже при наличии полной и точной информации об  $\vec{y}$  и  $\vec{a}$

# Работа с неопределенностью предсказания $\vec{y}(\vec{x}, \vec{a})$



Результат предсказания – не точное значение получаемого химического состава, а распределение его вероятности:



Интерес представляет та часть  
распределения, которая попадает  
в границы диапазона

# Оптимизация. Пример ограничений.

Ограничения – условия уверенного попадания содержания элементов в заданные диапазоны по химии:

$$\int_{y_{min}}^{y_{max}} \rho(y) dy \geq \alpha$$

$y$  – хим.состав;  $y_{min}, y_{max}$  – границы диапазона;

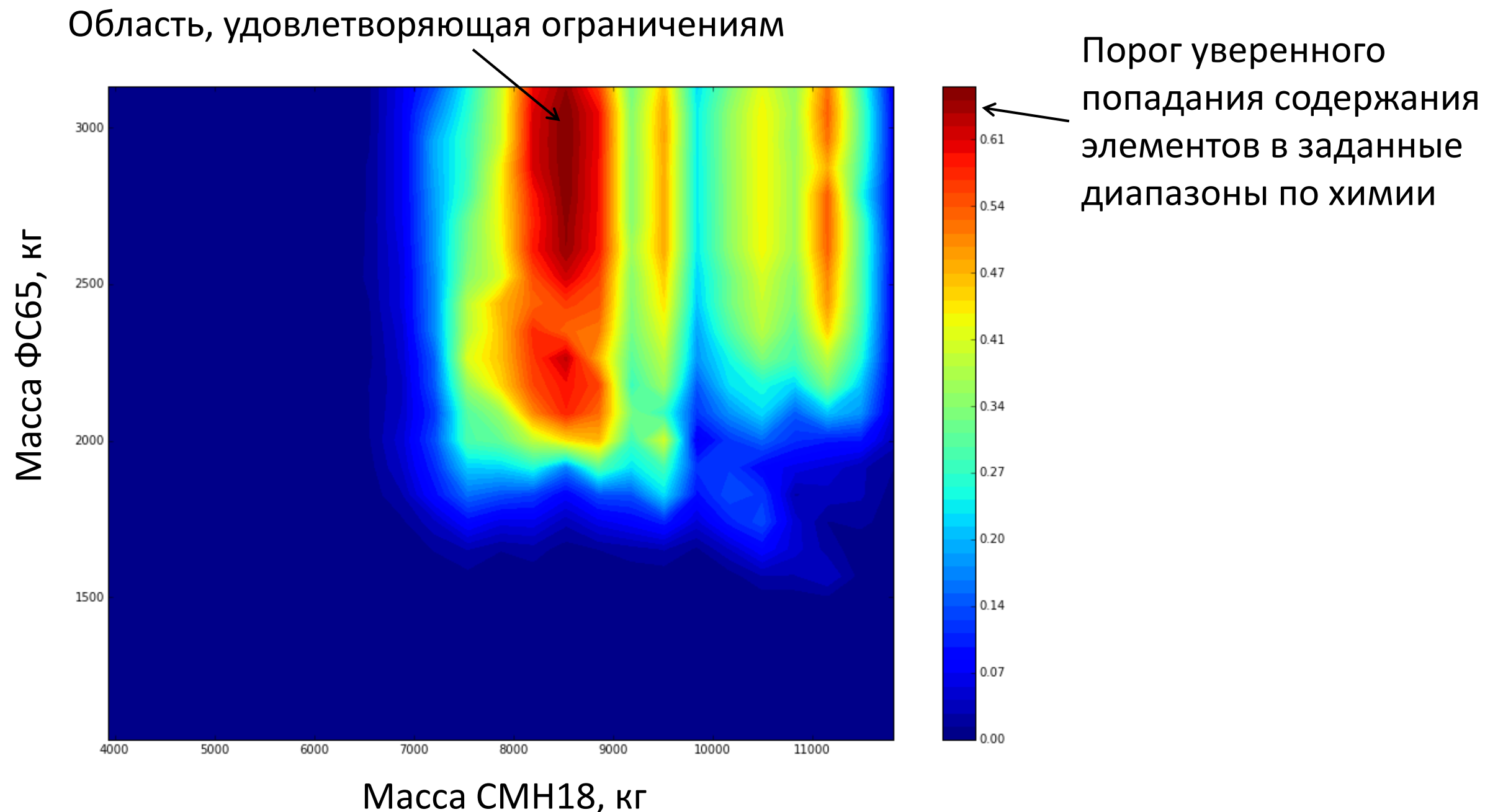
$\rho(y)$  – распределение плотности вероятности получения хим.состава  $y$ ;

$\alpha$  – порог уверенного попадания.



# Оптимизация.

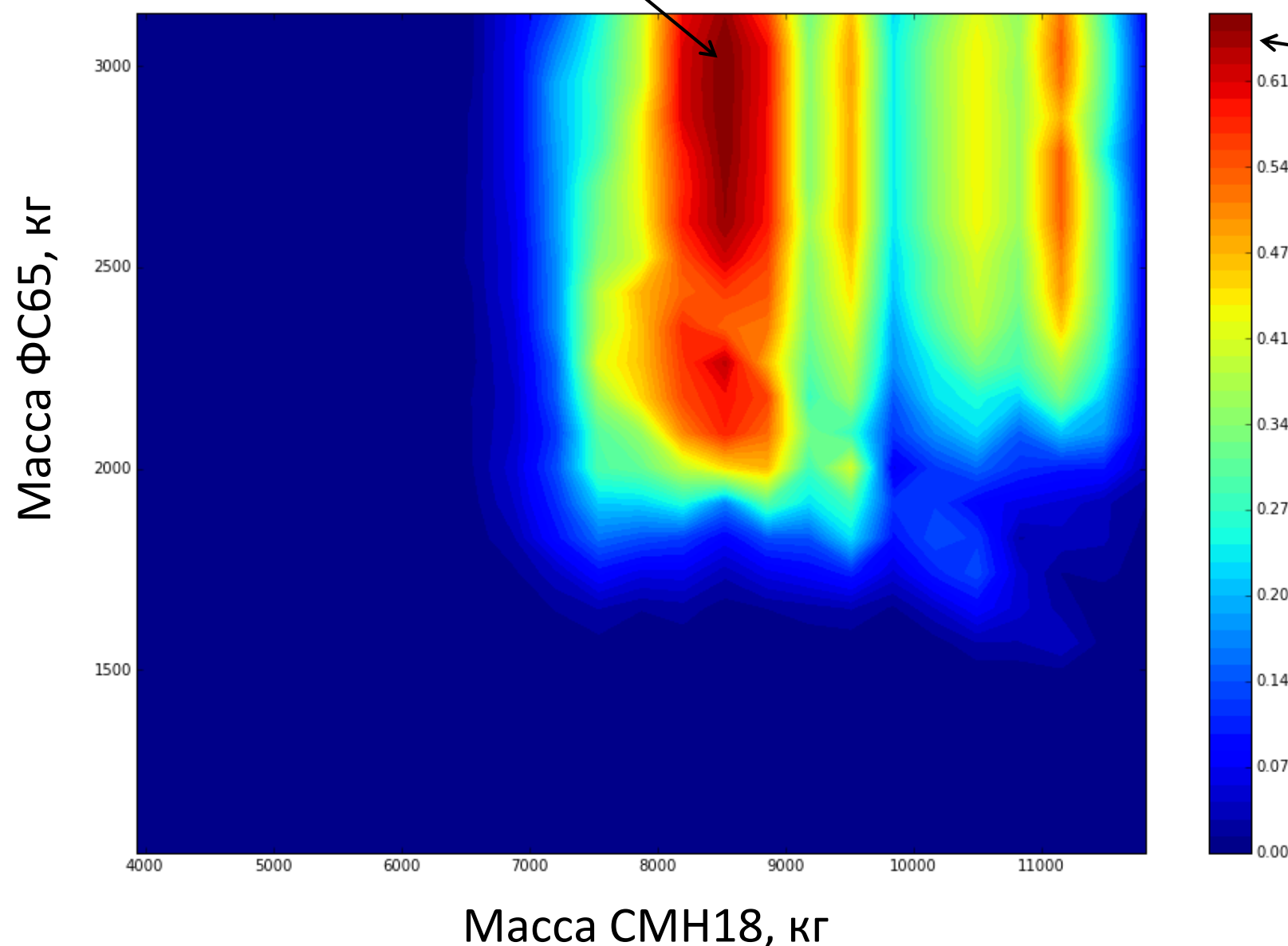
## Иллюстрация ограничений.



# Оптимизация.

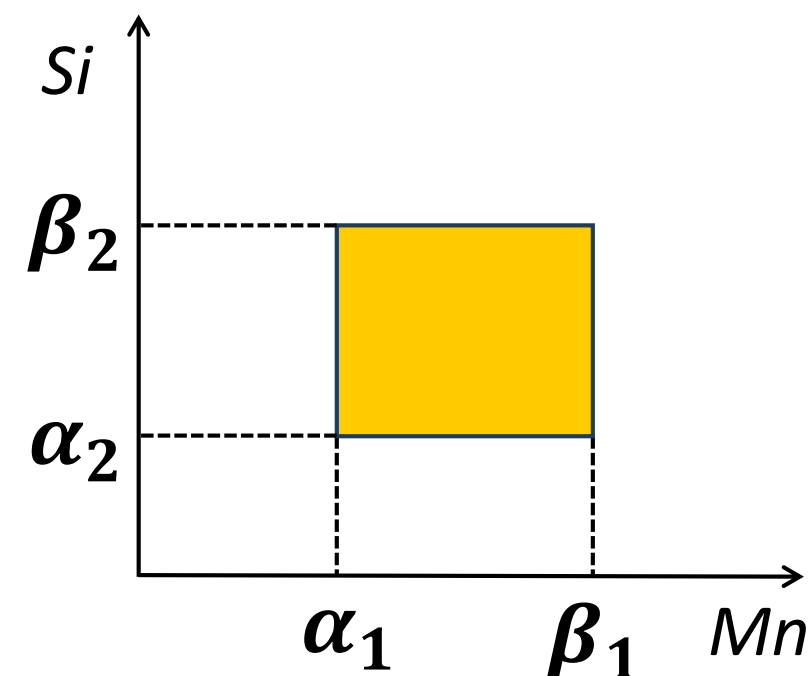
## Иллюстрация ограничений.

Область, удовлетворяющая ограничениям



Порог уверенного попадания содержания элементов в заданные диапазоны по химии

Это соответствует в определенном смысле области в пространстве, где заданы ограничения

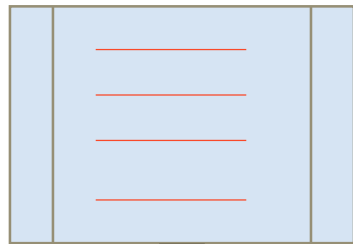


# Подход к решению. Общая идея.

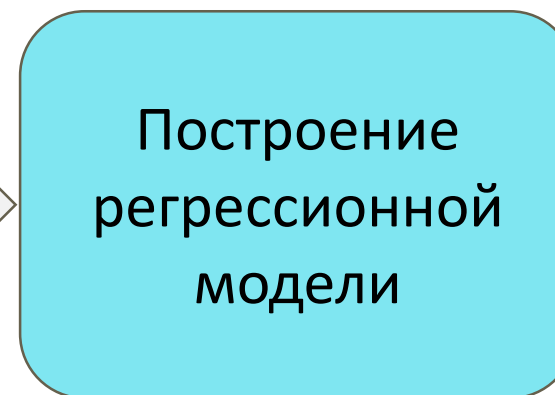
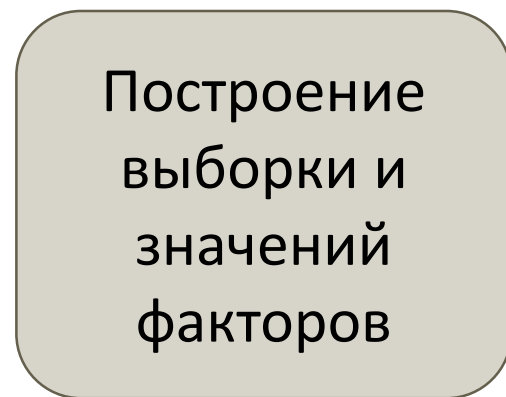
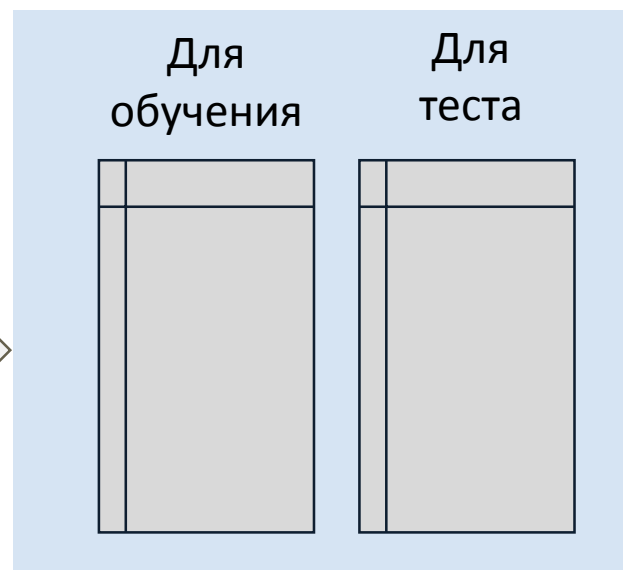
- Строим модель, которая предсказывает хим. состав стали.
- Оцениваем уровень погрешности модели.
- При помощи модели и оценки ее погрешности строим функцию вероятности попадания в диапазоны по химии; аргументы – количества добавляемых ферросплавов.
- Оптимизируем:
  - Вероятность попадания в диапазоны по химии
  - Дополнительные критерии: стоимость или близость к заданному химическому составу

# Подход к решению. Построение регрессии.

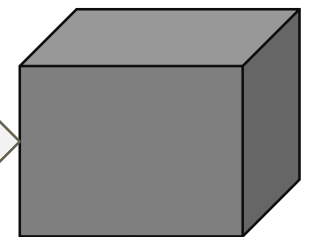
Список факторов и целей



Значения факторов для всех плавков.  
Плавки разбиты на две группы:

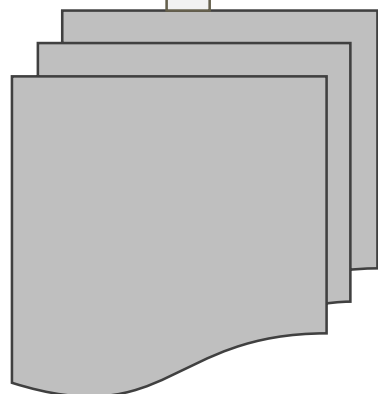


Регрессионная модель

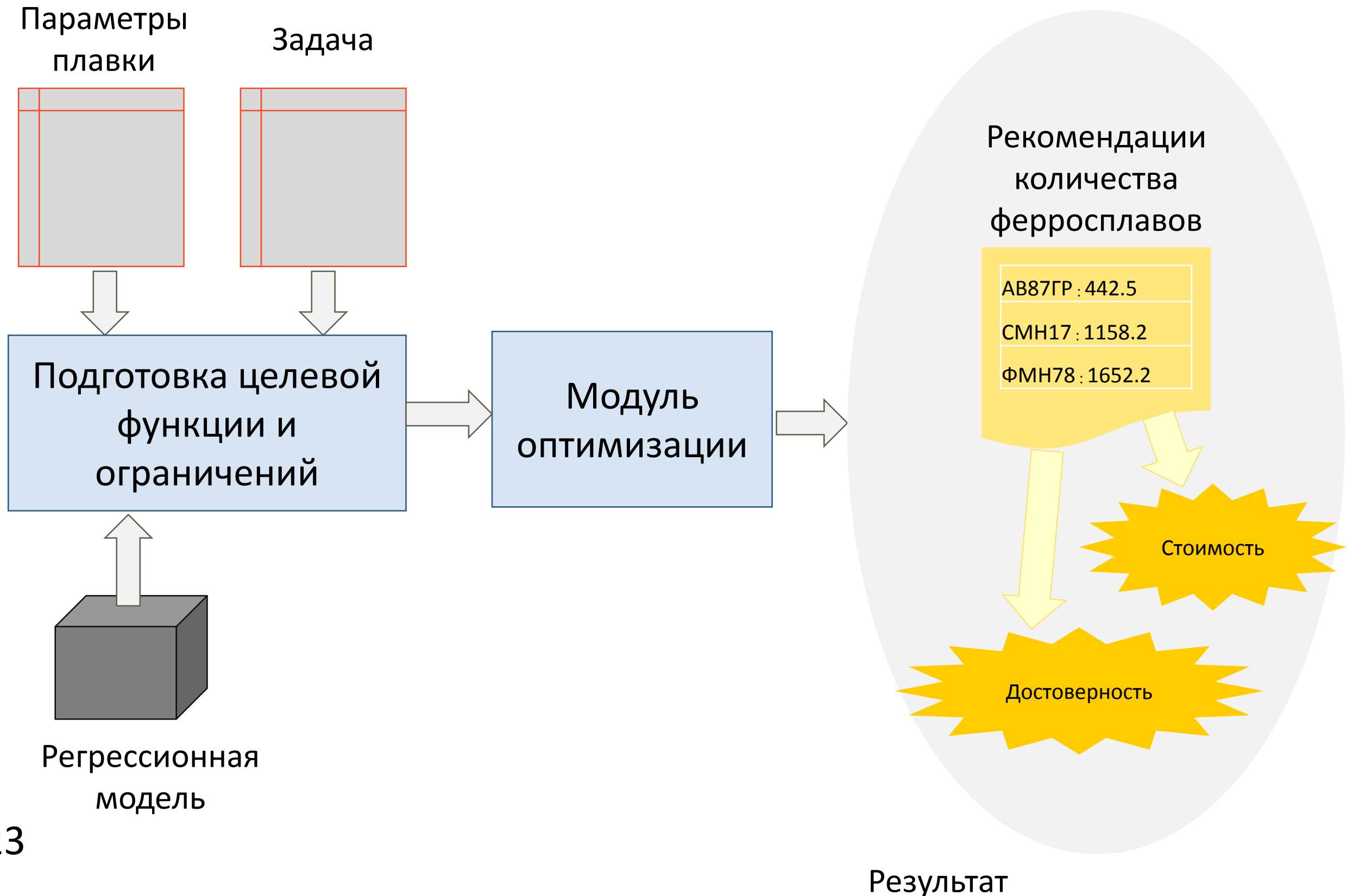


Регрессионная модель – «чёрный ящик».  
Зависимость целевых показателей от факторов  
определена, но трудноинтерпретируема.

Протоколы плавок



# Подход к решению. Оптимизация.





# Общая схема

## 1. Регрессия

- Объект – одна плавка
- Данные – разнотипные параметры плавки
- Целевое значение – хим. состав стали

## 2. Оптимизация

- Настраиваемые параметры – добавления ферросплавов
- Функционал качества – расстояние от результата регрессии до целевой точки
- Ограничение по вероятности попадания в заданный диапазон

# Данные для обучения регрессионной модели

- Данные – исторические, с 2011 по 2015 год
- Данные сырые, то есть в данных много ошибок
- Предполагается что технология выплавки за данный период, а также за период предсказания, не изменится
- Общее кол-во плавок (объектов) – 177273

# Построение признакового описания

- **Из каждой плавки извлекались :**
  - › Параметры плавки
  - › Параметры технологического процесса
  - › Добавления ферросплавов
- **Данные фильтровались:**
  - › исходя из физических соображений
  - › по остаткам регрессии

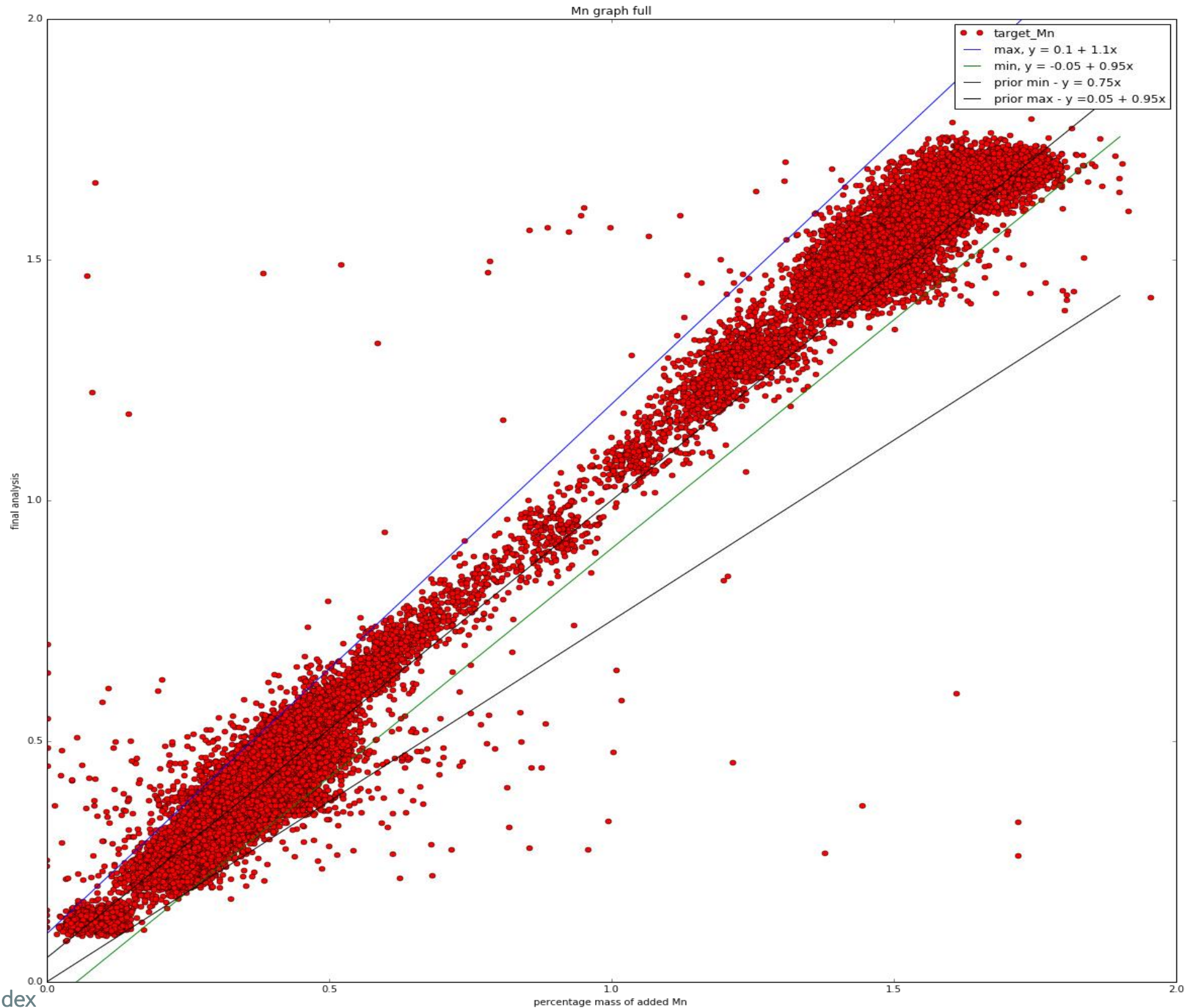
**Кол-во оставшихся плавок - 45000**

# Задача регрессии

- Хотим научиться моделировать процесс плавки
- Хотим уметь предсказывать погрешность
- Модель должна быть интерпретируемой
- Модель должна выдавать адекватный результат в областях, не покрытых историческими данными

# График Mn

Финальный анализ

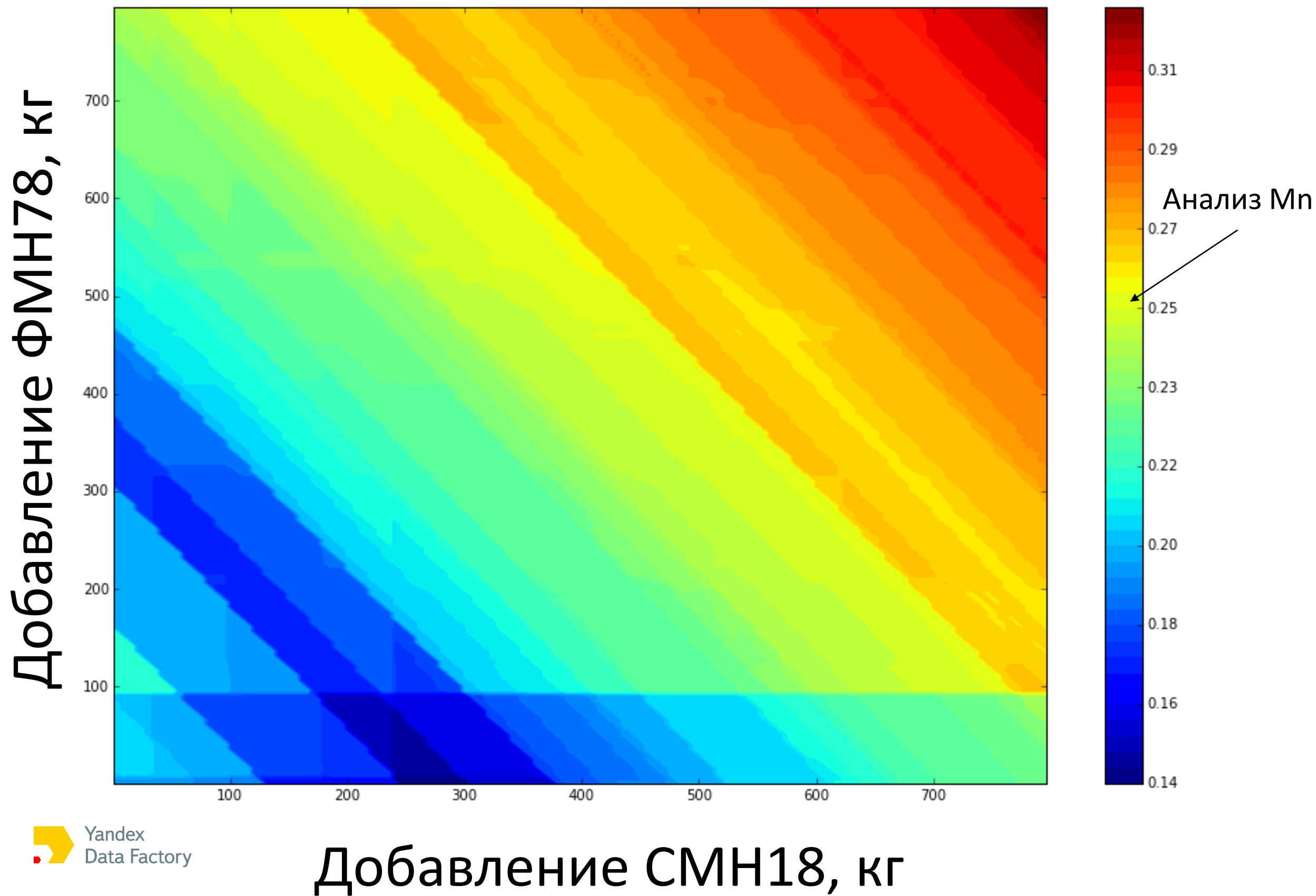




# Линейная регрессия

- Отдельная модель по каждому хим. элементу
- Простая модель
- Интерпретируемая модель
- Большие остатки

# Матрикснет



# Идея!

Совместим простую и  
интерпретируемую модель с  
матрикснетом

# Регрессия

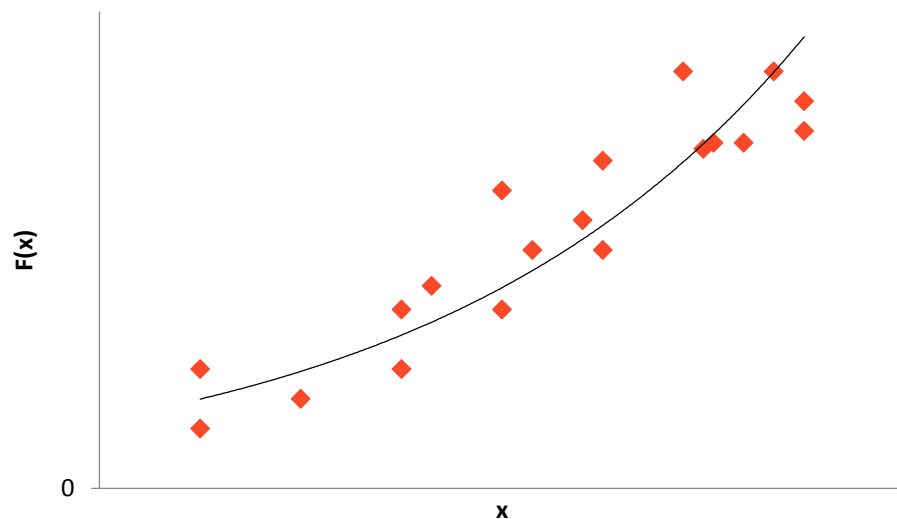
$\vec{x}$  — значения факторов

$y^k$  — цель (содержание элемента  $k$ )

$\{\vec{x}_j, y^k_j\}$  — исторические данные



$$y^k \approx F(\vec{x})$$



Предварительно рассматриваются наиболее важные факторы  $\vec{z}$  :

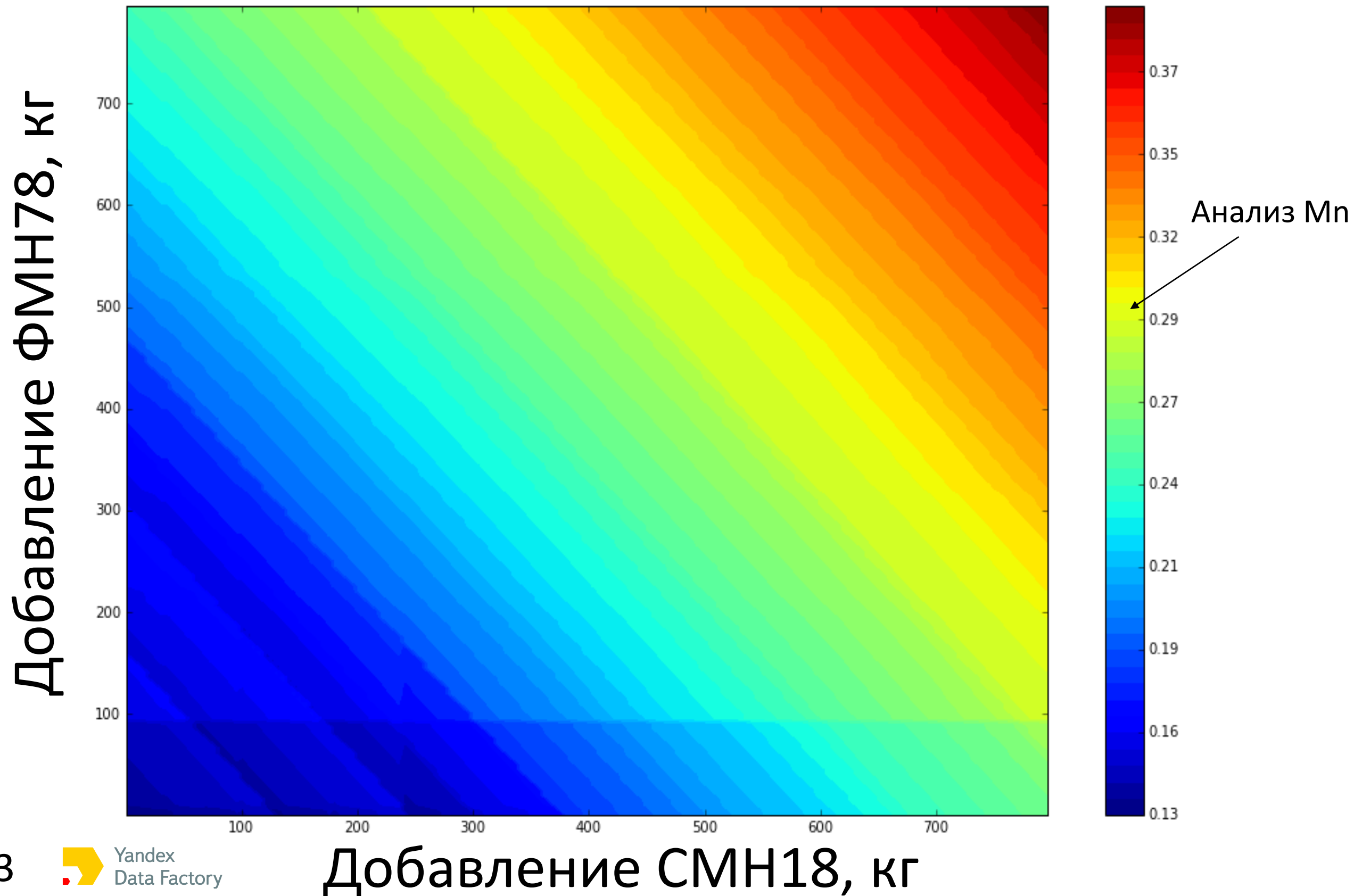
- содержание легирующего элемента в ферросплавах
- химия чугуна

2 этапа регрессии:

- Линейная зависимость от  $\vec{z}$
- Регрессия на остатки при помощи технологии Яндекса «Матрикснет»

$$y^k \approx L(\vec{z}) + M(\vec{x})$$

# Матрикснет + линейная регрессия



# Качество регрессии

элемент	SD	RMSE	RMSE/SD
C	0.0493	<b>0.0164</b>	0.3332
Mn	0.0444	<b>0.0272</b>	0.6117
P	0.0055	<b>0.0022</b>	0.3966
S	0.0062	<b>0.0030</b>	0.4783
Si	0.0287	<b>0.0167</b>	0.5807

# Проблемы регрессии

- **Не все признаки можно использовать**

**Добавления ферросплавов – плохие признаки:**

- › Названия ферросплавов меняются
- › Различные названия могут обозначать один и тот же ферросплав
- › Выбор некоторых ферросплавов обусловлен технологическим процессом

**Вместо добавлений ферросплавов – добавления элементов**

- **Не все элементы мы можем предсказывать**

# Плохой пример

Важность признаков для модели, предсказывающей углерод:

1. Масса добавленной серы — 11.18942894
3. Масса добавленного кремния — 8.734091967
4. Масса добавленного марганца — 8.67283899
- ...
11. Масса добавленного углерода — 2.059519198

**Признак, непосредственно влияющий на углерод  
находится на 11 месте!**

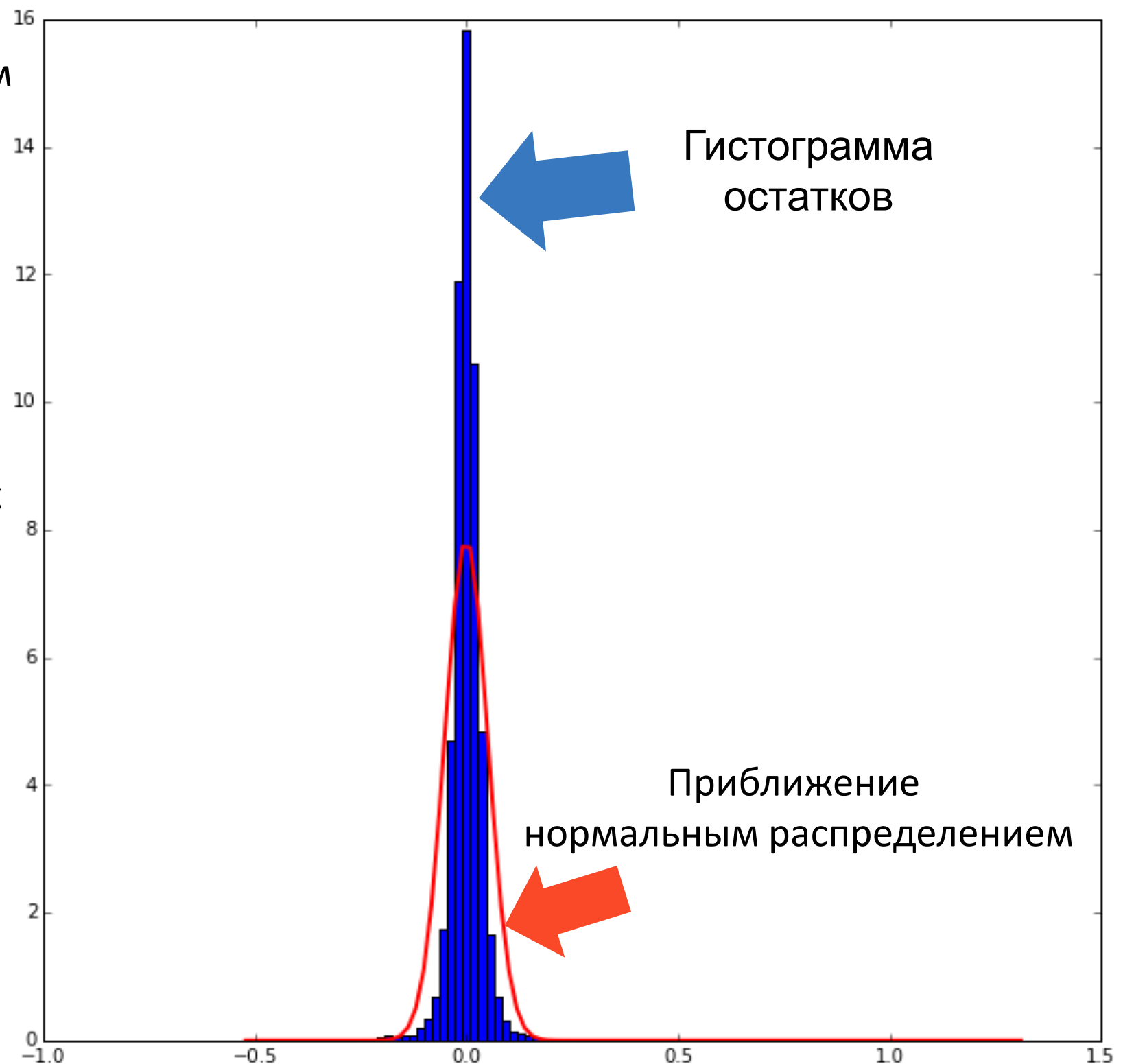


# Остатки регрессионной модели

- Остатки приближаются нормальным распределением
- Дисперсия рассчитывается по определенным подгруппам плавок (отдельно для каждой марки стали)
- Вероятность попадания считаем как интеграл по нормальному распределению:

$$P(y) = \int_a^b p(y'|y, \sigma) dy'$$

$p(y'|y, \sigma)$  - нормальное распределение



# Оптимизация

Исходная задача:  $f(y) = (y - y_G)^2 \rightarrow \min$

При ограничении  $P(y) = \int_a^b p(y'|y, \sigma) dy' \geq \textit{confidence}$

- $y$  - результат регрессии по рассматриваемому элементу
- $y_G$  - целевой анализ по заданному элементу
- $\sigma$  - дисперсия, рассчитанная по истории для заданной марки стали
- *confidence* - заданное значение вероятности попадания в диапазон  $[a, b]$

# Проблема

**$y$  – это результат матрикснета!**

# Алгоритм

**Data:** Параметры плавки

**Result:** Рекомендации ферросплавов

Генерируем сетку в пространстве ферросплавов;

**while** (*шаг сетки больше заданного значения*) **do**

**for** (*точка из сетки*) **do**

        считаем функционал;

**if** *функционал меньше текущего* **then**

            запоминаем точку;

**end**

**end**

    сдвигаем сетку;

    уменьшаем шаг;

**end**

# Функционал качества

$P(y)$  – вероятность попадания в диапазон  $[a, b]$

$$P(y) = \int_a^b p(y'|y, \sigma) dy'$$

Возможен случай, когда

$$\max_{y \in [a, b]} P(y) < \textit{confidence}.$$

Тогда ослабим ограничения:

$$P(y) \geq \textit{confidence}.$$



$$P(y) \geq \textit{confidence} \cdot \max_{y \in [a, b]} P(y)$$

# Функционал качества

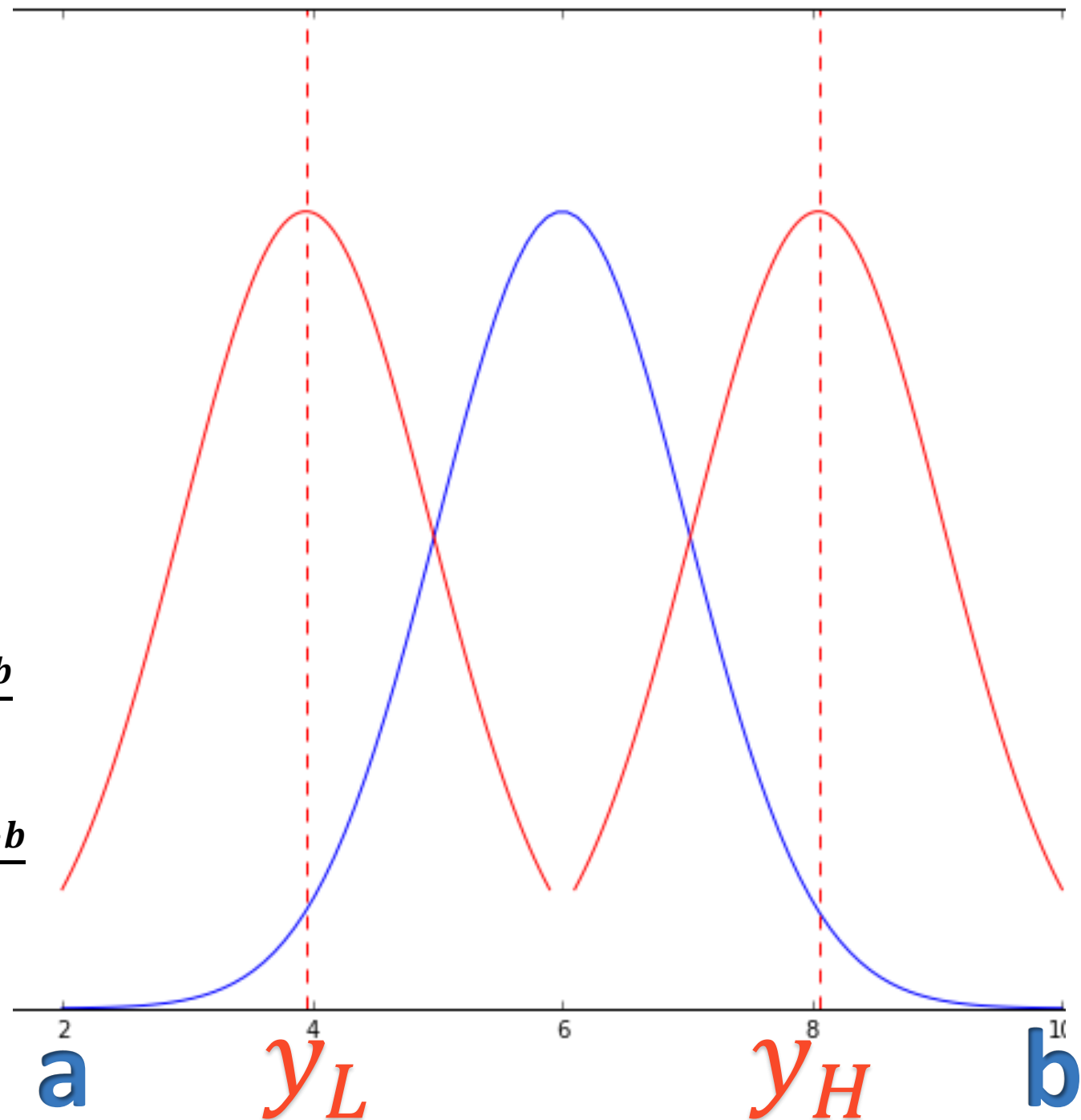
Перенесем ограничения в функционал:

$$f(y) = \omega_1(y - y_G)^2 + \\ \omega_2[y < y_L](y - y_L)^2 + \\ \omega_3[y > y_H](y - y_H)^2,$$

$\omega_{1,2,3}$  - веса компонент

$$y_L : P(y_L) = \text{confidence} \cdot P\left(\frac{a+b}{2}\right), y_L < \frac{a+b}{2}$$

$$y_H : P(y_H) = \text{confidence} \cdot P\left(\frac{a+b}{2}\right), y_H > \frac{a+b}{2}$$



# Результаты

- В большинстве случаев алгоритм выдает рекомендацию в течение 10 секунд
- По определенным группам (маркам) стали практически все рекомендации успешные
- Экономический эффект на этих марках всегда положительный

# Результаты

Кол-во плавок		Денежные затраты	
Принятых	Отвергнутых	В принятых плавках	В соответствии с историей
45	49	138469	147166

Использовано Mn		Использовано Si	
В принятых плавках	В соответствии с историей	В принятых плавках	В соответствии с историей
46689	49303	9024	11350

**Выигрыш по сравнению с историей:**

- По деньгам : 6.28%
- По Mn : 5.54%
- По Si : 25.78%



# Направления развития

## Постановка задачи:

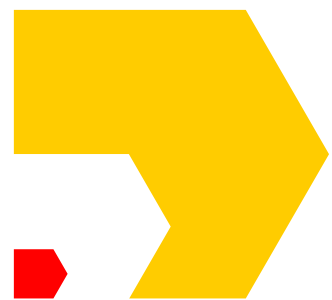
- › Одинаково ли важны все диапазоны химии?
- › Что делать с теми целями, на которые мы не влияем?
- › Как правильно задавать требования к концу конвертерного этапа?

## Модели регрессии:

- › Добавление в модель физически и химически обусловленных связей
- › Снижение погрешностей регрессии

## Качество оптимизации:

- › Как точнее оценить область поиска?
- › Ускорение алгоритмов



# Yandex Data Factory

Tel: +7 495 739-70-00

Fax: +7 495 739-70-70

[yandexdatafactory.com](https://yandexdatafactory.com)

[ydf-customer@yandex-team.ru](mailto:ydf-customer@yandex-team.ru)

Фабрика данных Яндекса,

119021, Москва, ул. Льва Толстого,  
16, Россия