



УНИВЕРСИТЕТ ИТМО

# Гибридные технологии предсказательного моделирования: концепции и приложения

С.В. Ковальчук

Новосибирск, 21-23 декабря 2017 г.

# Постановка проблемы

- Усложнение моделируемых систем ведет к усложнению структуры моделей
- Моделирование сложных систем зачастую требует комбинации моделей различных классов и масштабов в рамках одного решения
- Разработка обоснованной, полной и качественной модели становится сложной задачей
- Нетривиальность задачи управления сложными моделями требует разработки новых интеллектуальных решений для решения задач идентификации, калибровки, усвоения данных, оптимизации и пр.
- Решение прикладных задач на базе моделирования сложных систем требует автоматизации средств управления сложными моделями

# Классы моделей (1/3)

- **Класс I:** уникальные модели
- **Структура модели:** детерминированная, задана из первых принципов (структурой объекта)
- **Параметры:** не заданы
- **Функциональные соотношения:** не определены
- **Особенности моделирования:** численные методы, индивидуальная реализация

## Классы моделей (2/3)

- **Класс II:** модели сложных систем
- **Структура модели:** недетерминированная
- **Параметры:** не заданы
- **Функциональные соотношения:** экспертные знания, эвристические зависимости (правила, шаблоны)
- **Особенности моделирования:** комплексные сети, многоагентные модели и пр.

## Классы моделей (3/3)

- **Класс III:** модели на данных
- **Структура модели:** неизвестна
- **Параметры:** не заданы
- **Функциональные соотношения:** неизвестны
- **Особенности моделирования:** построение регрессионных или классифицирующих закономерностей, интеллектуальный анализ данных/процессов

# Функциональная гибридизация (1/3)

- Композиция моделей
  - Использование структуры системы
  - Моделирование отдельных элементов самостоятельными моделями
  - Внедрение сервисных моделей
- Естественное отображение на композитное приложение

- Ограничение 1: природа сложных систем (многомасштабность и дальние связи)
- Ограничение 2: сложность построения полной и качественной модели

# Функциональная гибридизация (2/3)

- Альтернативные модели
    - Альтернативные формальные модели
    - Альтернативные реализации
    - Альтернативные параметры
  - Отображение на КП – неестественное (несколько моделей для одного процесса)
- Агрегация результатов моделирования
  - Необходимость управления неопределенностью моделей и данных
  - Организация работы с разнородными моделями

# Функциональная гибридизация (3/3)

- Замещение моделей
  - Эвристические правила
  - Альтернативные по природе модели (например, многоагентные модели и ДУ)
  - Суррогатные модели (автоматизация выращивания моделей)
- Отображение на КП – противоестественное (внедрение моделей, не соответствующих процессам в исследуемом объекте)

- Эволюция и адаптация моделей
- Использование методов машинного обучения, искусственного интеллекта и т.п.
- Автоматизация управления структурой моделей

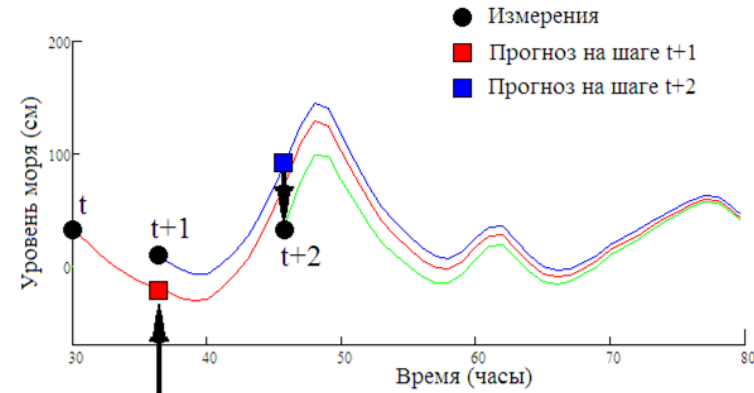


# Входные данные

- Вариабельность входных данных
  - Естественная – анализ данных как случайных величин
  - Техническая – данные из различных источников
  - Искусственная – внесение шума во входные данные и параметры
- Специфика качества данных
  - Неточность данных
  - Противоречивость альтернативных данных
  - Степень доверия к источнику данных
  - Субъективность данных (особенно – в социальных системах)
- Неопределенность данных
  - Управление неопределенностью

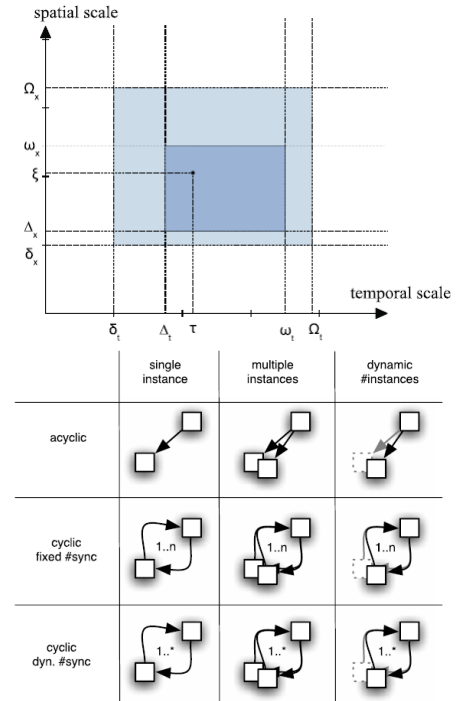
# Усвоение данных

- Различная природа данных
  - Данные измерений, модельные данные, оценки
- Усвоение – процедура коррекции модельных данных по данным измерений
- Сценарии использования
  - Коррекция входных/выходных данных
  - Идентификация параметров (в том числе функциональных)
  - Коррекция процедур регрессии/классификации
- Динамические и статические режимы (адаптация/эволюция или идентификация)



# Многомасштабная гибридизация

- Разнородные модели для масштабов
  - Многоагентные модели
  - Суррогатные модели: на уровень (масштаб) выше, на уровень ниже, «в сторону» – граничные условия, во времени, в группе
- Необходимы средства представления суррогатных моделей



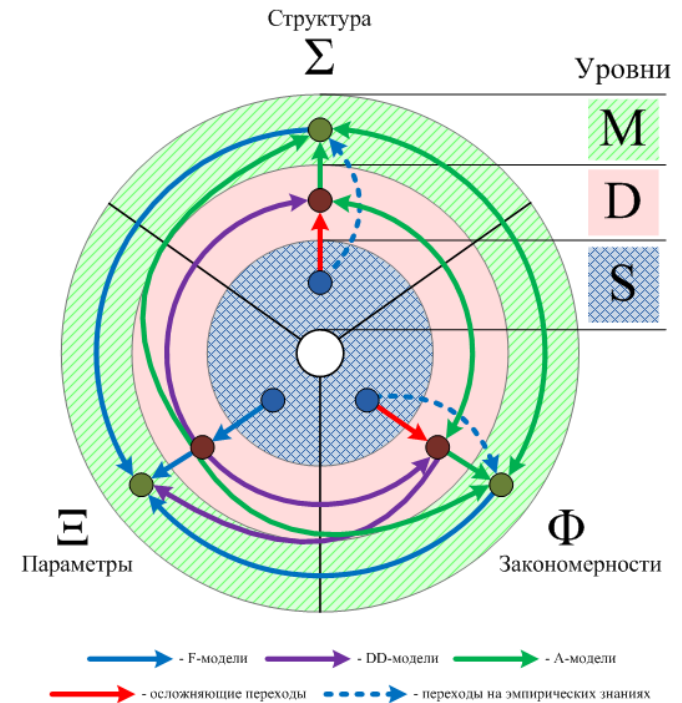
Borgdorff J. [et al.] *J. of Par. and Distr. Comp.*, 73(4), 2013, 465-483.

# Инструменты гибридизации

- Совместное использование разнородных моделей
  - Альтернативы: ансамблевое моделирование
  - Композиция: традиционные КП, суррогатные модели
- Совместное использование разнородных данных
  - Альтернативы: ансамблевое представление
  - Различная природа: усвоение
- Совмещение различных масштабов
  - Композиция: многомасштабное моделирование
  - «Не-замкнутость» системы: суррогатные модели

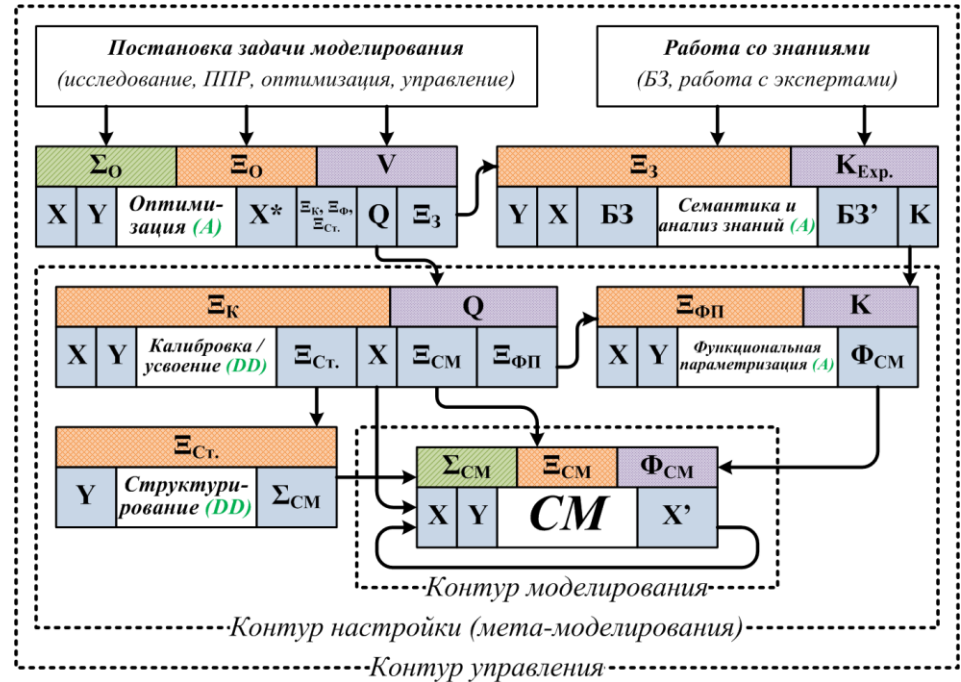
# Сложные модели

- Анализ на уровнях исследуемой системы, доступных данных и модели
- Идентификация структурных и функциональных параметров – затруднительна
- Работа с количественными параметрами – калибровка моделей и усвоение данных
- Результат: систематизация мета-модельных операций



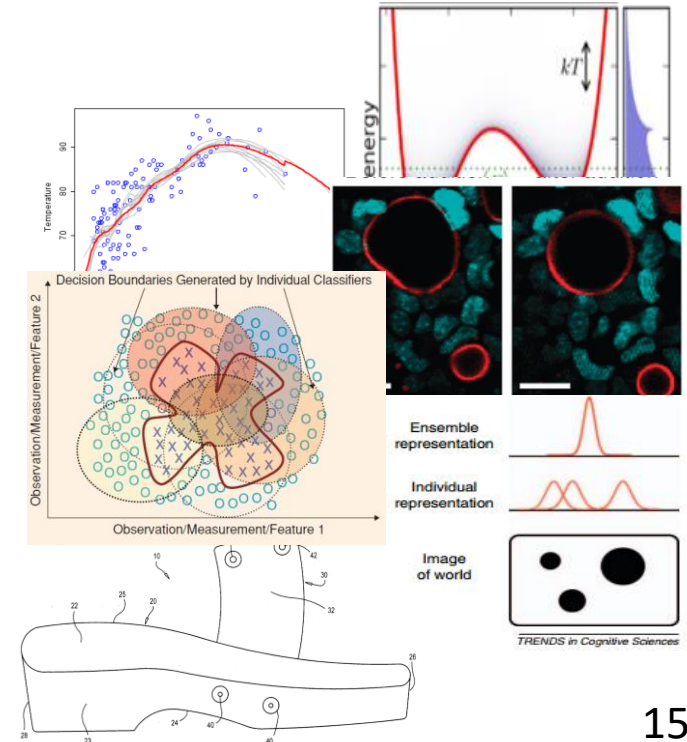
# Сложные модели

- Мета-модельные операции в составе иерархической сложной модели
- Контур моделирование – применение модели
- Контур настройки (мета-моделирования) – манипуляции с параметрами
- Контур управление – прикладное использование моделей и работа со знаниями



# Ансамблевое моделирование

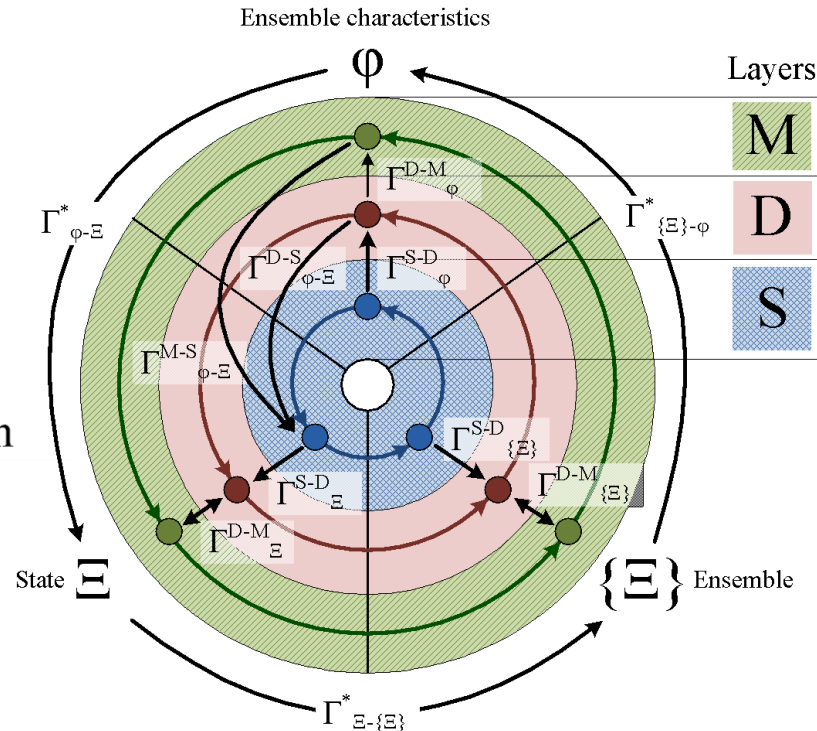
- Статистическая **физика** – набор всевозможных состояний системы, отвечающий определенным критериям
- **Моделирование:** климат, микробиология, моделирование лесных пожаров и пр.
- **Машинное обучение:** комбинация слабых классификаторов для получения «сильного»
- Экзотические области: например, когнитивная визуализация



# Цикл эволюции ансамбля

## Layer basic operations

- $\Gamma_1$  - diversity creation
- $\Gamma_2$  - ensemble analysis
- $\Gamma_3$  - ensemble aggregation

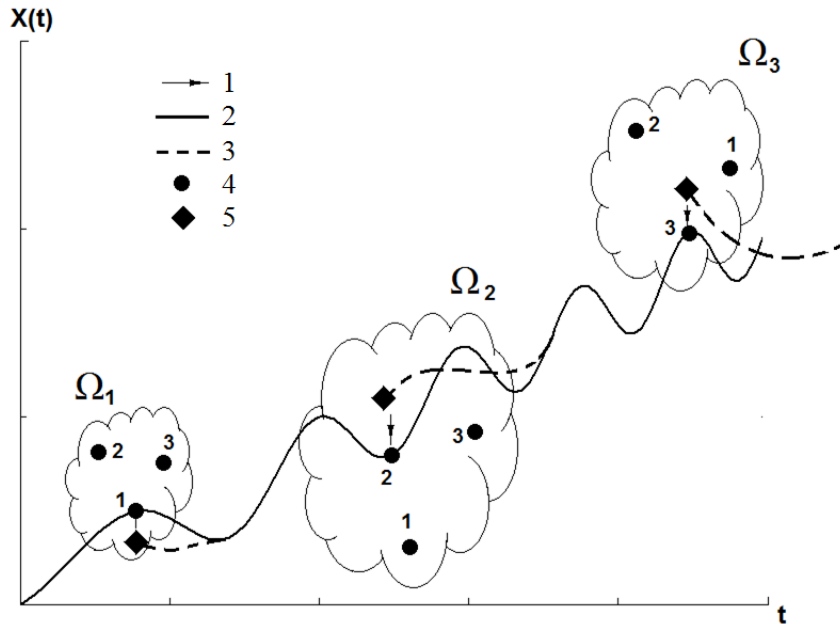


## Inter-layer operations

- $\Gamma_4$  - assimilation
- $\Gamma_5$  - model application
- $\Gamma_6$  - assessment
- $\Gamma_7$  - state refinement



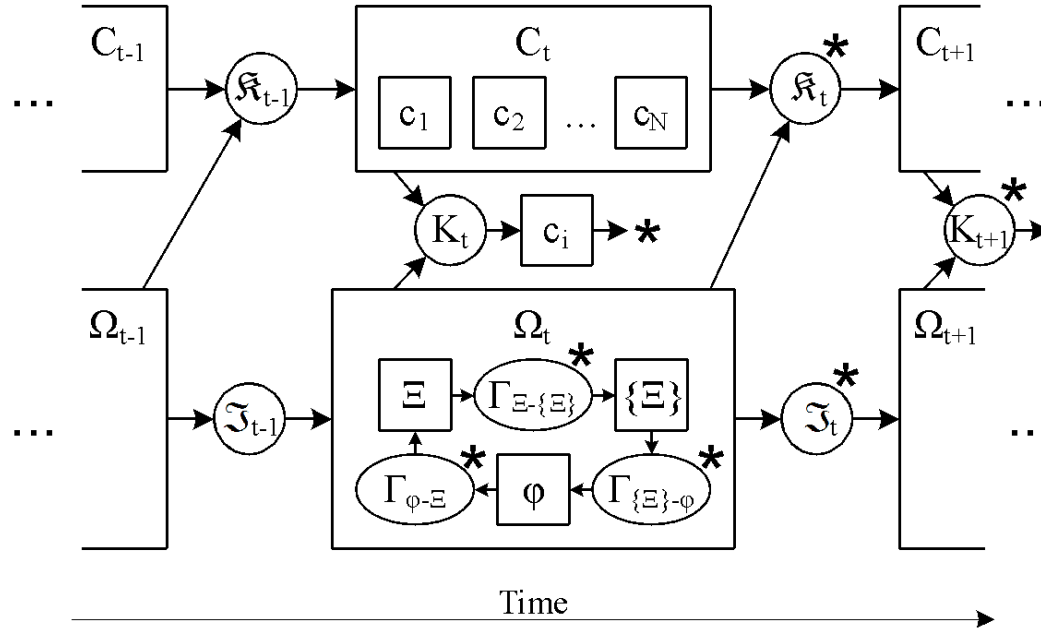
# Цикл эволюции ансамбля



$$\left. \begin{aligned} X_{t+1} &= \Phi(A, \varepsilon) X_t \\ Y_{t+1} &= \Psi(B, \delta) Y_t \\ \Omega_{t+1} &= \Theta(C, \mathcal{G}) \Omega_t \end{aligned} \right\} \Rightarrow$$

$$\left\{ \begin{aligned} A_{t+1} &= \langle \mathfrak{I}_A \Omega_t \rangle, & \varepsilon_{t+1} &= G_{XY}(Y_t, \Omega_t) \varepsilon_t; \\ B_{t+1} &= \langle \mathfrak{I}_B \Omega_t \rangle, & \delta_{t+1} &= G_Y(\Omega_t) \delta_t; \\ C_{t+1} &= F(Z_t), & \mathcal{G}_{t+1} &\in \Omega_Z. \end{aligned} \right.$$

# Обобщенная классификация в ансамблевом прогнозировании



\* - possible influence of class selection

# A1. Декомпозиционный ансамбль

- Принципы применения
  - Идентификация моделей компонентов в составе комплексной модели
  - Работа с объектами большего или меньшего масштаба (по сравнению с расчетной сеткой): вихри на поверхности воды, облака
- Варианты решения
  - Параметризация внутренних моделей
  - Внедрение новых моделей
- Технологии
  - Модификация кода
  - Композиция

## A2. Ансамбль с вариацией моделей

- Принципы применения
  - Различные реализации моделей
  - Различные параметры (сетки, под-модели и пр.)
  - Использование суррогатных моделей (как отдельно, так и с классическими)
- Варианты решения
  - Регрессия/классификация моделей
- Технологии
  - Агрегация данных
  - Композиция

## А3. Ансамбль с вариацией данных

- Принципы применения
  - Анализ и управление неопределенностью данных и моделей
  - Внесение возмущений во входные данные
  - Анализ распределения выходных данных ансамбля
- Варианты решения
  - Регрессия/классификация моделей
- Технологии
  - Агрегация данных
  - Композиция
  - Анализ больших данных

## A4. Ансамбль состояний системы

- Принципы применения
  - Состояние исследуемой системы
  - Состояние системы моделей
- Варианты решения
  - Суррогатные модели
  - Классификация состояний системы
- Технологии
  - Композиция
  - Дополнительная обработка данных

## A5. Мета-ансамбль

- Принципы применения
  - Работа с различными вариантами ансамбля
  - Высокоуровневая модификация моделей
- Варианты решения
  - Суррогатные модели
  - Регрессия/классификация на уровне моделей
  - Специализированные методы
- Технологии
  - Композиция
  - Дополнительная обработка данных

# Пример декомпозиционного ансамбля

**Задача:** определение коэффициента передачи энергии ветра волнам

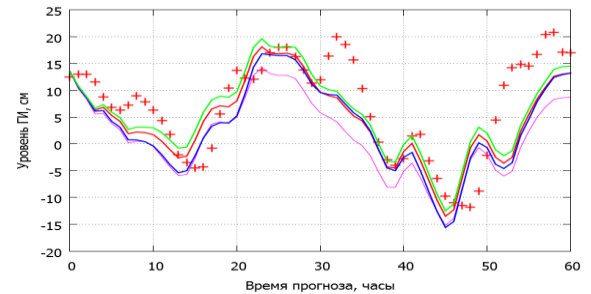
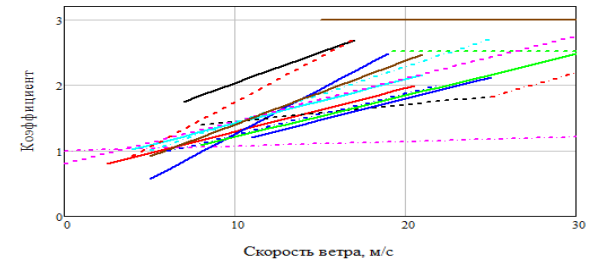
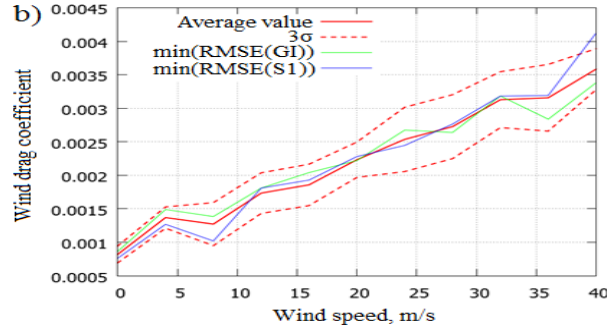
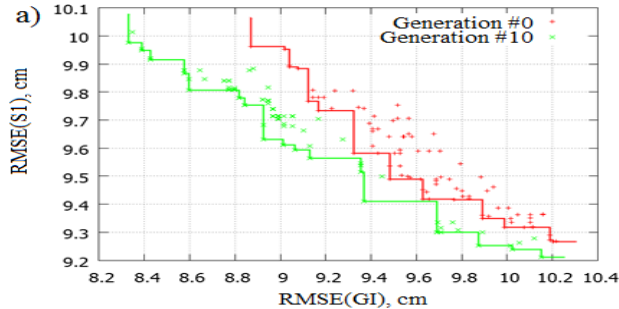
Природа возникновения ветровых волн – трудно формализуема

Разнообразие эмпирических формул

Пространственная и временная изменчивость: зависимость от места и ситуации

**Решение:** ансамбль реализаций на базе ГА (популяция – ансамбль)

- Многокритериальная оптимизация
- Учет особенностей акватории
- Учет особенностей модели (две модели: Balt-P и BSM)

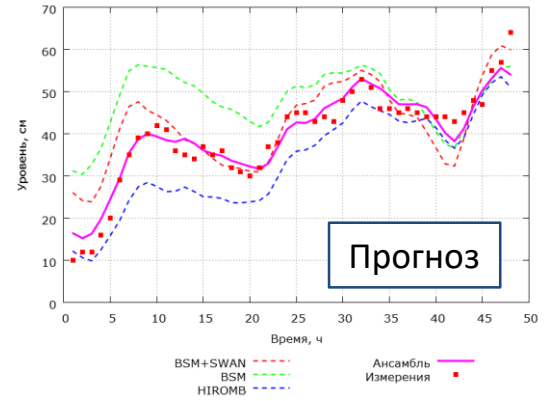
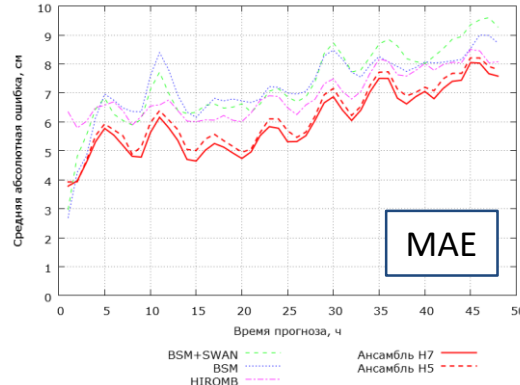
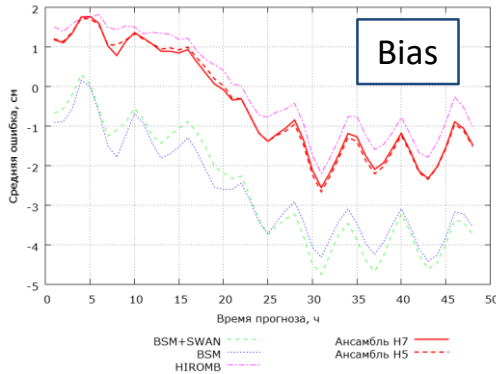
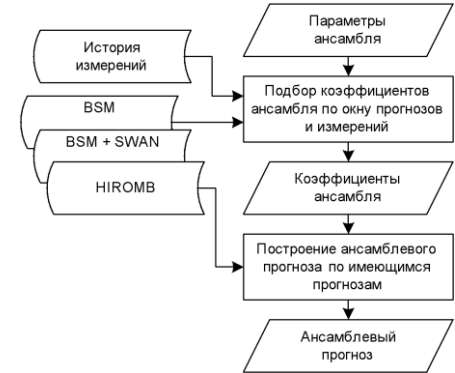


Измерения +  
BSM-BS (RMSE<sub>ГИ</sub>: 8.1 см)  
Среднее по ансамблю (RMSE<sub>ГИ</sub>: 6.2 см)  
#32 - min(C1) (RMSE<sub>ГИ</sub>: 6.5 см)  
#93 - min(ГИ) (RMSE<sub>ГИ</sub>: 5.8 см)



# Пример ансамбля на альтернативных моделях

**Задача:** прогнозирование уровня Балтийского моря  
 Множество источников метеопрогнозов (ветер, давление)  
 Множество моделей уровня (внешние – источники данных, внутренние – программные модули)  
 Дополнительно: данные измерений для проведения усвоения  
 Простейший вариант: линейная регрессия



# Учет объектов предметной области

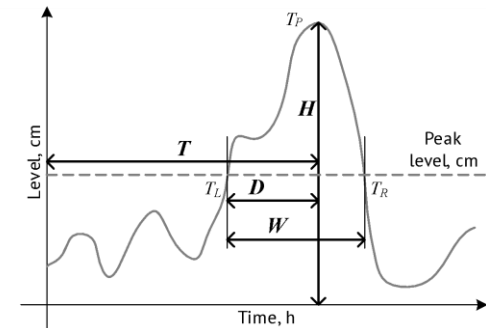
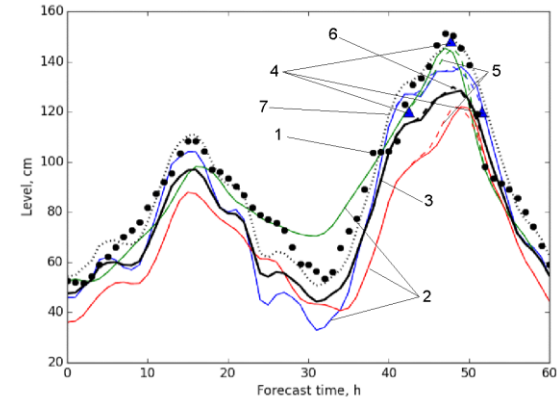
**Задача:** модификация структуры ансамблевого прогноза с учетом прогноза пика наводнения

Необходимо для выработки плана маневрирования затворами

“Сглаживание” пика при использовании линейной комбинации прогнозов

**Решение:** параметрический ансамбль для контроля структуры пика

- Двухуровневый ансамбль
- Учет особенностей моделей



# Выбор класса ансамбля

**Задача:** предсказание оптимальной структуры ансамбля

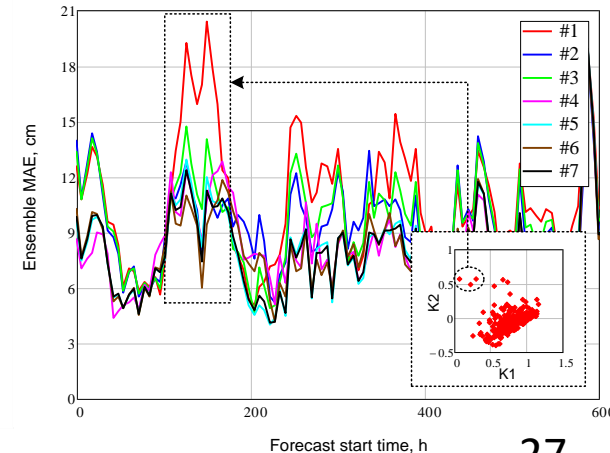
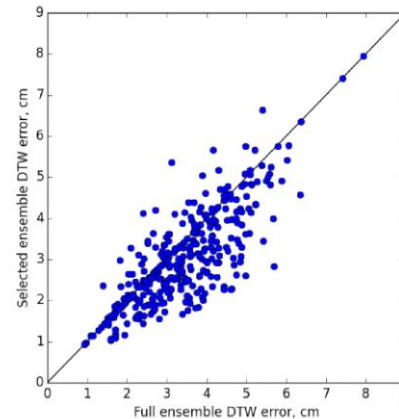
Повышение качества прогноза за счет исключения элементов с низким качеством

Усложнение формы регрессионного соотношения

**Решение:** выбор лучшего варианта регрессионного соотношения

- Использование символьной регрессии для идентификации возможных структур ансамбля
- Выбор лучшей структуры модели в рамках классификации
- Моделирование переходов между классами

	Individual	#6	$(C0 \times h \times b - C1)$
#0	$C0$	#7	$(C0 \times b - C1 \times h)$
#1	$b$	#8	$(C0 - C1 \times \max(b, h))$
#2	$h$	#9	$((C0 \times b - C1 \times h) - C2)$
#3	$\max(b, h)$	#10	$(C0 \times \max(C1 \times b, C2 \times h) - C3)$
#4	$(C0 \times b - C1)$	#11	$((\max(h, b) - \min(C0 \times \max(b, h), C1)) - C2)$
#5	$(C0 \times h - C1)$	#12	$((C0 \times \frac{(C1 \times h - C2)}{(b - 100.000)} - C3) - \max(C4 \times b, C5 \times h))$

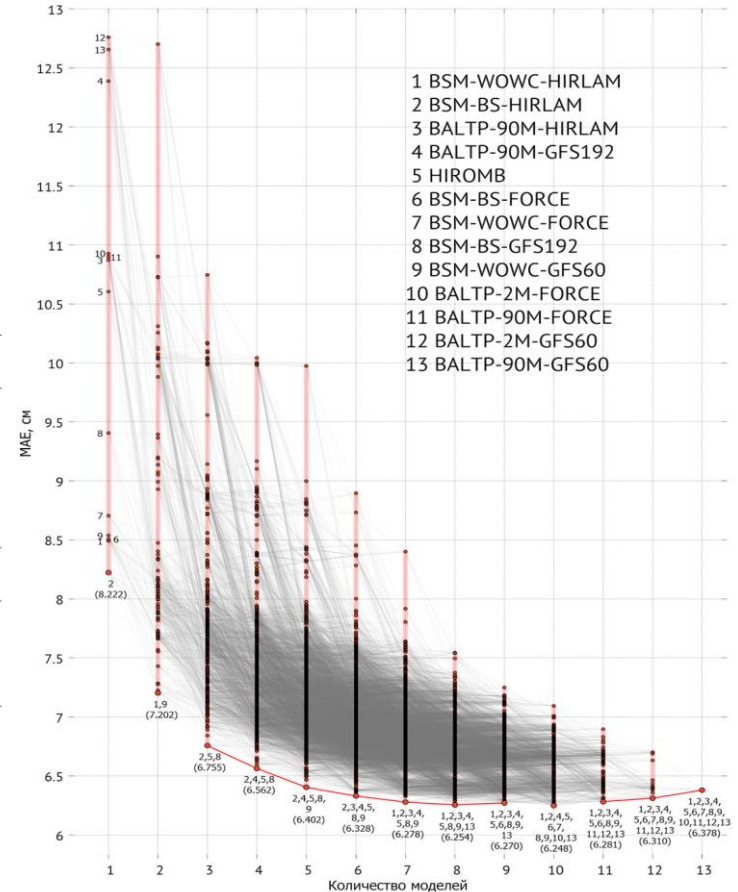
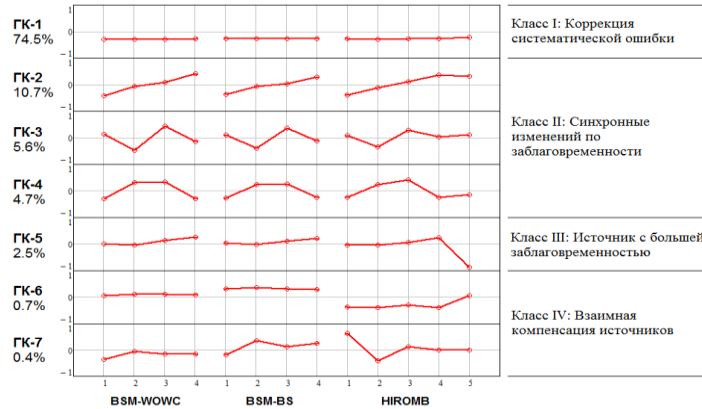


# Мета-ансамбль

**Задача:** развитие ансамбля моделей уровня моря

**Особенность:** слишком большое количество моделей приводит к проблеме мультиколлинеарности

**Решение:** применение регрессии на главные компоненты



# Моделирование сложных сетей

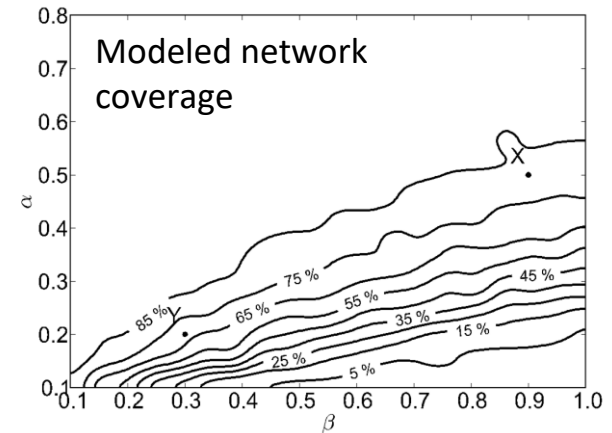
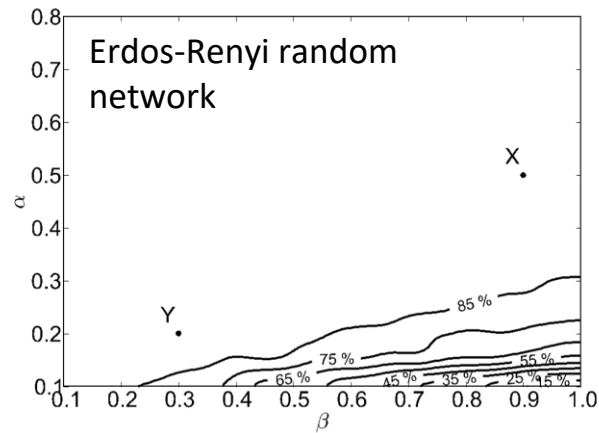
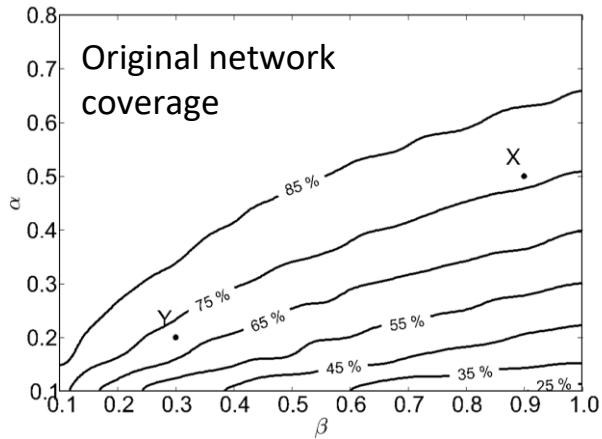
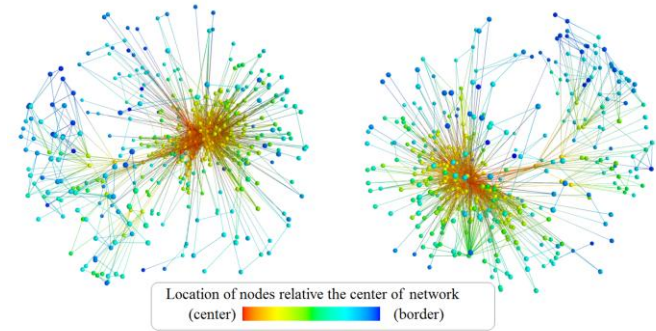
Идентификация сетей по функциональным характеристикам

**Задача:** определение структуры сетевой модели аэропортов США по распространению инфекции

Ансамбль: эволюционирующий ансамбль сетевых структур

Гибридизация: модель распространения инфекции и оптимизация сети

Усвоение: идентификация параметров сети по наблюдению процесса в этой сети

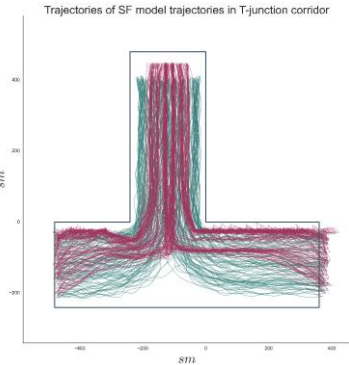
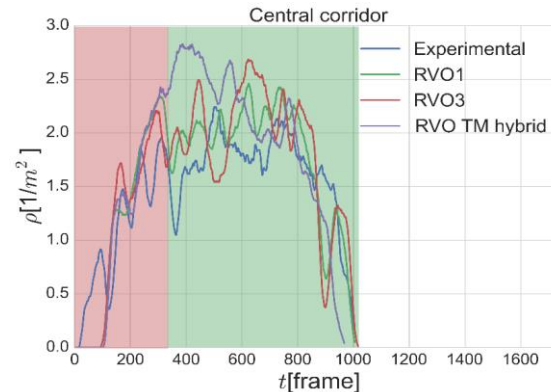
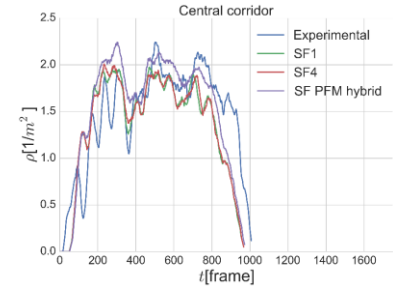
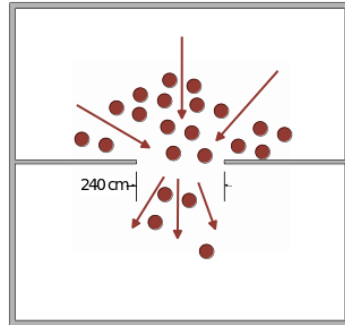
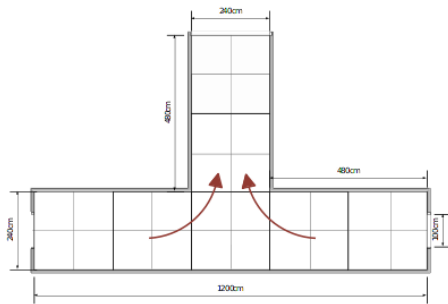


# Построение многоагентных моделей

Построение ансамблей многоагентных моделей

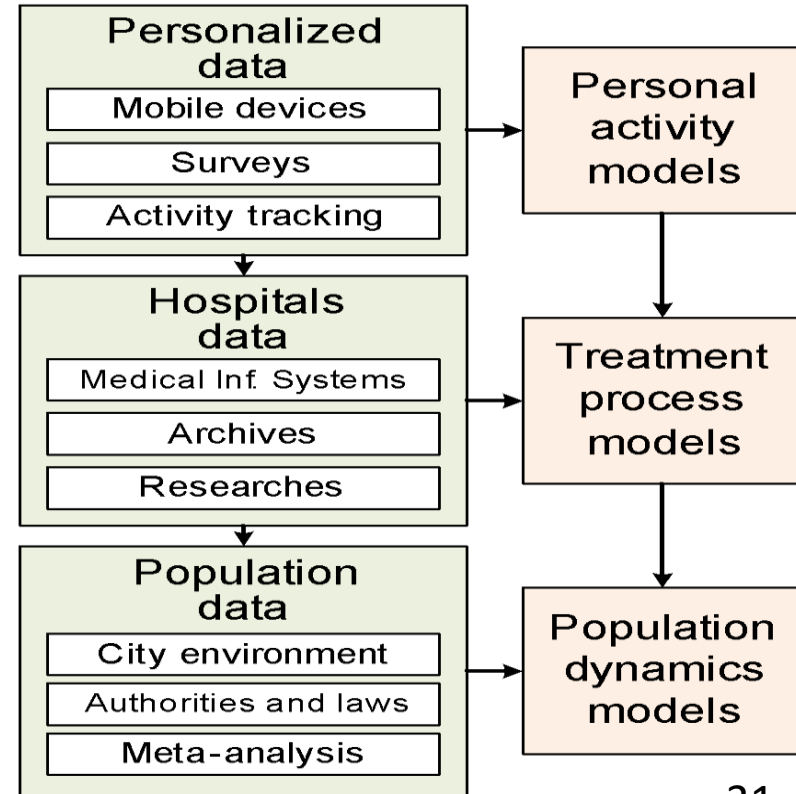
**Задача:** построение комбинирующих правил для микромоделей движения толпы  
 Ансамбль альтернативных модели перемещения (RVO, Social Forces), различные параметры

**Решение:** Козволюционная калибровка моделей. Комбинация моделей: по пространству, по времени, по агентам.

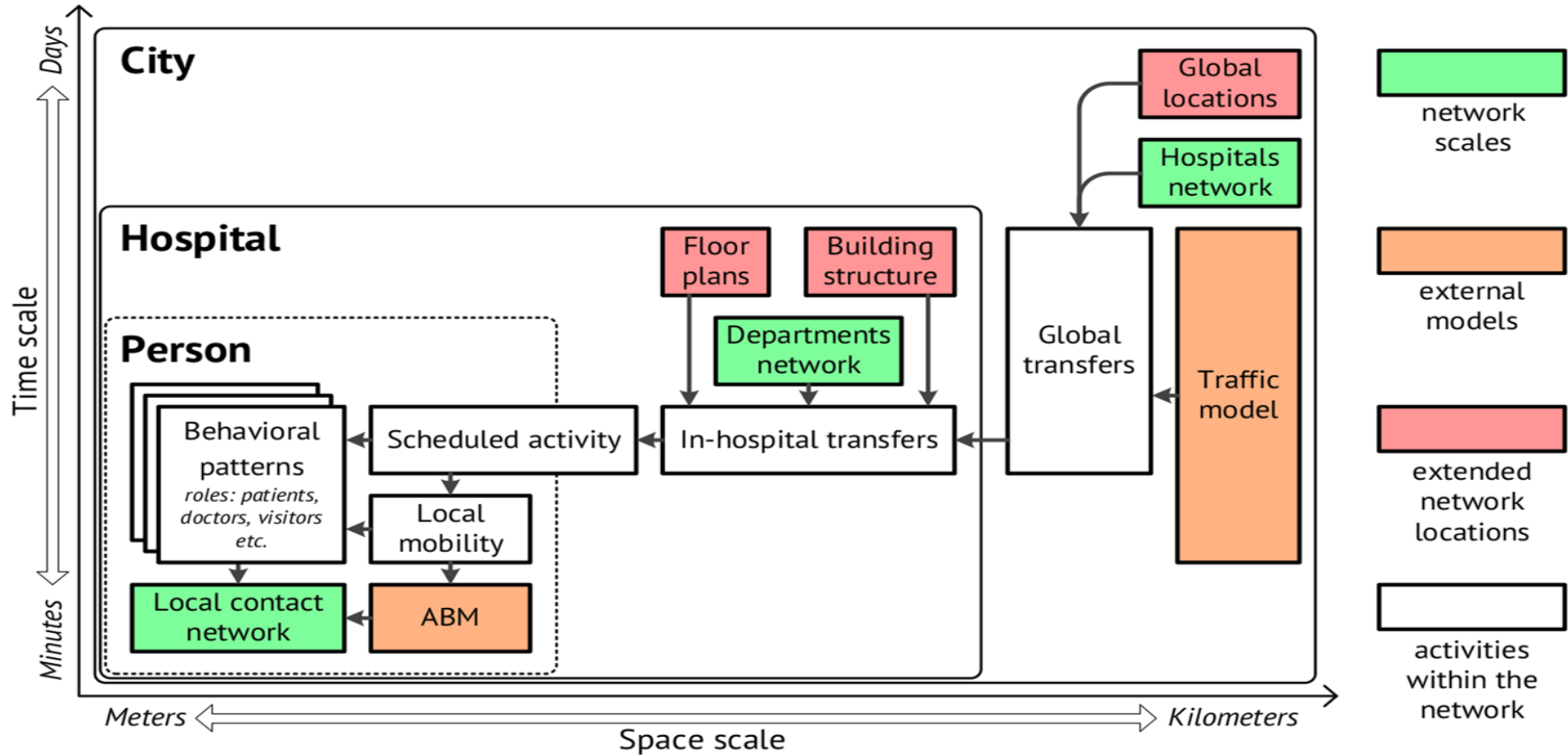


# Моделирование процессов в медицине и здравоохранении

- Высокая неопределенность
- Интеграция разнородных моделей и источников данных
- Множество масштабов рассматриваемых систем
- Наличие различных сценариев оказания медицинской помощи
- Разнородных поток пациентов и ценностно-ориентированный подход



# Многомасштабное моделирование сервисов здравоохранения

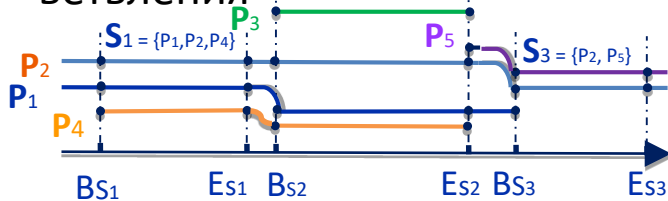




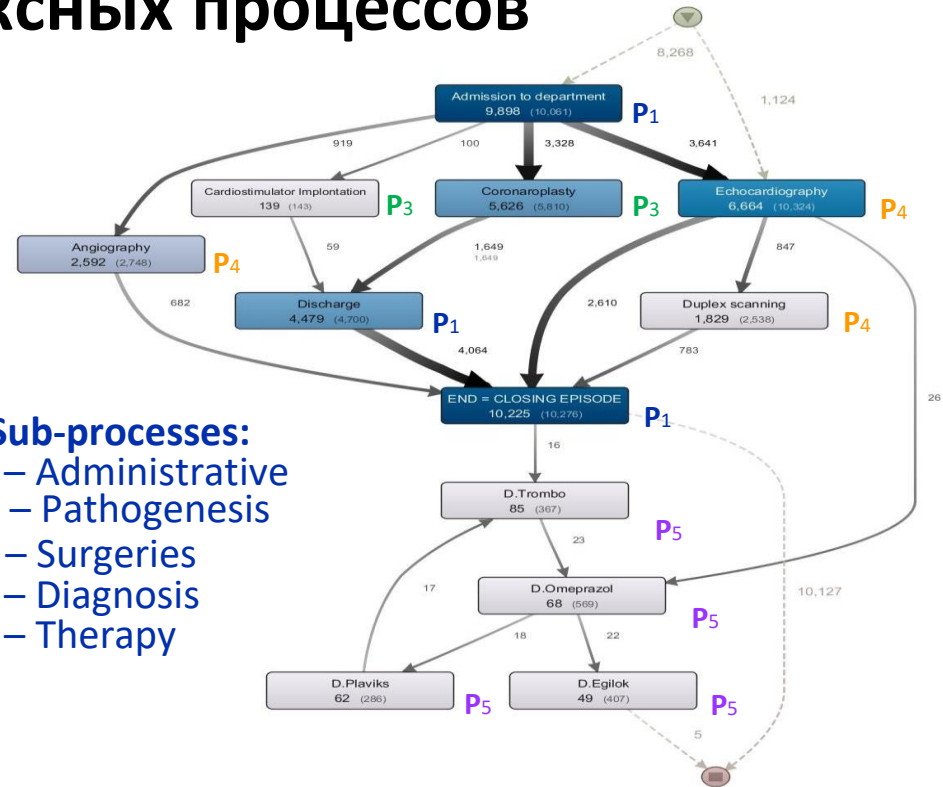
# Идентификация комплексных процессов

- Применение технологий process mining для идентификации структуры модели процесса
- Комбинация и синхронизация множества под-процессов
- Классы процессов и под-процессов
- Проблема предсказания

ветвления

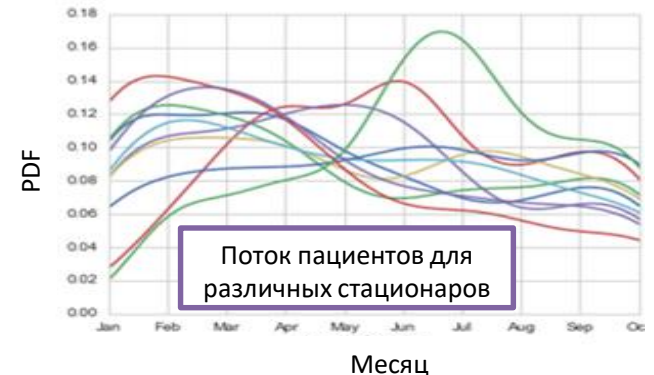
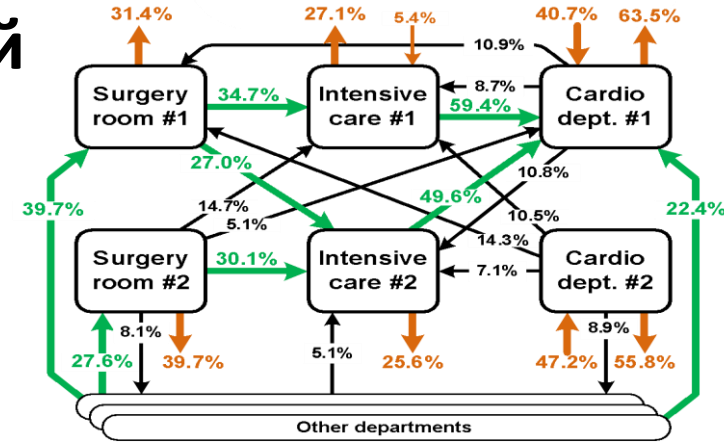
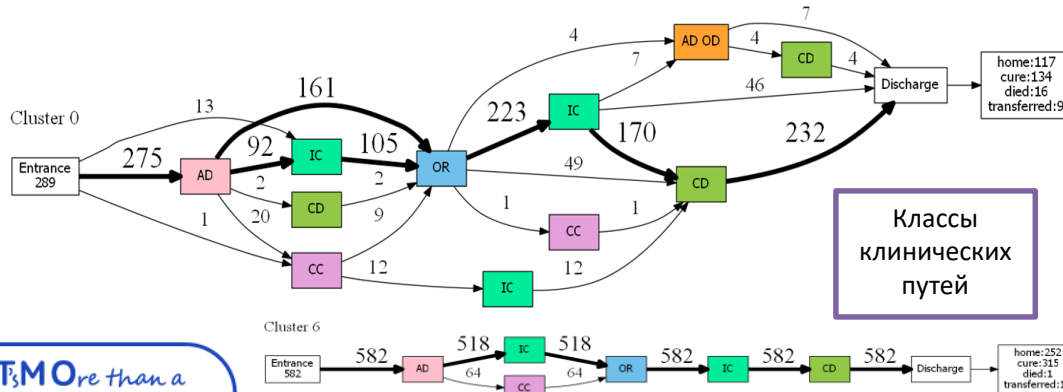


**Sub-processes:**  
**P<sub>1</sub>** – Administrative  
**P<sub>2</sub>** – Pathogenesis  
**P<sub>3</sub>** – Surgeries  
**P<sub>4</sub>** – Diagnosis  
**P<sub>5</sub>** – Therapy



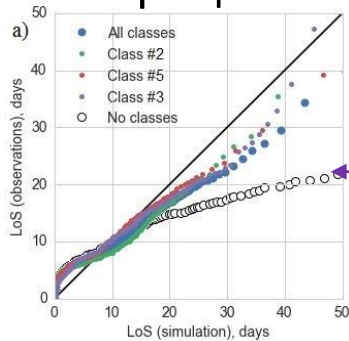
# Исследование клинических путей

- Идентификация клинических путей по данным медицинских карт
- Расширение и коррекция журнала событий
- Оценка близости клинических путей
- Классификация клинических путей



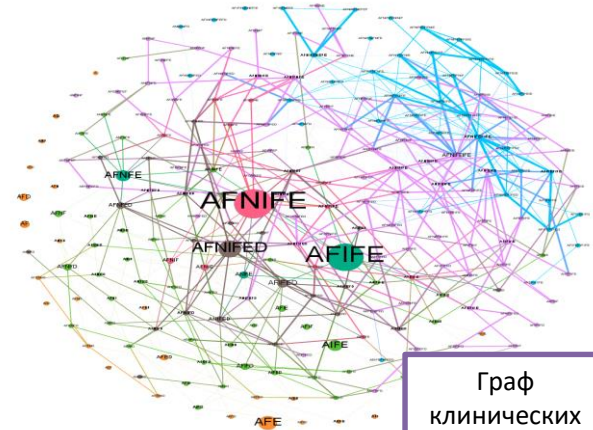
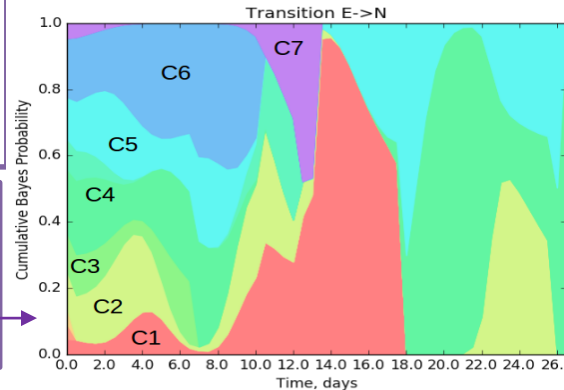
# Моделирование клинических путей

- Предсказание классов и переходов
- Имитационное моделирование разнородного потока пациентов
- Исследование пространства возможных процессов

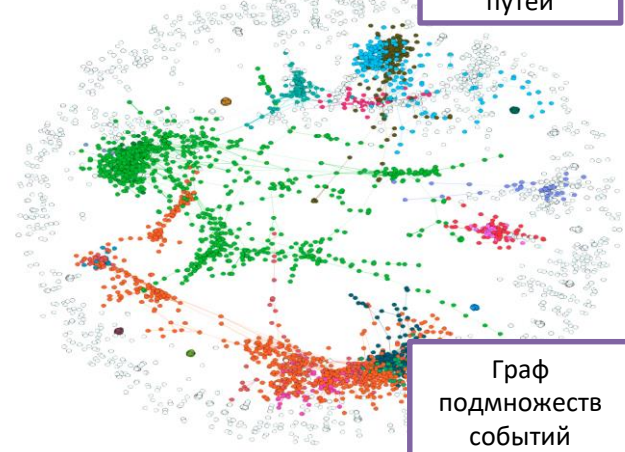


Дискретно-событийное моделирование потока пациентов

Предсказание класса по событиям и времени пребывания

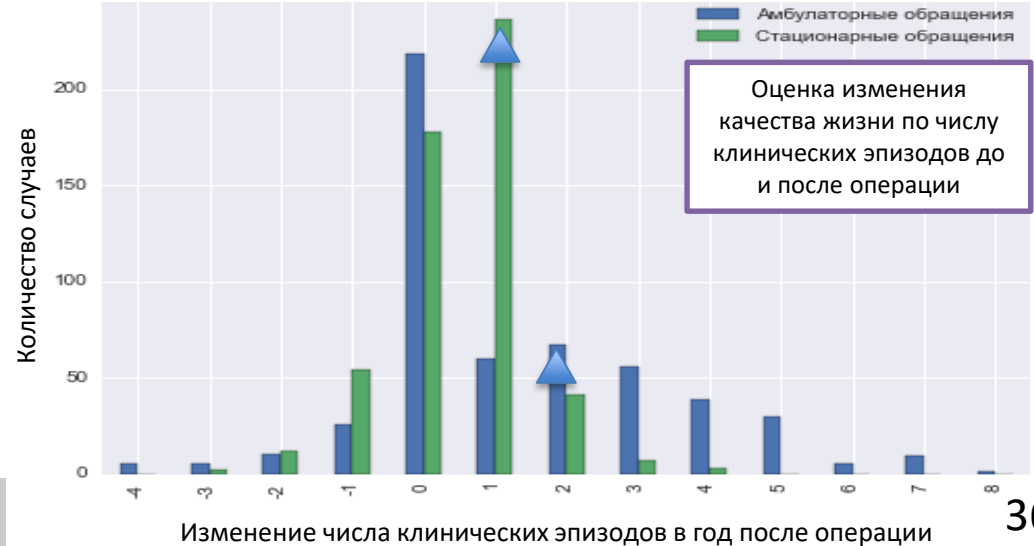
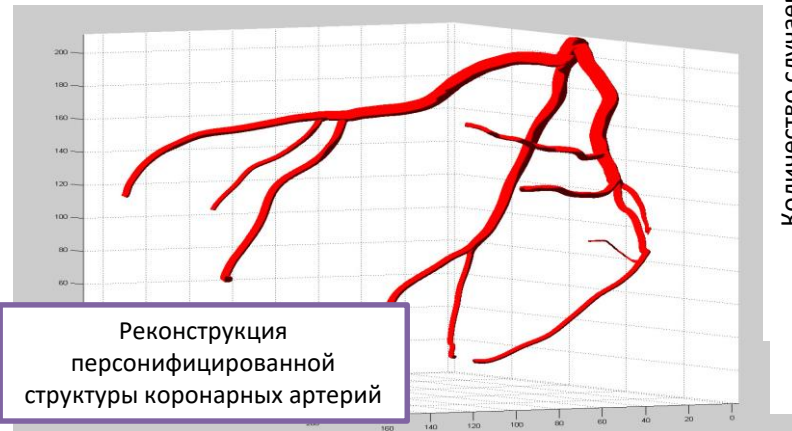
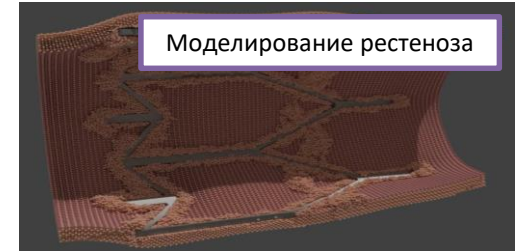


Граф клинических путей



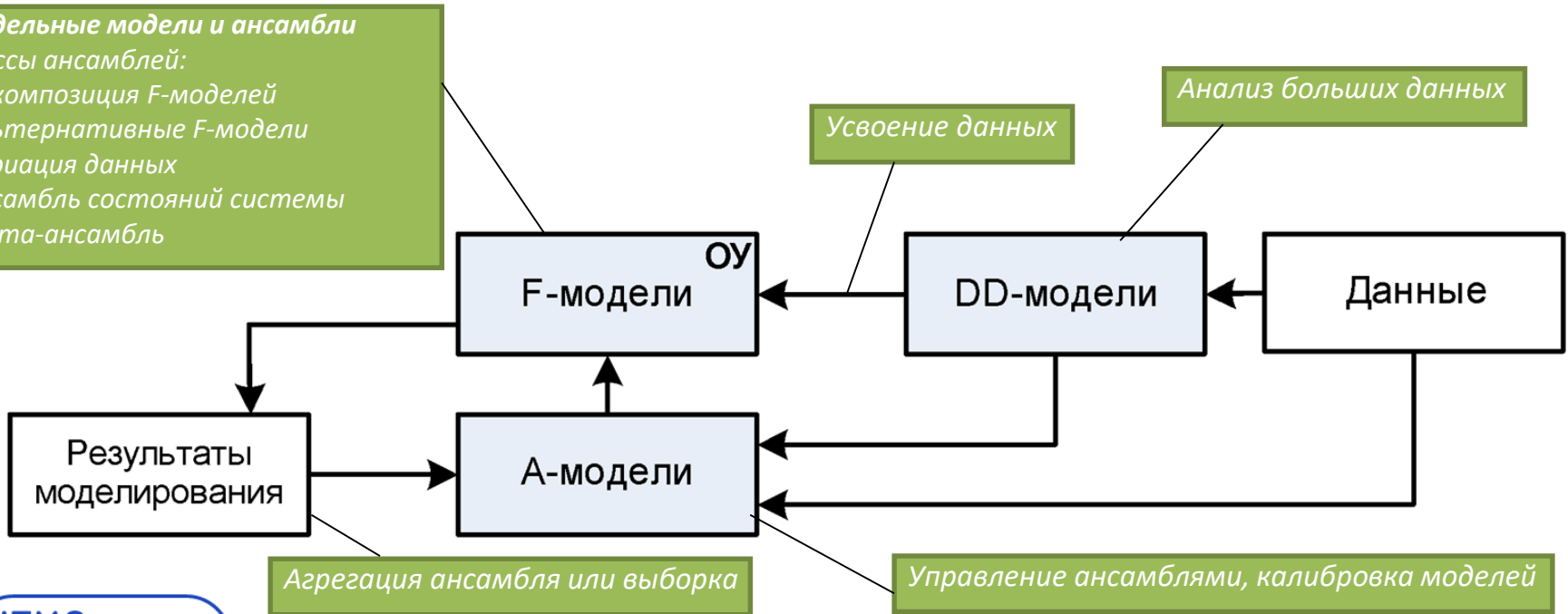
Граф подмножеств событий

# Многомасштабное моделирование гемодинамики и стентирования

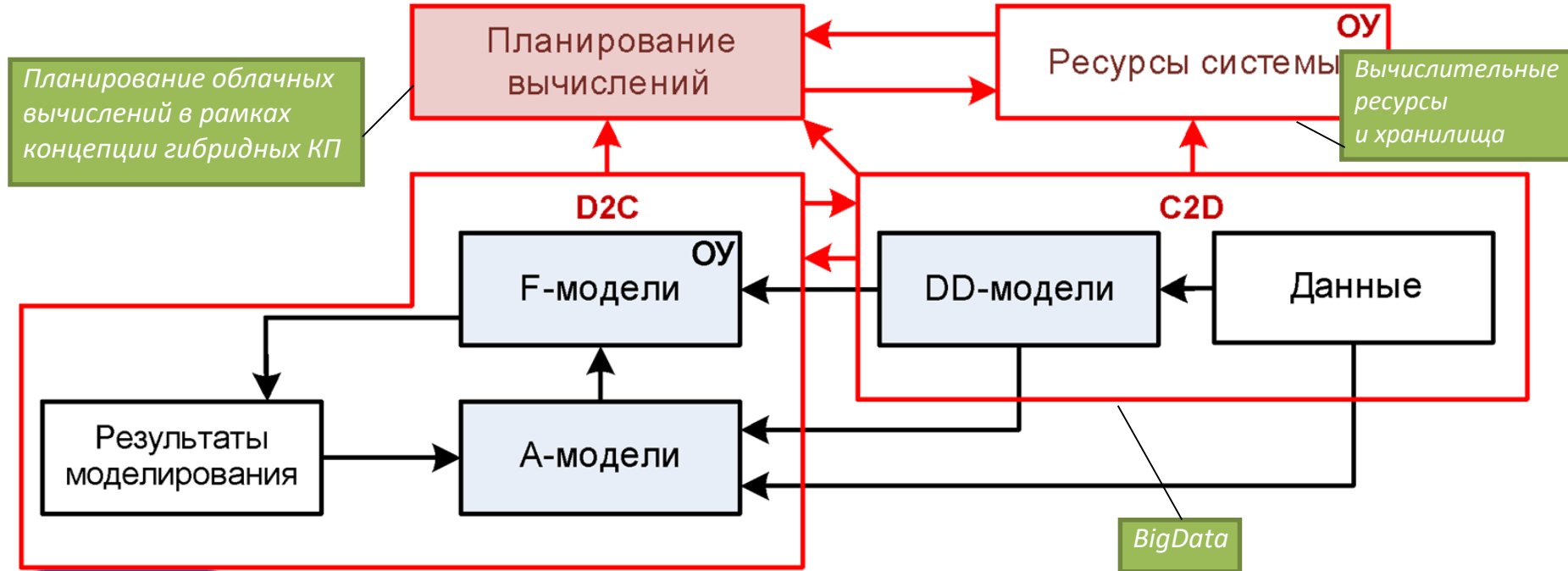


# Управление в задачах гибридного моделирования (1/3)

- Отдельные модели и ансамбли*  
 Классы ансамблей:
- Декомпозиция F-моделей
  - Альтернативные F-модели
  - Вариация данных
  - Ансамбль состояний системы
  - Мета-ансамбль

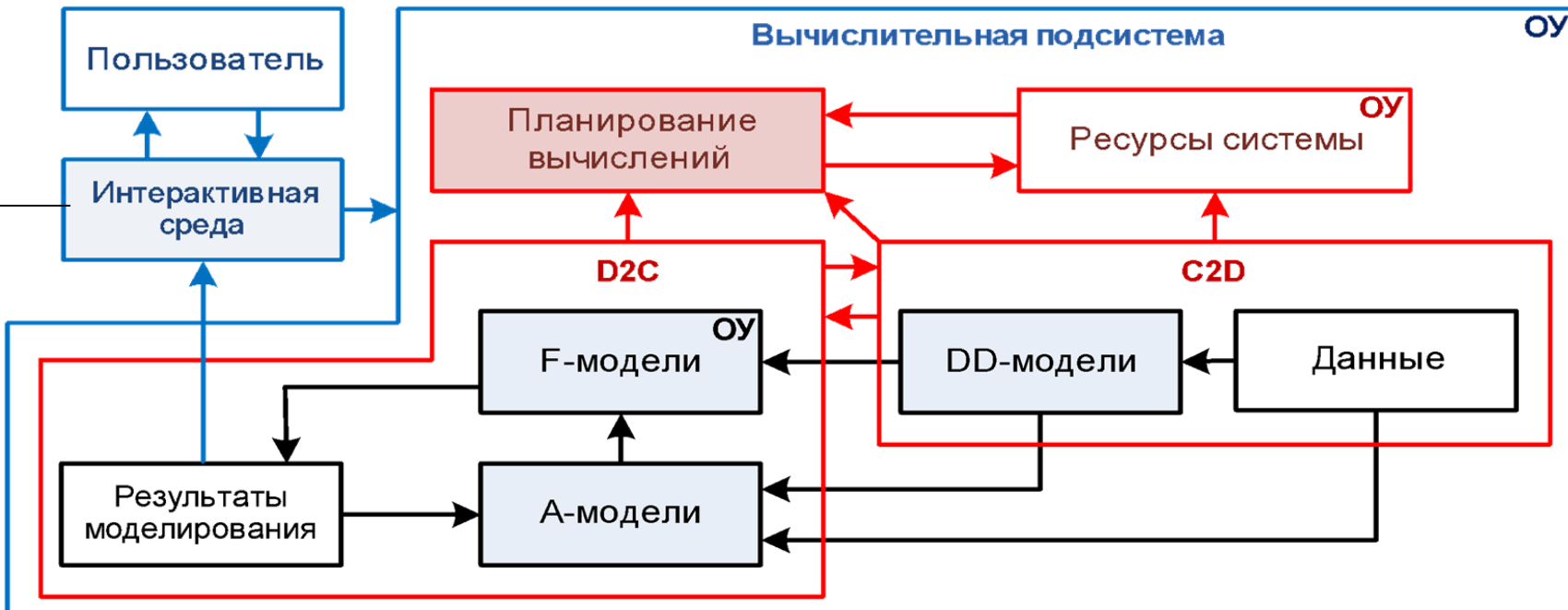


## Управление в задачах гибридного моделирования (2/3)



## Управление в задачах гибридного моделирования (2/3)

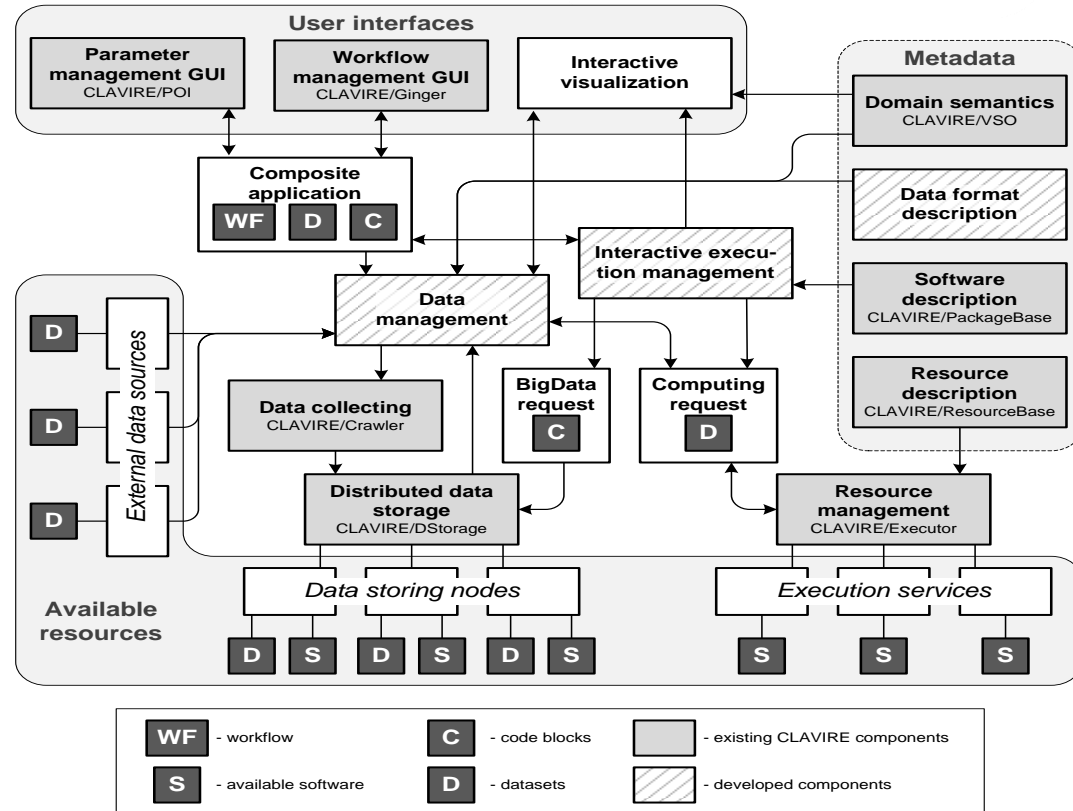
Человеко-компьютерное взаимодействие



# Гибридные композитные приложения

Развитие структуры КП

- Бесшовная интеграция с задачами анализа BigData
  - Интеграция средств интерактивной работы с пользователем
  - Развитие средств ансамблевого моделирования
- Развитие инфраструктуры
- Семантическая интеграция разнородных источников данных
  - Управление локальными пакетами на узлах хранилища





# Интеллектуальные средства построения композитных приложений

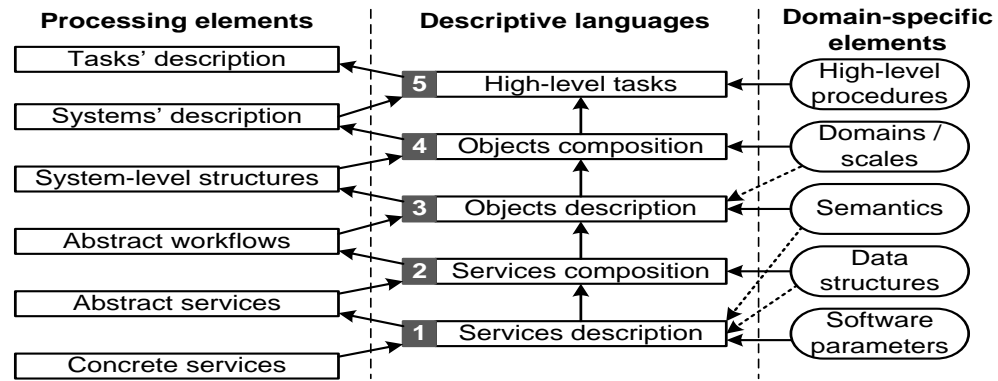
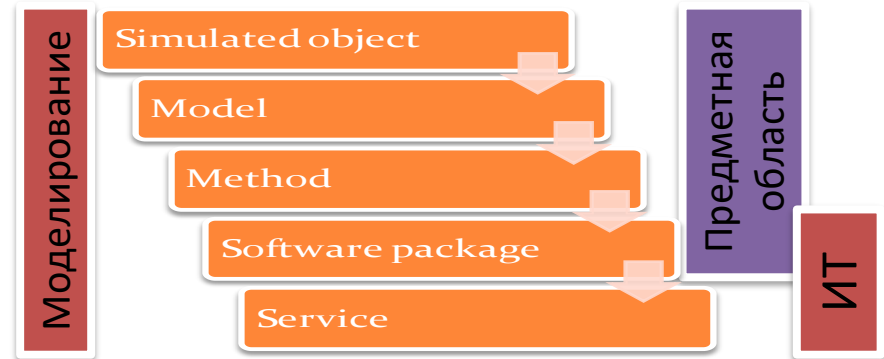
Три **области знаний**: предметная область, ИТ, принципы моделирования

Пять **базовых концептов** знаний, объединенных в иерархию

Два подхода к **формализации**: DSL и знание-ориентированные технологии

Пять **уровней** описания задач

Два подхода к **интерпретации**: описание (со стороны пользователя), интерпретация (со стороны системы)



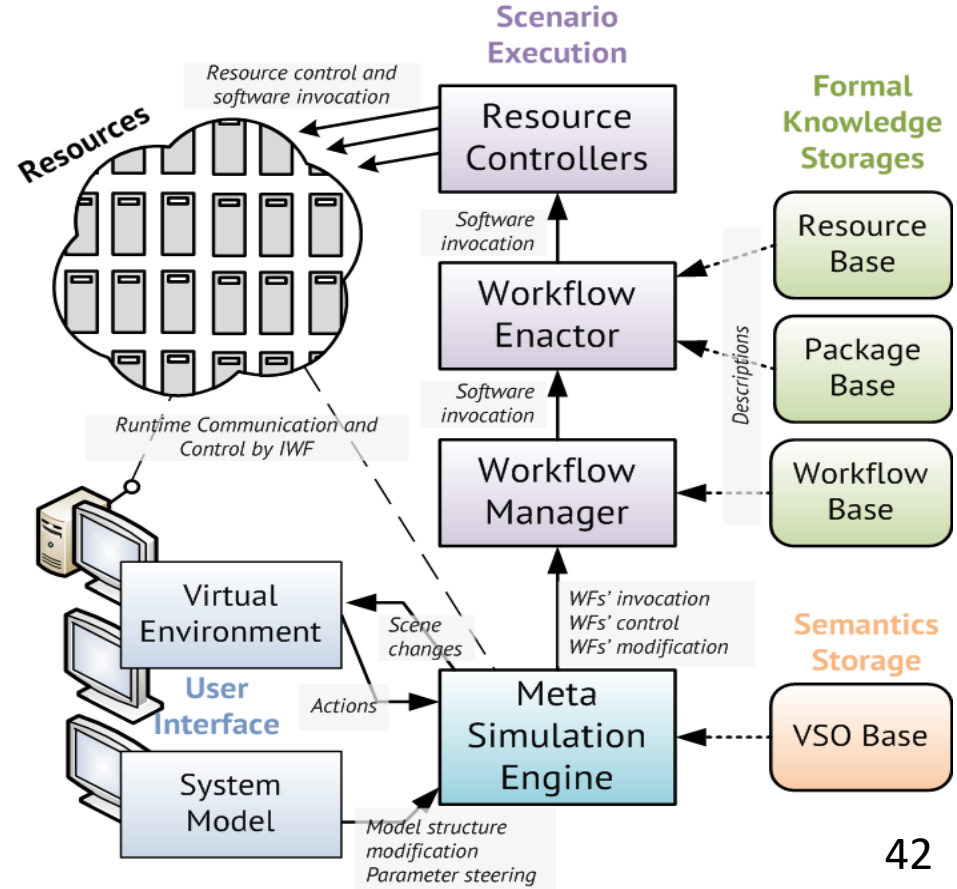
# Интерактивное ансамблевое моделирование

Поиск и агрегация в архивах данных

Формирование и эволюция ансамбля

Усвоение и контроль моделирования в соответствии с архивными данными

Интерактивное управление составом ансамбля (предсказание, аппроксимация запросов)



# Заключение

Гибридное моделирование ориентировано на совмещение разнородных объектов (моделей, данных) в рамках одного решения

Использование интеллектуальных технологий при построении мета-модельных операций позволяет выйти на новый уровень в автоматизации решении прикладных задач моделирования

Ансамблевые подходы, усвоение, суррогатные модели, эволюционные методы могут быть использованы в качестве инструментов гибридного моделирования

Гибридное моделирование и его автоматизация предъявляют специфические требования к инструментальному технологическому обеспечению компьютерного моделирования

**Спасибо за внимание!**

kovalchuk@corp.ifmo.ru  
escience.ifmo.ru

IT'sMO *re than a*  
UNIVERSITY