

System Lab



СООРУЖЕНИЕ ОИАЭ НАПРАВЛЕНИЯ ВНЕДРЕНИЯ МЕТОДОВ DATA SCIENCE

ОЦКС Росатома

Москва. Январь 2018 г.

Дмитришин Юрий Михайлович
dmitrishin@system-lab.ru

РЕЗУЛЬТАТЫ РАБОТ



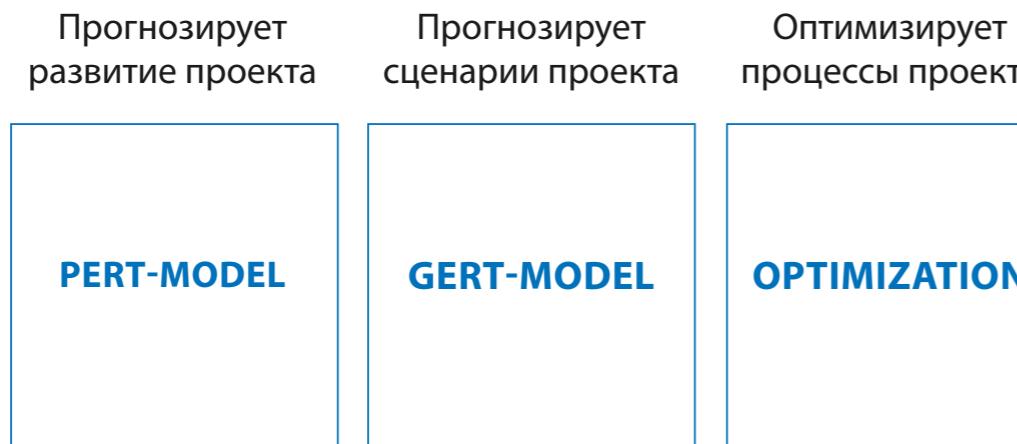
- Комплексная прогнозная PERT-модель сооружения АЭС, которая основана на анализе данных опыта строительства АЭС с 1980 по 2018 г. (с перспективой дальнейшей обработки данных).
- Сценарная GERT-модель сооружения АЭС, которая учитывает возможные альтернативные сценарии реализации проекта в контрольных точках критического пути.
- Алгоритмы оптимизации по многим критериям ключевых организационно-технологических процессов находящихся на критическом пути графика сооружения АЭС и возможных его отклонений.
- Алгоритмы и IT-решения на базе передовых технологий Data Science, для анализа опыта строительства АЭС с 1970 по 2018 г. с перспективной дальнейшего развития.
- Создание баз данных и документов ОТП АЭС-2006 и ОТВ ВВЭР-ТОИ в виде облачной цифровой платформы с Web-интерфейсом, оснащенной инструментами анализа и визуализации.

ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ



- Снижение стоимости и сокращение сроков сооружения АЭС до значений предусмотренных ТЗ на АЭС с ВВЭР-ТОИ за счет оптимального планирования на всех уровнях управления.
- Снижение неопределенности и повышение гибкости управления при реализации проекта за счет увеличения точности прогнозных оценок и управления рисками. По мере накопления новых данных прогнозная точность моделей будет увеличиваться.
- Совершенствование методов ценообразования и нормирования материальных и трудовых ресурсов.
- Методики и алгоритмы сравнения альтернатив по многим критериям.
- Повышение эффективности деятельности участников проектов сооружения ОИАЭ за счет online-доступа к результатам анализа данных и средствам оптимального планирования.
- Дополнение существующих методик управления проектами, стоимостного инжиниринга, экономики и ценообразования информацией, которая получена на основе анализа большого количества данных.

НАУЧНЫЙ ПОТЕНЦИАЛ



ИНСТРУМЕНТЫ DATA SCIENCE
Описывает, исследует, прогнозирует, визуализирует.

ДАННЫЕ ОБ ОПЫТЕ СООРУЖЕНИЯ

- Создается комплекс моделей, которые могут быть использованы для большого количества исследований в сферах: технологии и организации производства, стоимостного инжиниринга, экономики строительства и сметного дела, управления проектами и портфелями, нормирования и ценообразования, передовых ИТ-технологий для нужд атомной отрасли.
- Практически испытываются и отбираются наиболее хорошо себя зарекомендовавшие модели и методы Data Science и машинного обучения.
- За счет инструментов Data Science могут быть найдены неочевидные корреляции и закономерности в процессах, событиях, явлениях скрытые большом объеме данных.
- В результате исследований данных и виртуального эксперимента на моделях, можно будет находить новые функциональные зависимости и новые алгоритмы расчета для процессов и специфики реализации проектов ОИАЭ.

НАУЧНЫЙ ПОТЕНЦИАЛ



- Централизуется, восстанавливается, обобщается и анализируется огромный опыт сооружения ОИАЭ за период минимум 40–60 лет. Этот объем данных в дальнейшем будет передан для нужд сооружения новых АЭС и для вывода из эксплуатации.
- Создается почва для подготовки научных кадров в сферах IT-технологий и Data Science для нужд атомной отрасли в сфере управления проектами.
- Создается технологический задел для дальнейшего перехода от прогнозной и описательной к предписывающей аналитике (так называемая аналитика 3-го поколения).
- Часть решений может адаптироваться и использоваться в других отраслях при реализации крупных инвестиционно-строительных проектов.

ДАННЫЕ О СТРОИТЕЛЬСТВЕ

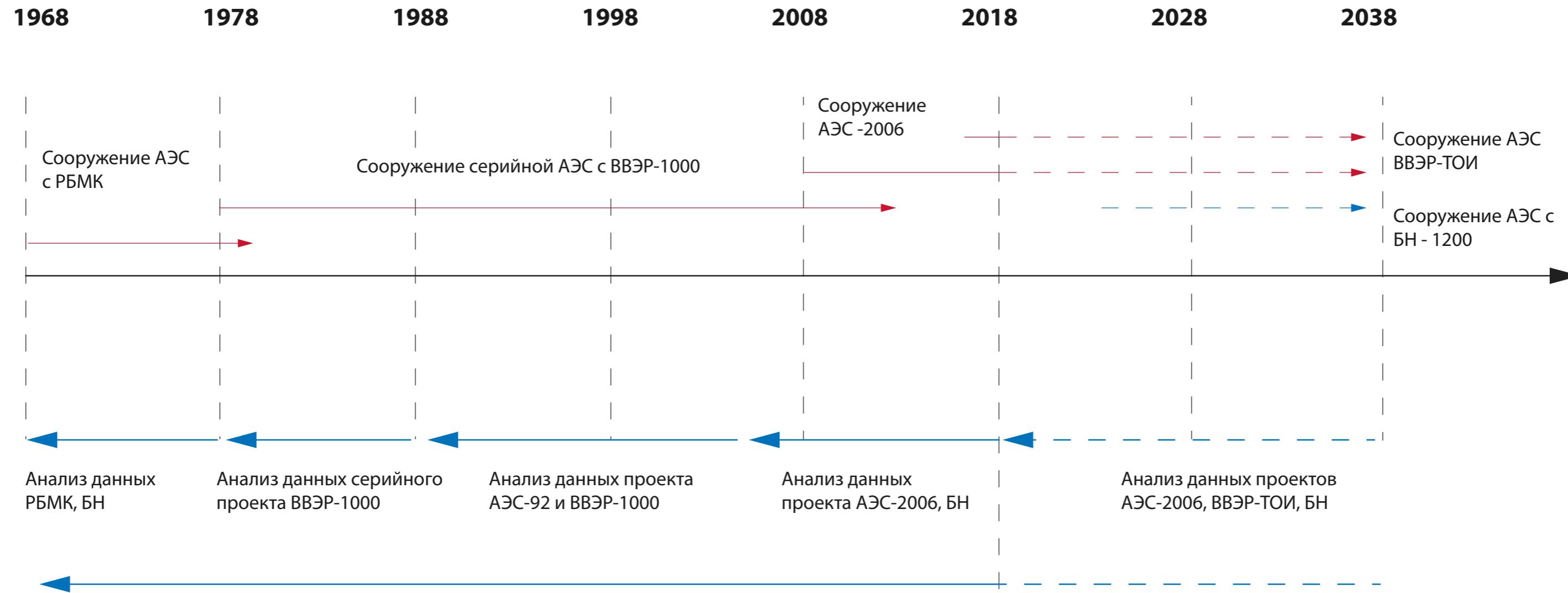


Данные об опыте сооружения АЭС. Обобщение опыта сооружения АЭС с РБМК и ВВЭР с 1970-х по 2018 год с перспективной дальнейшего накопления и анализа данных по наиболее значимым инвестиционно-строительным проектам ОИАЭ.

Целесообразно собирать доступную информацию о ходе строительства реализуемых АЭС, причем от разных участников. Так например большое количество ценных данных находится у монтажных организаций и экспертов, в исполнительной и проектной документации. Полезные данные содержатся как в официальной отчетной документации и базах данных, так и в различных локальных базах и файлах.

Было бы полезно организовать сбор данных по специальным методикам для всех строящихся АЭС и обработать данные по АЭС вводимым и введенным в эксплуатацию.

АНАЛИЗ ОПЫТА СООРУЖЕНИЯ АЭС



АНАЛИЗ ДАННЫХ ОПЫТА ПОТОЧНОГО СООРУЖЕНИЯ АЭС С ВВЭР-1000



81-4
1465

УТВЕРЖДАЮ
Министр энергетики и электрификации СССР
П. С. НЕПОРОЖНИЙ
20 июня 1980 г.

РЕШЕНИЕ
ВСЕСОЮЗНОЙ КОНФЕРЕНЦИИ «ПОТОЧНОЕ
СТРОИТЕЛЬСТВО АЭС, СОКРАЩЕНИЕ ТРУДОЗАТРАТ
И ПРОДОЛЖИТЕЛЬНОСТИ СТРОИТЕЛЬСТВА»

12—15 июня 1980 г.

г. Энергодар

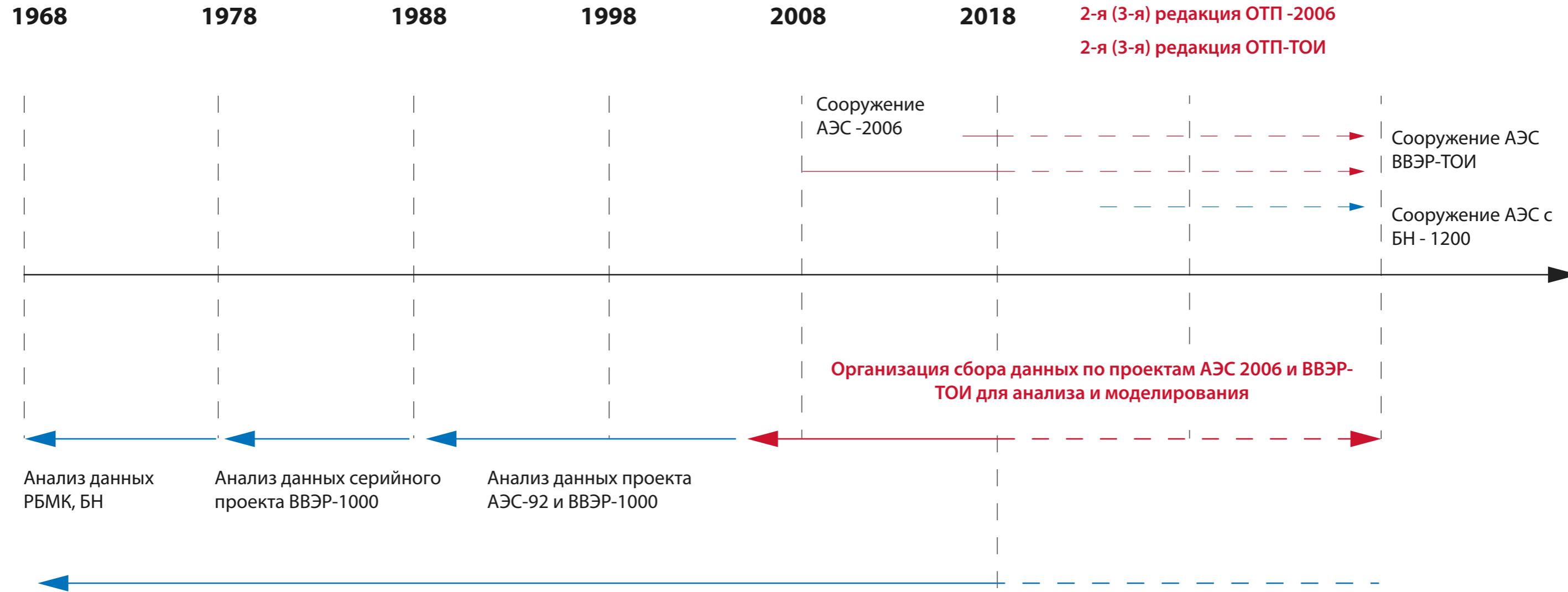
Москва 1980

ОПЫТ СООРУЖЕНИЯ ВВЭР-1000

Некоторые выдержки по результатам Решения всесоюзной конференции «Поточное строительство АЭС, сокращение трудозатрат и продолжительности строительства» от 12-15 июня 1980 года:

- Атомэнергостройпроекту совместно с другими заинтересованными организациями разработать и утвердить у руководства Минэнерго СССР комплексный сетевой график строительства Запорожской АЭС. Выполнить его оптимизацию по материальным и трудовым ресурсам, исходя из продолжительности строительства до ввода блока 4 года, со сроком ввода первого энергоблока Запорожской АЭС в 1983 г.
- Главниипроекту и ГПТУ по строительству для сооружения группы АЭС с однотипным оборудованием и аналогичными проектными решениями разработать эталонный ПОС и ППР, в которых предусмотреть и технически проработать основные мероприятия, обеспечивающие сокращение трудозатрат и сроков строительства.
- До начала строительства каждой АЭС выполнять расчет трудозатрат и численности рабочих по объектам, видам работ, потокам. Союзатомэнергострою и главным производственным управлением вести систематический учет и анализ фактических трудозатрат на физические объемы основных работ.

АНАЛИЗ ДАННЫХ ОПЫТА СООРУЖЕНИЯ АЭС-2006, АЭС ВВЭР-ТОИ И БН-1200



ТЕКУЩЕЕ СООРУЖЕНИЕ АЭС В РОССИИ И ЗА РУБЕЖОМ



DATA SCIENCE И СТАТИСТИКА



Кратко можно выделить следующие методы анализа данных:

- Описательная статистика (descriptive statistics) или разведочный анализ данных — это статистические методы обработки данных, их систематизации, наглядного представления в виде таблиц и графиков, а также количественное описание данных с помощью системы статистических показателей.
- Дисперсионный анализ (analysis of variance) — статистический метод изучения взаимосвязи. Применяется для исследования влияния одной или нескольких качественных переменных на одну зависимую количественную переменную.
- Кластерный анализ (cluster analysis) — совокупность многомерных статистических методов классификации объектов по характеризующим их признакам, разделение совокупности объектов на однородные группы, близкие по определяющим критериям, выделение объектов определенной группы.
- Регрессионный анализ (regression analysis) — это метод изучения статистической взаимосвязи между одной зависимой количественной зависимой переменной от одной или нескольких независимых количественных переменных. Зависимая переменная в регрессионном анализе называется результатирующей, а переменные факторы — предикторами или объясняющими переменными.

МАШИННОЕ ОБУЧЕНИЕ



ИНСТРУМЕНТЫ DATA SCIENCE

Описывает, исследует, прогнозирует, визуализирует.

ДАННЫЕ ОБ ОПЫТЕ СООРУЖЕНИЯ

Задача классификации состоит в том, что требуется определить к какому из известных классов относятся исследуемые объекты, то есть классифицировать эти объекты. Причем каждый из этих объектов имеет некоторое количество характеристик. Например рассматривается большое количество объектов имеющих несколько фиксированных характеристик и требуется дать ответ «да» или «нет» по каждому из этих объектов.

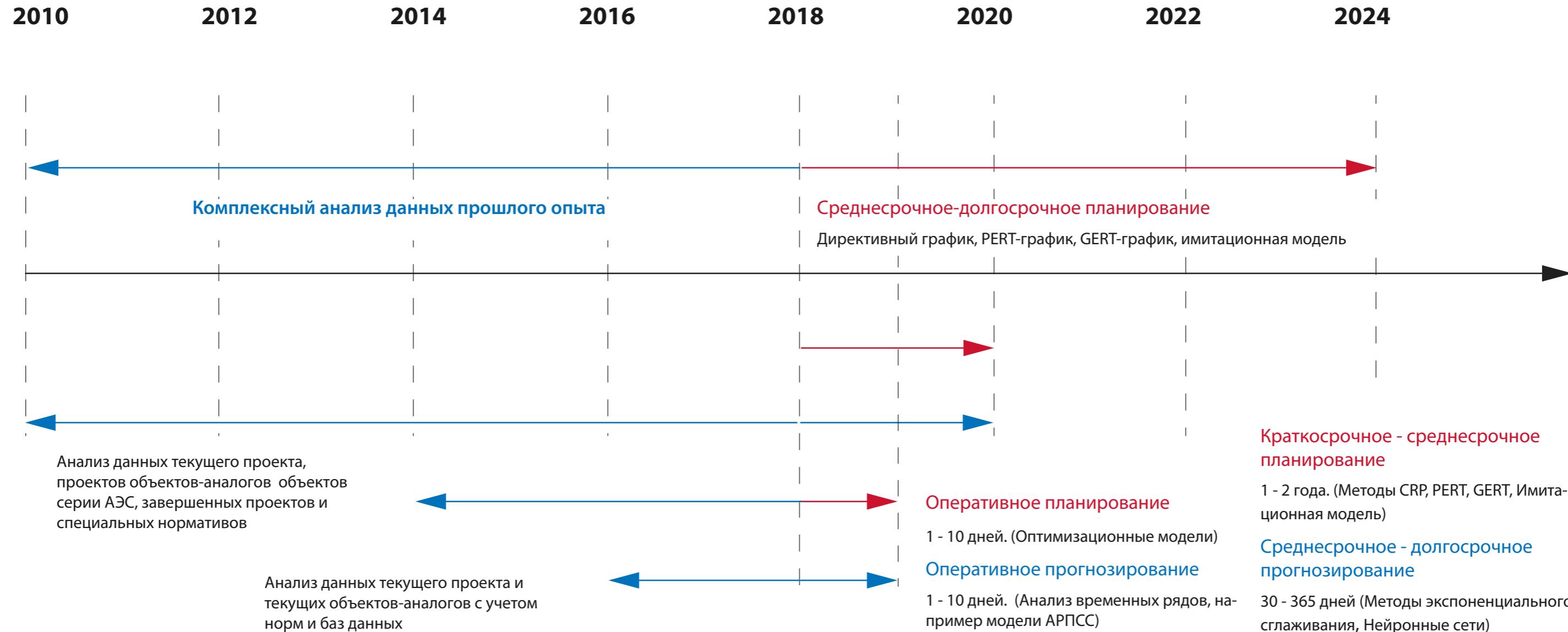
Задачу классификации рассматривают как задачу определения значений одного из параметров анализируемого объекта на основании значений других параметров. Определяемый параметр — это зависимая переменная, а параметры, участвующие в его определении, — независимыми переменными. Если значениями независимых и зависимой переменных являются действительные числа, то такая задача называется задачей регрессии.

Задача кластеризации состоит в разделении исследуемого множества объектов на группы «похожих» объектов, называемых кластерами. Слово кластер переводится как сгусток, пучек, группа. Задача разбиения множества элементов на кластеры называют кластерным анализом. Кластеризация может применяться практически в любой области где необходимо исследование экспериментальных или статистических данных.

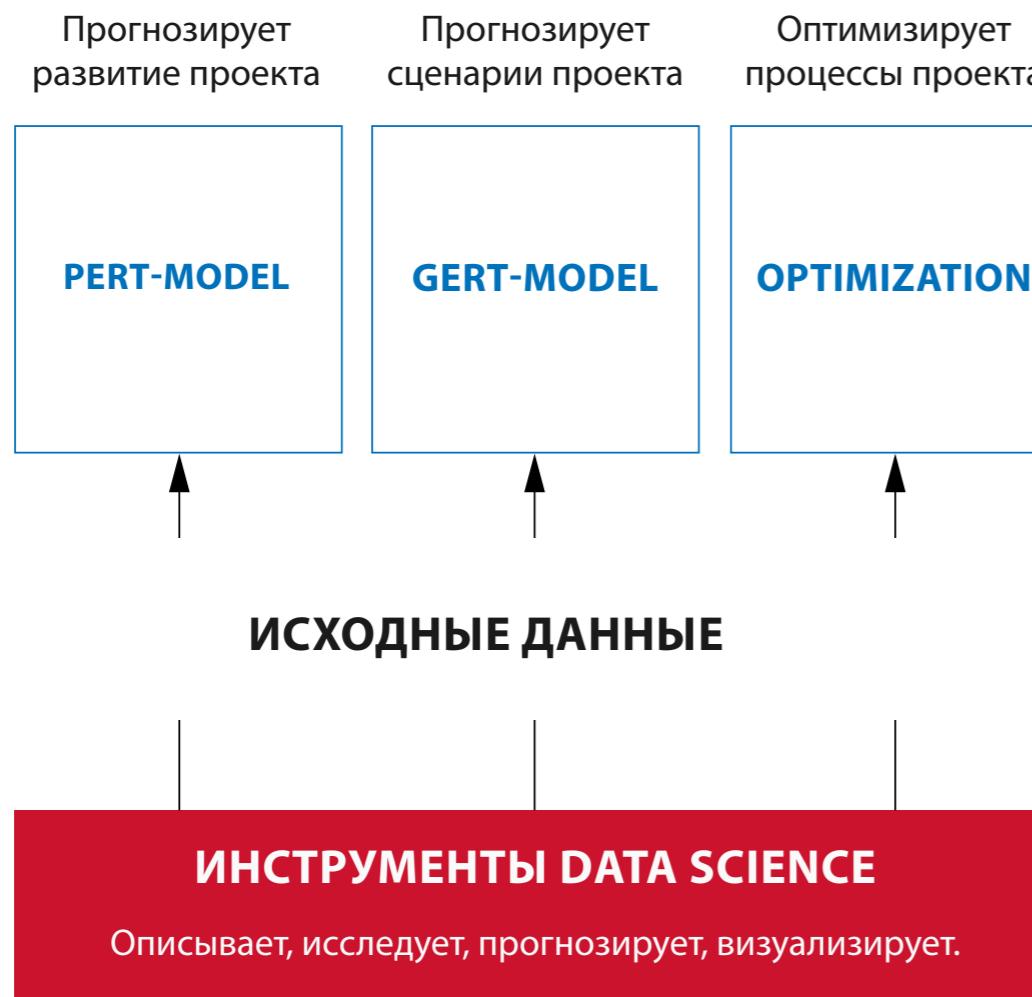
МЕТОДЫ DATA SCIENCE ДЛЯ АНАЛИЗА ОПЫТА СООРУЖЕНИЯ АЭС



ПРОГНОЗИРОВАНИЕ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ DATA SCIENCE



ВЗАИМОДЕЙСТВИЕ МОДУЛЕЙ

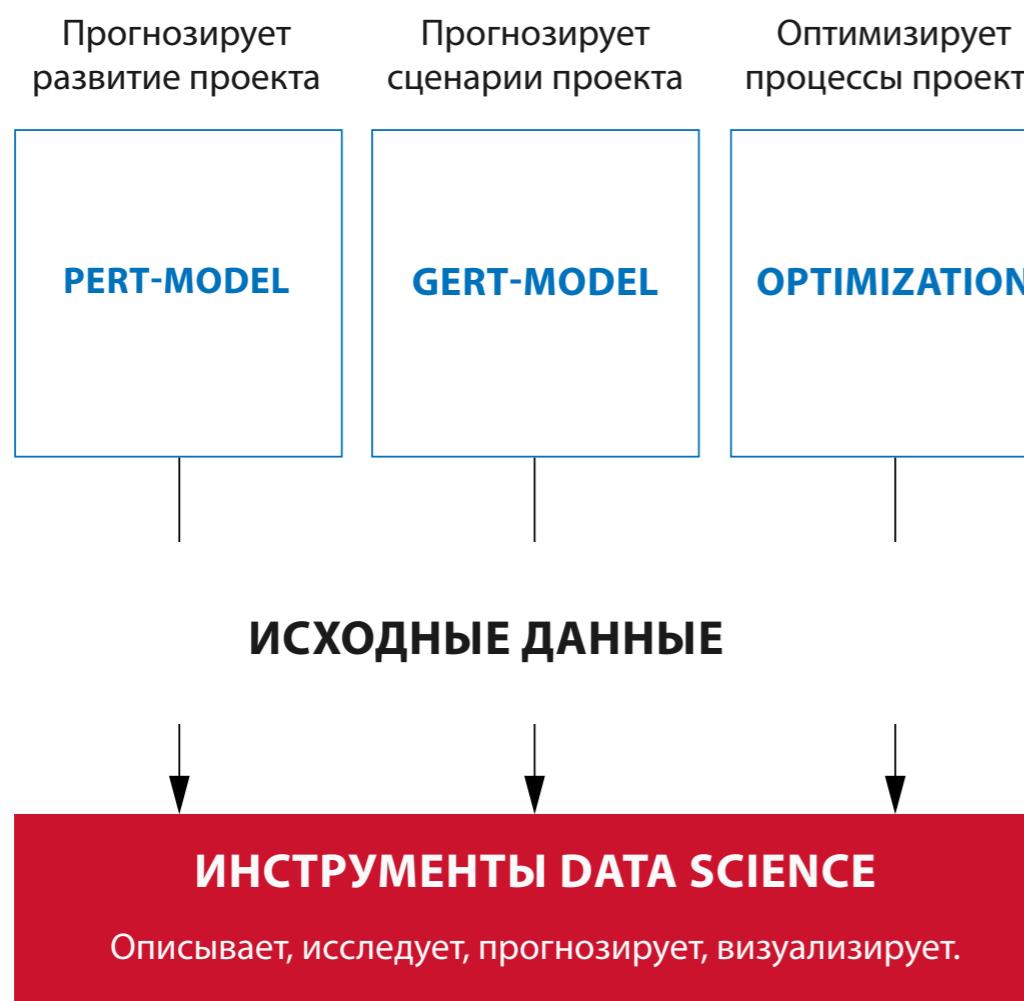


PERT- и GERT- модели, а также модели оптимизации получают исходные данные после того, как они обработаны методами и инструментами Data Science. Это может быть решение статистических задач, решение задач машинного обучения. Основной смысл PERT модели заключается в том, что она основана на реальных наблюдениях или любых других фактических данных.

В общих чертах можно охарактеризовать это взаимодействие следующим образом:

- PERT-модель — это ключевая целостная структура, в которую «вкладываются» данные как атрибуты. Целостность этой структуры заключается в том, что она охватывает все работы по сооружению АЭС.
- GERT-модель — это структура ключевых процессов, которая «вырезана» из общей структуры PERT и дополнена альтернативными сценариями. Атрибуты в такой модели также как и в PERT присваиваются на основании обработки данных реального опыта.
- Модуль моделей оптимизации работает как со структурами PERT- и GERT-модели так и с отдельными группами атрибутов этих моделей.
- Инструменты и методы Data Science готовят данные для их дальнейшего использования в моделях. По сути это тоже совокупность моделей, на входы которым подаются некоторые данные, а из них извлекается полезная информация.

ВЗАИМОДЕЙСТВИЕ МОДУЛЕЙ



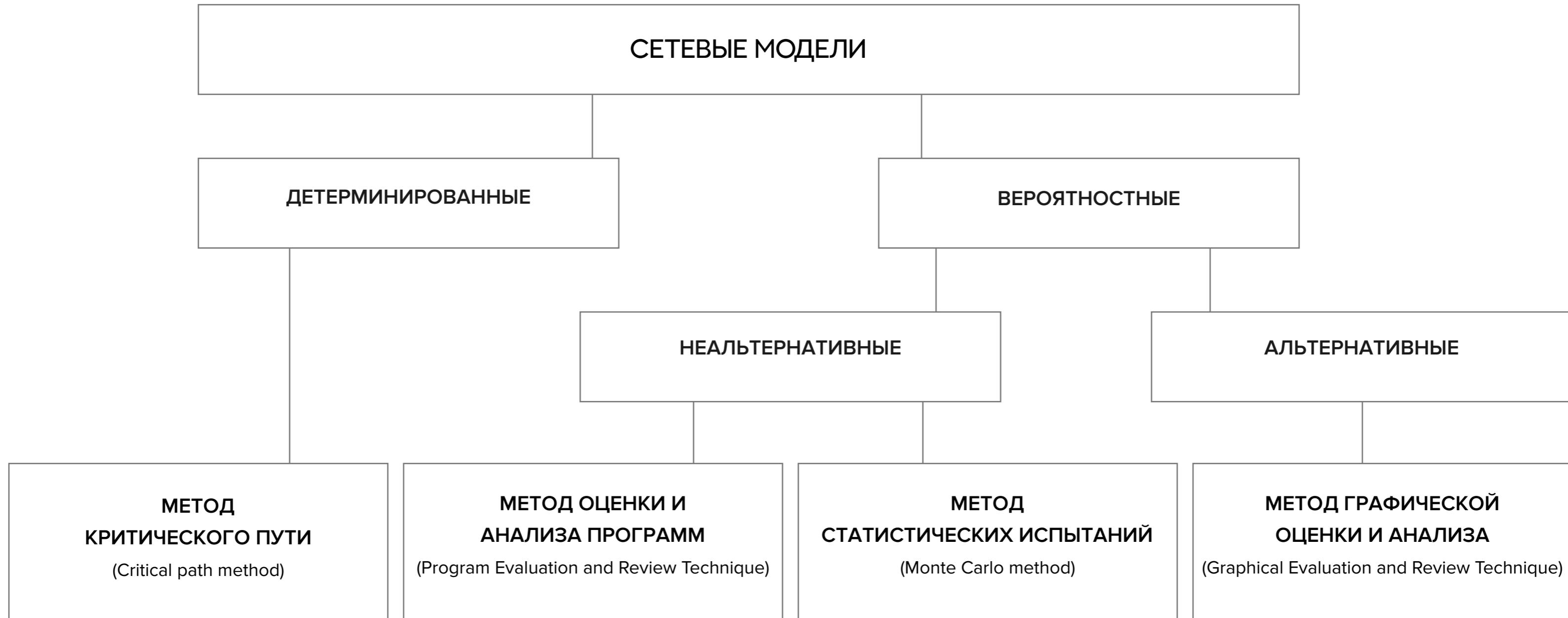
Смысл обратного процесса обмена данными заключается в том, что инструменты Data Science могут использоваться в направлении «от модели к инструментам Data Science».

Например PERT — модель может изменять свои параметры в заданных пределах с использованием генератора псевдослучайных чисел для некоторого набора атрибутов. Это могут быть модели типа Монте-Карло, модели некоторых процессов. Имея базовую структурную модель, её можно использовать как исследовательскую модель. На ней можно проверять гипотезы, выполнять виртуальные эксперименты. Такая модель может генерировать достаточное количество статистически значимых данных.

Например можно моделировать, как повлияет одна или несколько новых технологий на весь процесс сооружения АЭС. Множество параметров этой технологии учитываются как в структуре работ WBS, так и в качестве атрибутов. Затем выполняется перебор и подстановка значений и производится наблюдение. Результаты наблюдений могут сохраняться и обрабатываться.

Наличие инструментов Data Science дает широкое поле для исследований и моделирования.

КЛАССИФИКАЦИЯ СЕТЕВЫХ МОДЕЛЕЙ



PERT-МОДЕЛЬ



ИНСТРУМЕНТЫ DATA SCIENCE
Описывает, исследует, прогнозирует, визуализирует.

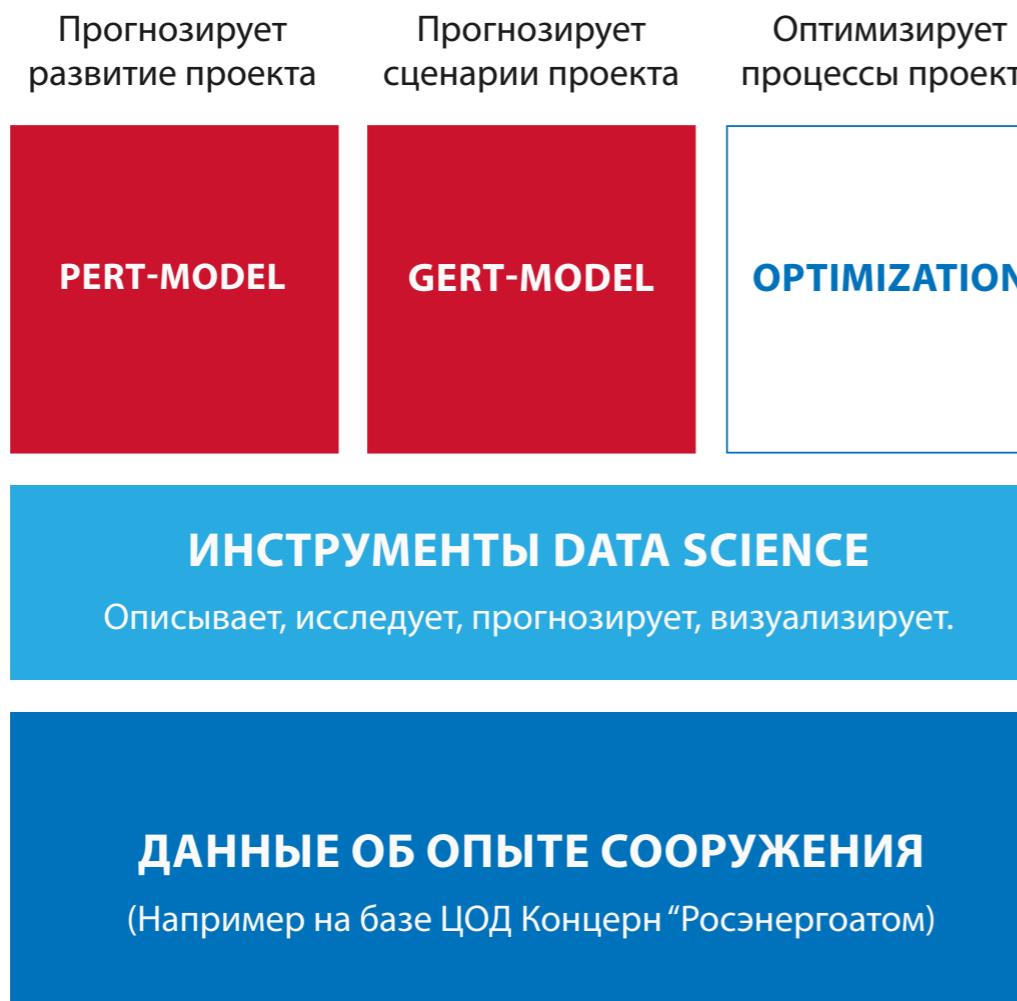
ДАННЫЕ ОБ ОПЫТЕ СООРУЖЕНИЯ
(Например на базе ЦОД Концерн "Росэнергоатом")

Комплексная прогнозная вероятностная PERT-модель сооружения АЭС. Основана на анализе данных опыта строительства АЭС с 1980 по 2018 г (с перспективой дальнейшей обработки данных). Модель позволит прогнозировать реализацию проекта по многим критериям, оценивать риски, изучать и ликвидировать различные проектные отклонения. Основная ценность модели заключается в том, что она учитывает прошлый опыт сооружения АЭС и совершенствуется за счет накопления новых данных.

PERT предназначен для очень масштабных, единовременных, сложных, нерутинных проектов. Метод в особенности нацелен на анализ времени, которое требуется для выполнения каждой отдельной задачи, а также определение минимального необходимого времени для выполнения всего проекта.

Методология PERT хорошо известна, но с этой методологией связан ряд сложностей. Поскольку в таких моделях необходимо задавать вероятностные веса ребрам ориентированного графа в сетевой модели плана работ. Для решения такой задачи целесообразно использовать методы Data Science.

GERT-МОДЕЛЬ

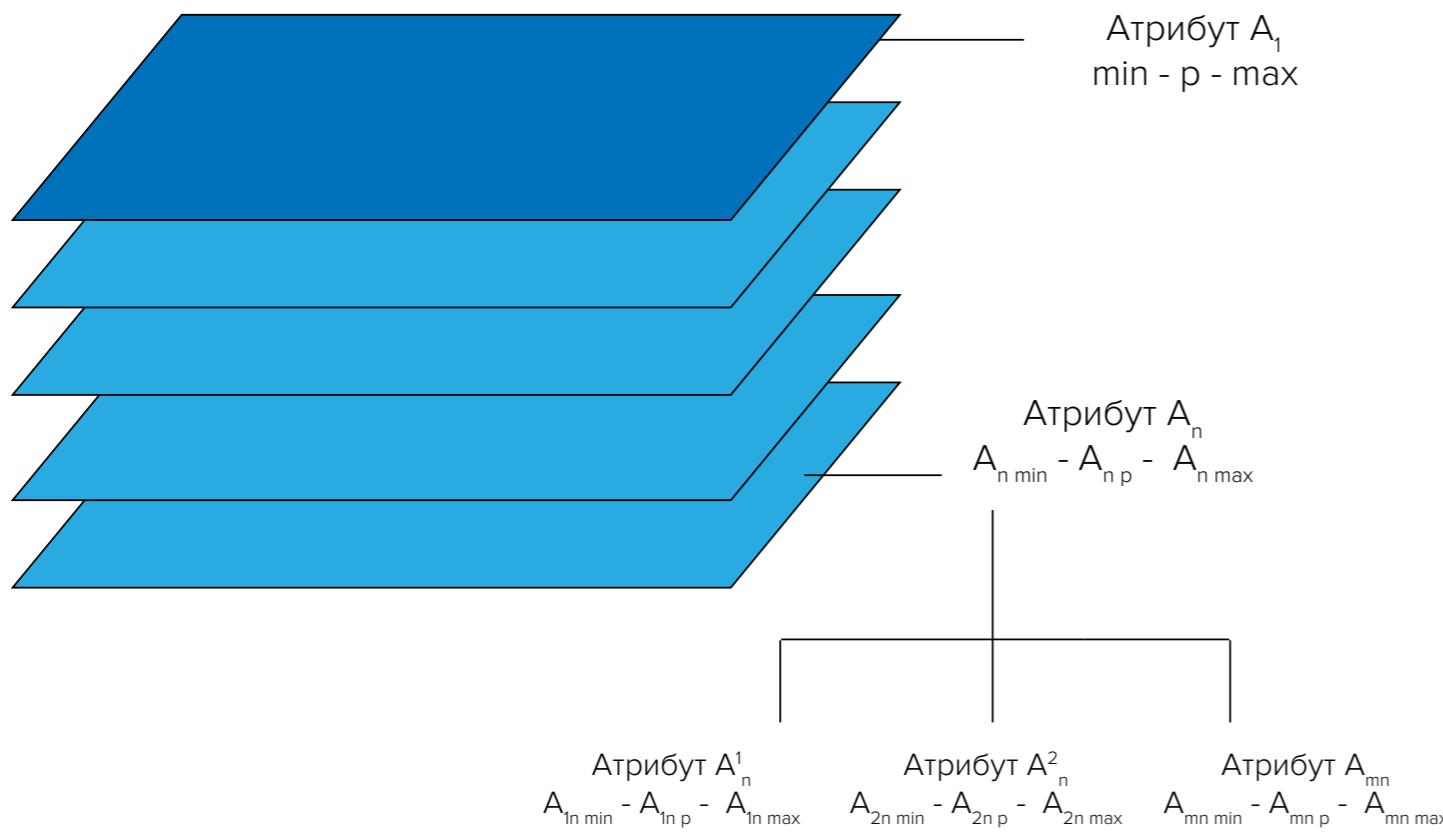


Наиболее вероятные критические пути PERT-модели передаются в GERT-модель. Метод графической оценки и пересмотра программ (GERT) позволяет учесть риск изменения состава работ при наступлении определенных событий или по результатам выполнения предшествующих работ. В сетевой модели GERT могут создаваться точки ветвления или точки выбора, после которых планируются несколько независимых цепочек работ, не все из которых выполняются.

Учитывает ключевые технологии сооружения АЭС и возможные альтернативные сценарии реализации проекта в контрольных точках критического пути. Дополняет PERT-модель альтернативными сценариями развития проекта в ряде ключевых точек вероятного критического пути.

АТРИБУТЫ И СТРУКТУРА

Атрибуты в сетевой модели



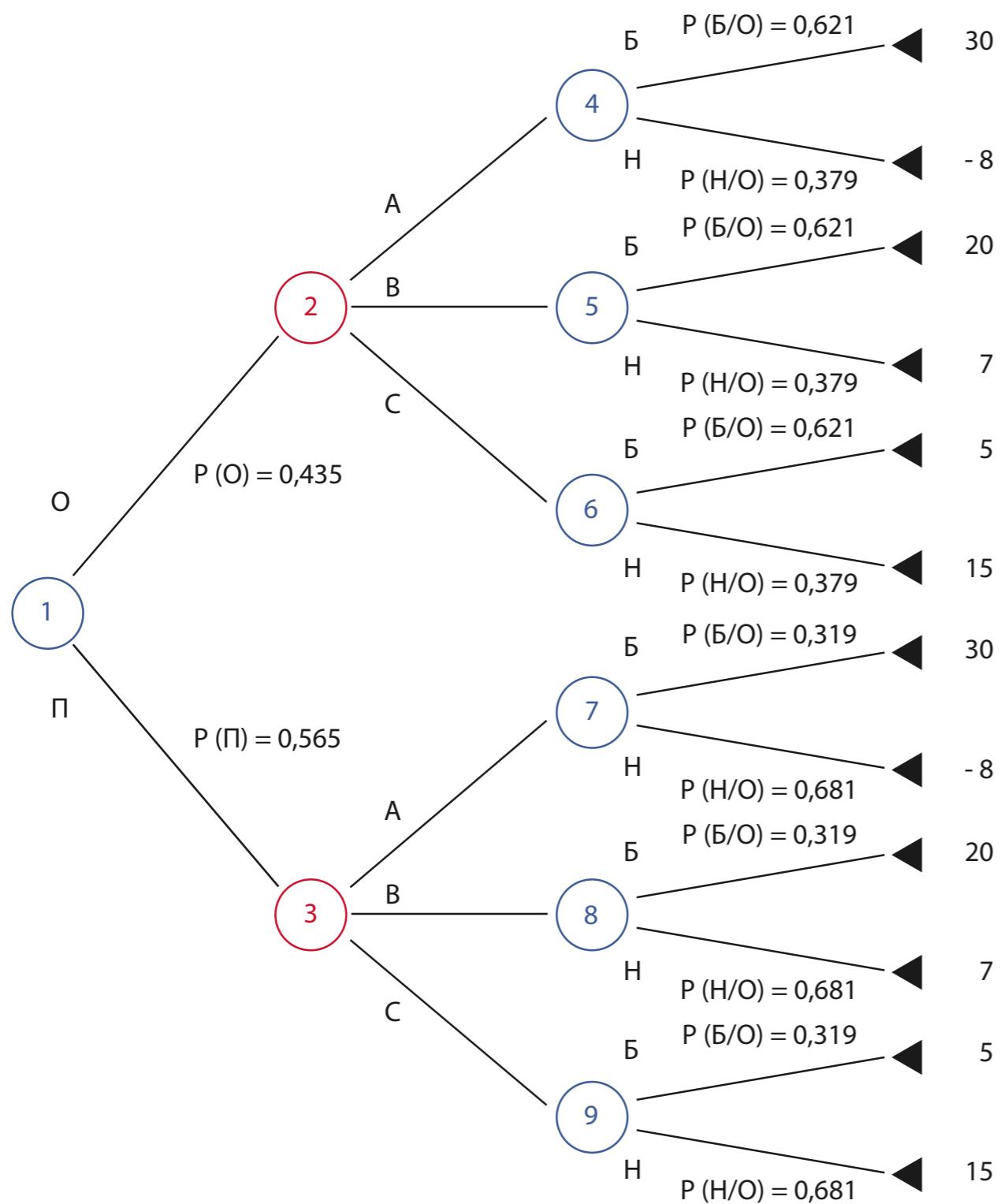
Сетевая модель PERT представляет собой направленный граф. Каждой дуге данного графа соответствует некоторый атрибут. Количество атрибутов данного графа достаточно велико. Атрибуты представляют собой такие характеристики проекта как длительность, трудовые и материальные ресурсы, стоимости, кодировки и.т.д.

Используя методы Data Science представляется возможным задавать значения атрибутов как для продолжительности работ, так и для других характеристик которые описывают этот тип работ не просто нормативно, а на основе обработки данных.

Выделяя некоторый атрибут в сетевой модели PERT, получаем модель, которая определена данным типом атрибута. Причем каждый атрибут может быть представлен в виде дерева с родительским элементом.

PERT-модель с декомпозицией по ресурсам, может быть представлена в виде ресурсной сетевой модели, где веса дуг графа заданы вероятностными величинами, или функциями, а сами ресурсы могут быть представлены в виде классификации или дерева.

PERT-модель также учитывает структуру работ WBS, организационную структуру (OBS – Organizational Breakdown Structure) и структуру статей затрат (CBS – Cost Breakdown Structure).



ДЕРЕВЬЯ РЕШЕНИЙ

Данные из PERT и GERT моделей для той или иной точки ветвления могут быть представлены в виде дерева решений, а задача может быть сформулирована как задача принятия локальных решений для тех или иных ситуаций.

Деревья — графическое средство анализа в условиях риска. Деревья создаются для использования в моделях, в которых выполняется последовательный выбор альтернатив. Каждое решение ведет к своему результату. Модель используют как визуальный аналитический инструмент, где рассчитываются ожидаемые значения или ожидаемую полезность конкурирующих вариантов.

Основные идеи модели такого типа:

- Для каждого решения находят полезность всех возможных результатов.
- Определяют вероятность каждого результата. Для этого используют методы теории вероятностей, математической статистики и Data Science.
- Находят ожидаемую полезность каждого решения.
- Оптимальное решение то, которому соответствует наибольшая ожидаемая полезность.

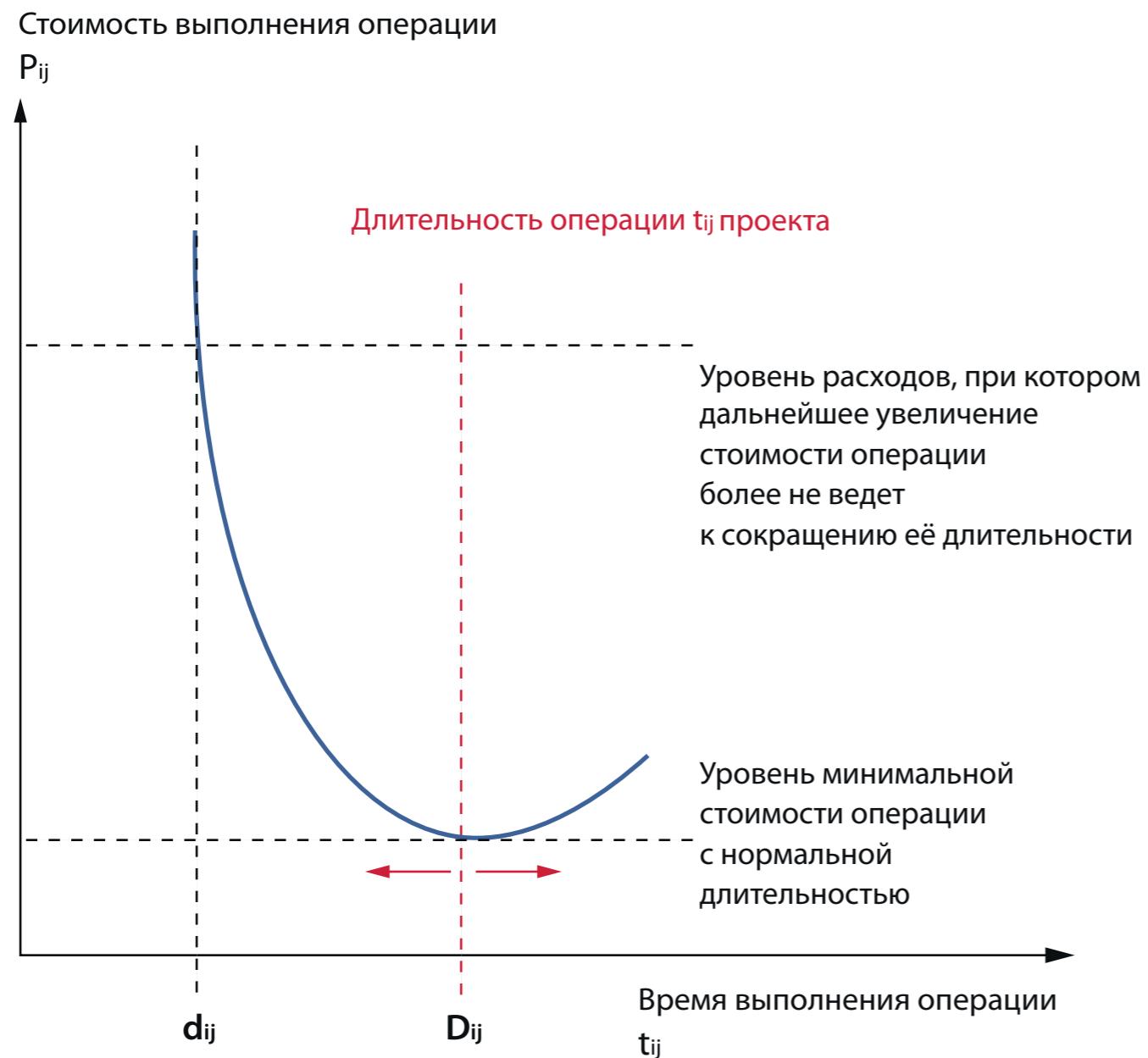
ОПТИМИЗАЦИЯ СЕТЕВОЙ МОДЕЛИ



Оптимизация по критерию стоимости проекта, может выполняться на основе следующих предположений:

- Функция времени-стоимости описывается кривой.
- Продолжительность каждой операции t_{ij} изменяется между двумя границами d_{ij} и D_{ij} , которые обоснованы техническими или экономическими соображениями.
- Граница D_{ij} соответствует минимальным расходам и приемлемой длительности выполнения операции.
- Граница d_{ij} соответствует минимальной продолжительности выполнения операции, однако в этом случае увеличиваются расходы.
- Сроки выполнения операций можно сокращать, одновременно увеличивая стоимость, но только до величины, после которой увеличение расходов не приводит к сокращению длительности операции.
- Если продолжительность операции возрастает, становится больше D_{ij} , то стоимость операции так же возрастает за счет накладных расходов, простоев и прочих затрат, что изображено на графике как отклонение вправо.
- Если срок выполнения каждой операции будет соответствовать D_{ij} , то продолжительность проекта «нормальная». Если ускорить до максимума каждую операцию и срок её выполнения будет равен d_{ij} , то это «срочный» план проекта.

ОПТИМИЗАЦИЯ СЕТЕВОЙ МОДЕЛИ



В проекте, выполняется ограничение на длительность работ так что:

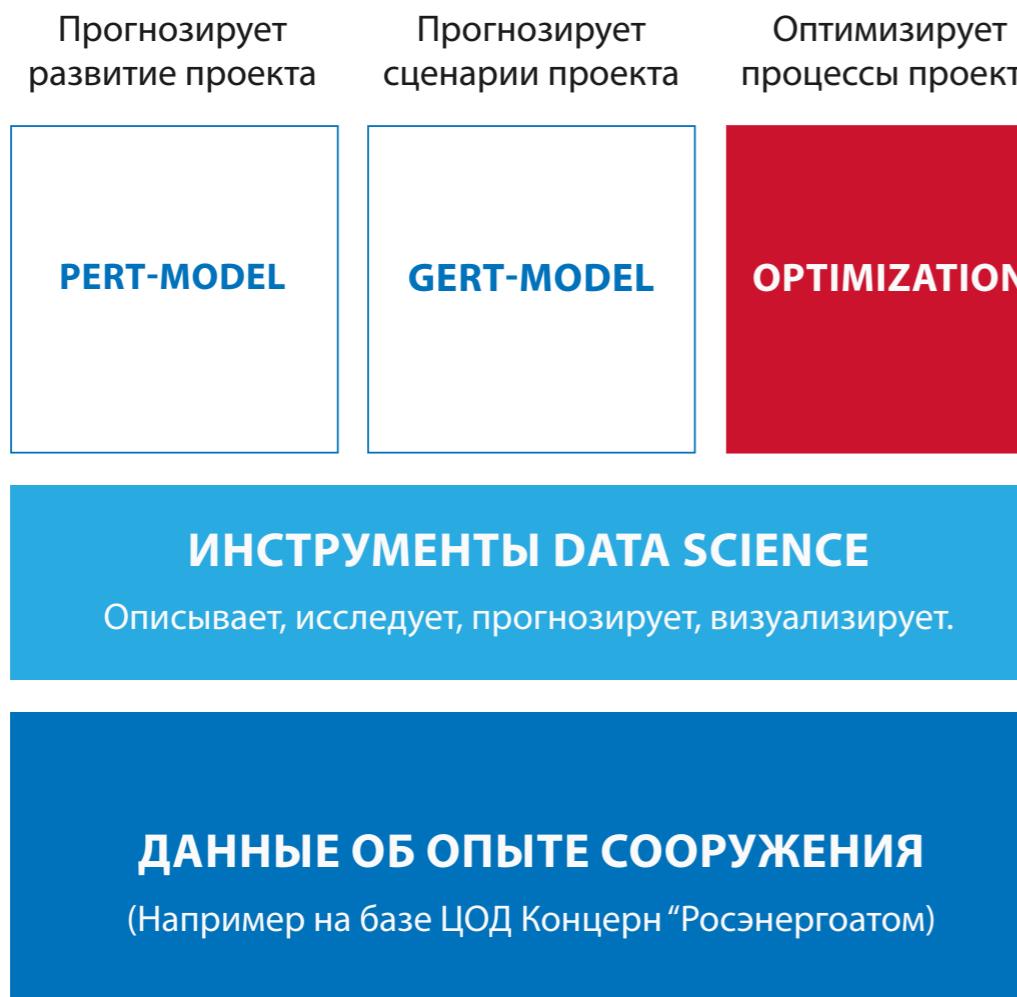
$$d_{ij} \leq t_{ij} \leq D_{ij}$$

то есть, продолжительность каждой работе в проекте расположена в пределах двух границ.

Вычислив критический путь, определим стоимость проекта. Полная стоимость проекта равна сумме стоимостей всех операций. Для уменьшения полной стоимости проекта при тех же сроках выполнения, уменьшим свободные резервы некритических операций в пределах, установленных ограничениями. Уменьшение свободного резерва, ведет одновременно и к возрастанию длительности, и к уменьшению стоимости.

Продолжительность операций увеличивается до границы D_{ij} не пересекая её. Вместе с тем, увеличение сроков проекта производится только по тем работам, которые находятся вне критического пути, то есть общая продолжительность проекта не изменяется. Тем не менее, при таких изменениях, может измениться критический путь проекта. Такая оптимизация — своеобразная «тонкая настройка». В этом случае, целесообразно проверить решение на устойчивость. Для ряда процессов, функция длительность-время разная, не всегда представляется возможным определить точную функциональную зависимость.

МЕТОДЫ ОПТИМИЗАЦИИ



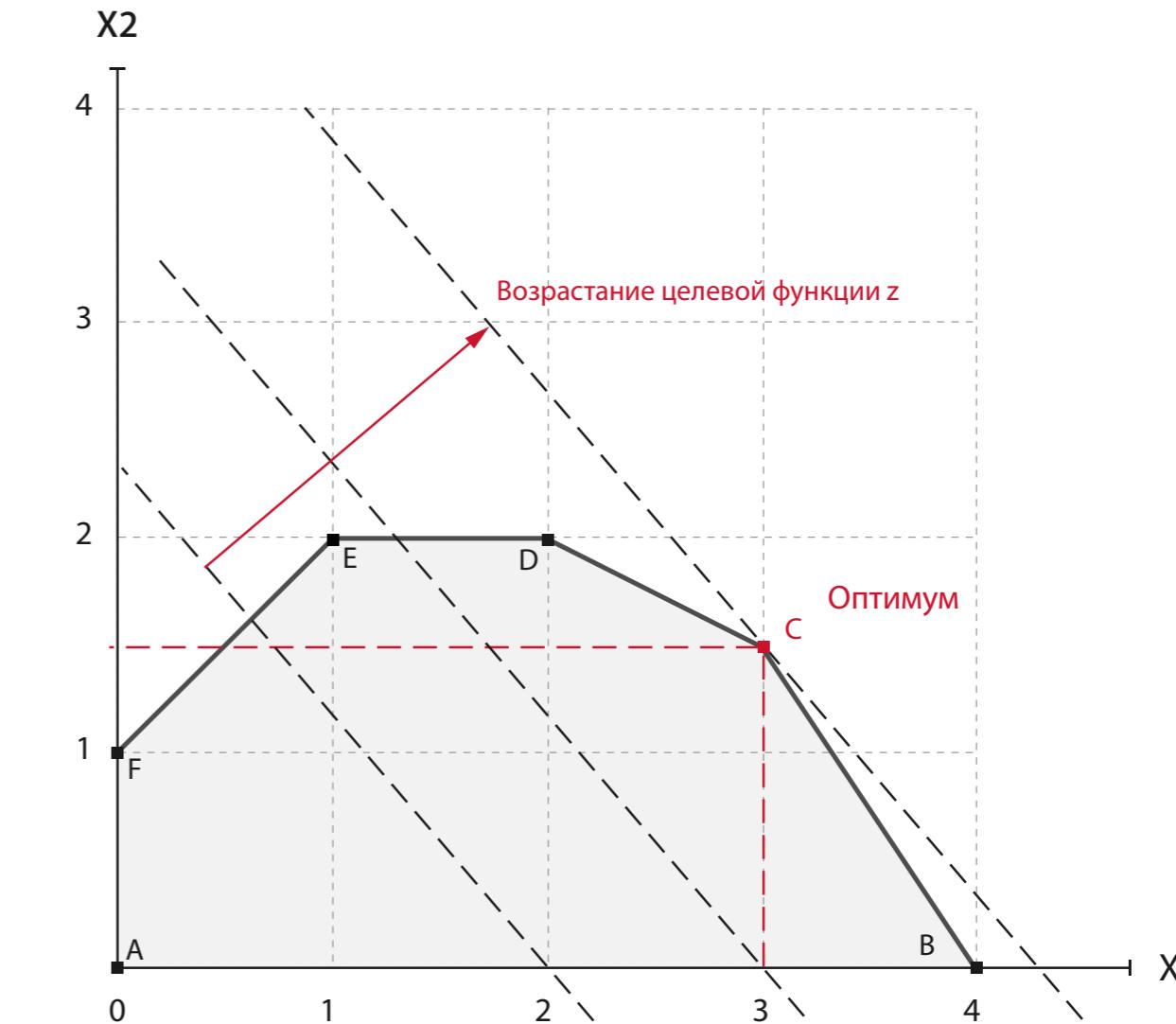
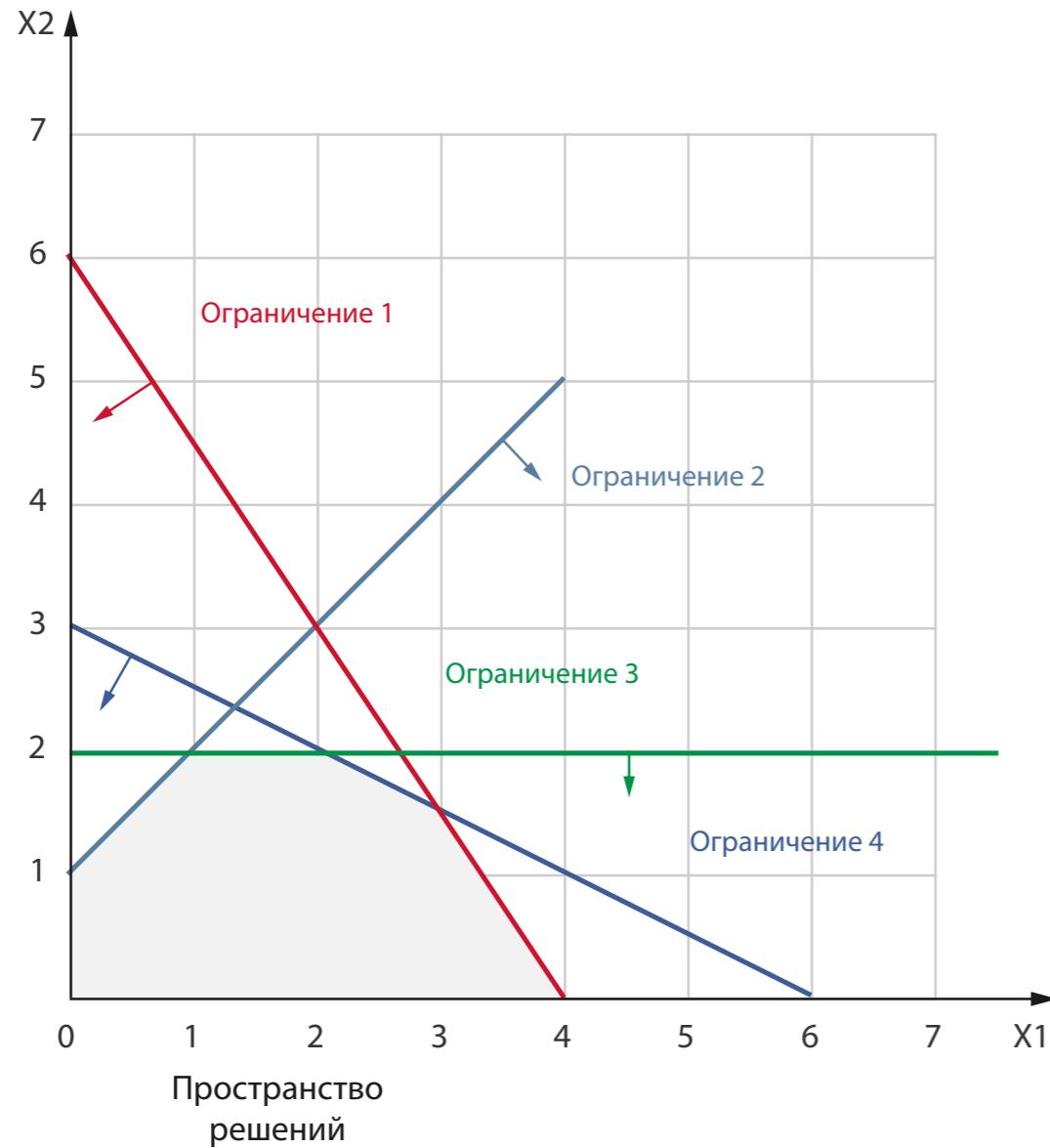
На основе данных, которые получены в PERT и GERT-модели выполняется оптимизация для организационно-технологических решений возможных критических путей.

Для этого используются методы и модели исследования операций. Моделирование с использованием подходов исследования операций, которые включают математические модели следующих видов:

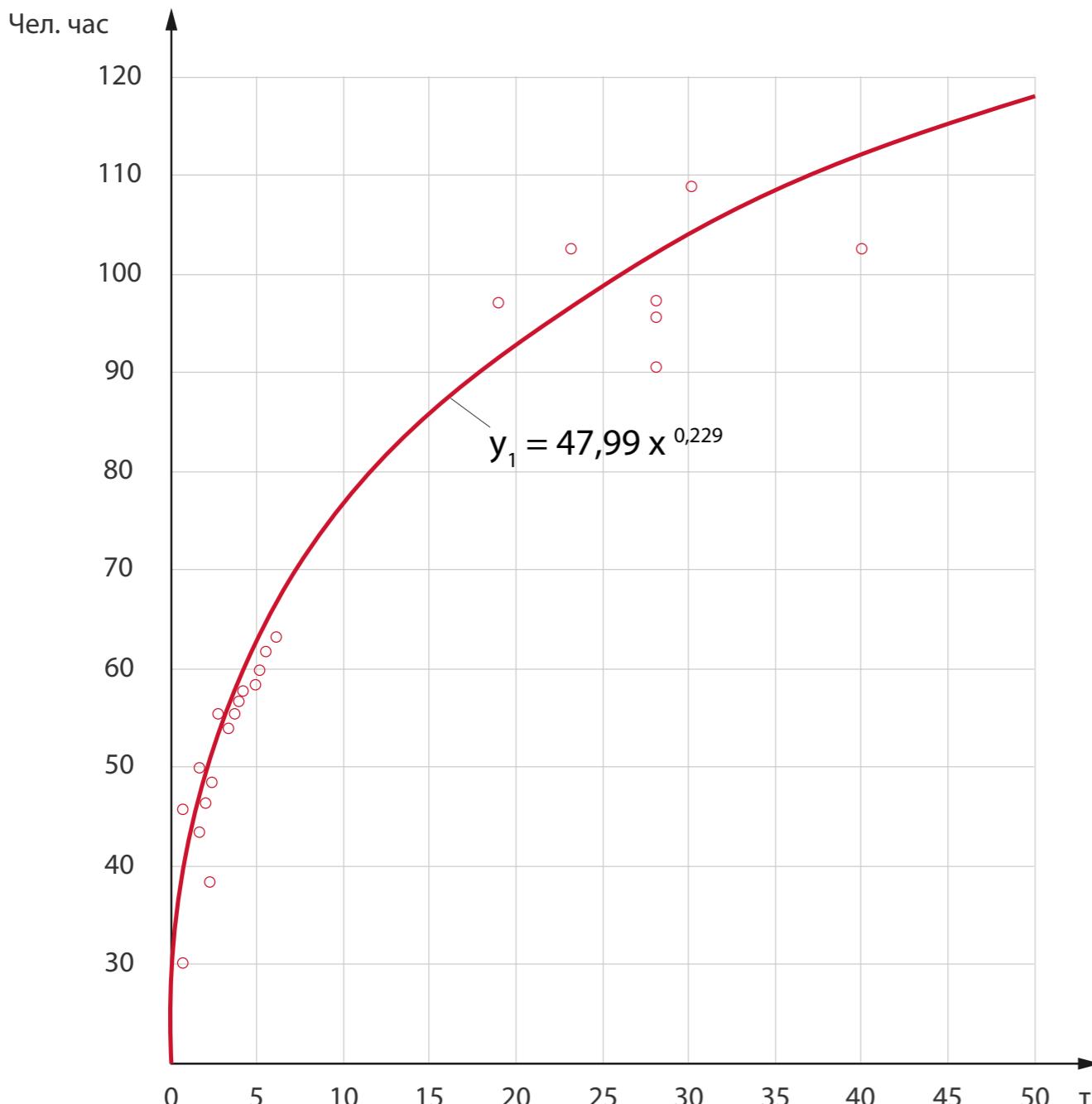
- задачи оптимизации перевозок;
- задачи оптимального распределения объемов работ;
- задачи массового обслуживания (например работа ведущего комплекта машин);
- задачи управления запасами и хранения;
- задачи оптимального распределения заданий между участками;
- задачи недельно-суточного планирования;
- задачи о назначениях ресурсов на различные типы работ
- задачи оптимального раскroя материалов;
- и.т.д.

Все 4 блока, а именно блок инструментов Data Science для анализа данных, блоки PERT и GERT а так же блок оптимизационных моделей работают итерационно и параллельно.

МОДЕЛИ ОПТИМИЗАЦИИ



ОПТИМИЗАЦИЯ. АРМОБЛОКИ.



Крупноблочное строительство сборно-монолитных конструкций. Информэнерго. 1990 г. «Поскольку не имеется единой методики разбивки на монтажные блоки, на каждой стройплощадке применяется своя разрезка сборно-монолитных конструкций. В этих условиях трудно установить нормы и расценки на изготовление, сборку с укрупнением и монтаж блоков. ВНИР не охватывает всего перечня работ, применяемых на АЭС. На каждой стройплощадке разработаны индивидуальные калькуляции на монтажные блоки разной массы на стадиях изготовления, сборки и монтажа. Для определения зависимости между массой монтажных блоков и трудоемкостью работ на стадии изготовления, сборки и монтажа проанализированы калькуляции, разработанные на Смоленской и Курской АЭС, а также в институте «Атомэнергостройпроект».

Анализ статистических данных показал, что исследуемые зависимости могут аппроксимироваться степенными функциями. Вычисленные зависимости между массой монтажных блоков и трудоемкостью на стадиях изготовления, сборки с укрупнением и монтажа представлены в таблице. Полученные корреляционные зависимости и графики позволяют определить трудоемкость работ в зависимости от массы монтажных блоков для используемых типов монтажных блоков на стадиях изготовления, сборки с укрупнением и монтажа. Используя полученные зависимости, имеется возможность определить время монтажа монтажных блоков и приведенные затраты.»

System Lab