

НТУУ «КПИ»  
Межуниверситетский медико-инженерный факультет

Национальный институт сердечно-сосудистой хирургии  
им. Н.М. Амосова

Распознавание функциональных  
зависимостей с помощью процедур  
кластерного анализа и их использование  
для идентификации математических  
моделей биологических систем

Е.А. Настенко

Киев 2012

# **Зачем нужно математическое моделирование?**

1. Модели – аппроксимации (класс модели задается априори).
2. Модели, учитывающие ограниченный набор свойств и исследующие их (свойств) проявление.

Моделирование исследует то, что нельзя измерить или получить экспериментально.

# Построена математическая модель. Как выполнить ее идентификацию?

Нужны функциональные характеристики.  
А как их получить, например, для  
организма человека?

- Синтезировать по группе наблюдений ?
  - Получить в результате тестов для конкретного пациента? А насколько это безопасно?

# Проблема статистического анализа данных

Сравнение групп однородных наблюдений.

- Как их получить?
- Что такое «однородность»?

Однородность:

- (i) близость показателей по величине
- (ii) однородность объектов по (виду) функциональной связи между показателями

# **Функциональная связь:**

- Единственная или семейство функциональных характеристик (тогда имеем пространство сниженной размерности, редуцированное)?
  - Если семейство, то многомерная функциональная зависимость?
- Все ли переменные доступны для измерения?

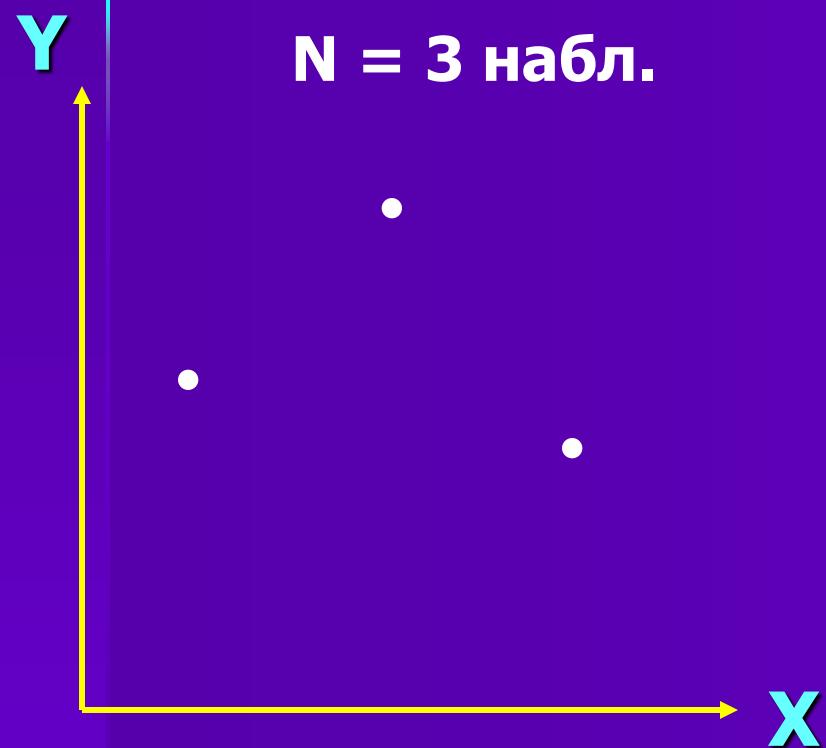
**На сегодня  
распознавание типов  
функциональной связи  
между показателями  
биологических объектов  
является далекой от  
окончательного решения  
задачей**

# Цель

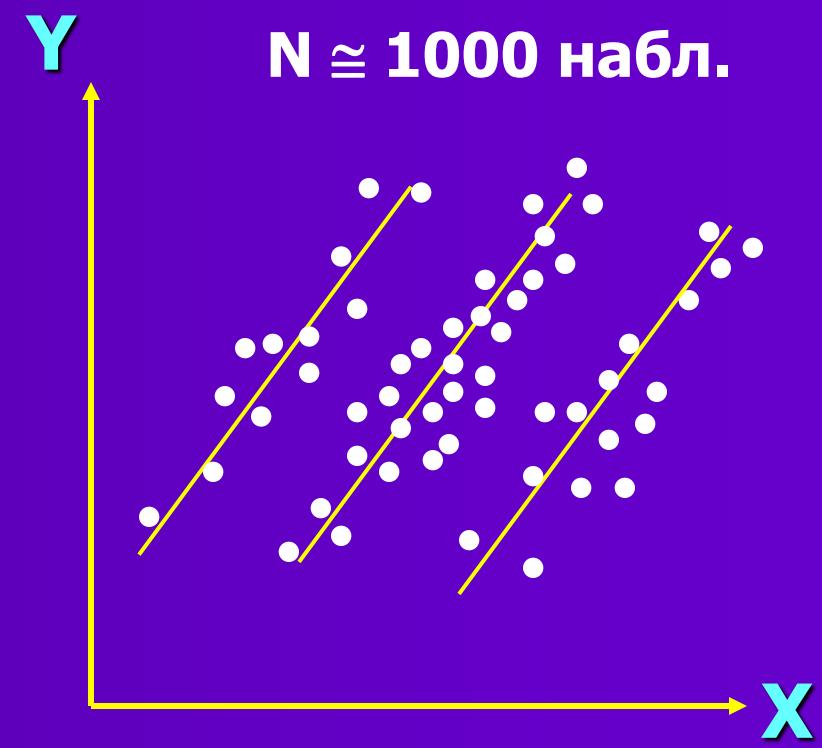
Разработать ИТ  
безопасного восстановления  
(распознавания) функциональных  
связей количественных  
показателей сложных объектов на  
основе специальных алгоритмов  
кластер-анализа

**Объекты биологических и  
медицинских исследований обладают  
рядом специфических свойств,  
которые должны либо приниматься во  
внимание,  
либо непосредственно использоваться  
при обработке биомедицинских  
данных**

# «Упорядочивание» (самоорганизация) наблюдений в пространстве состояний при увеличении числа испытаний



$N = 3$  набл.

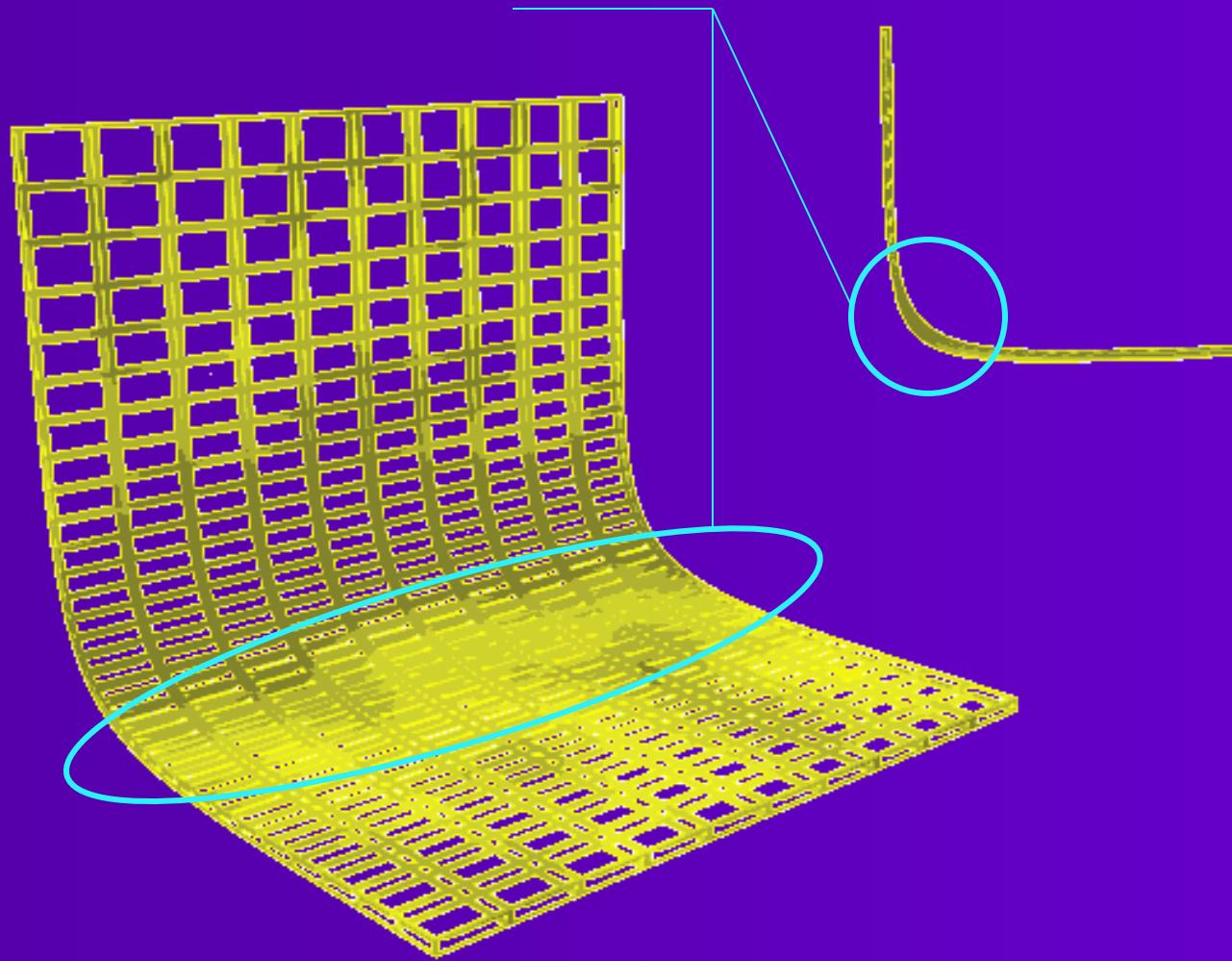


$N \approx 1000$  набл.

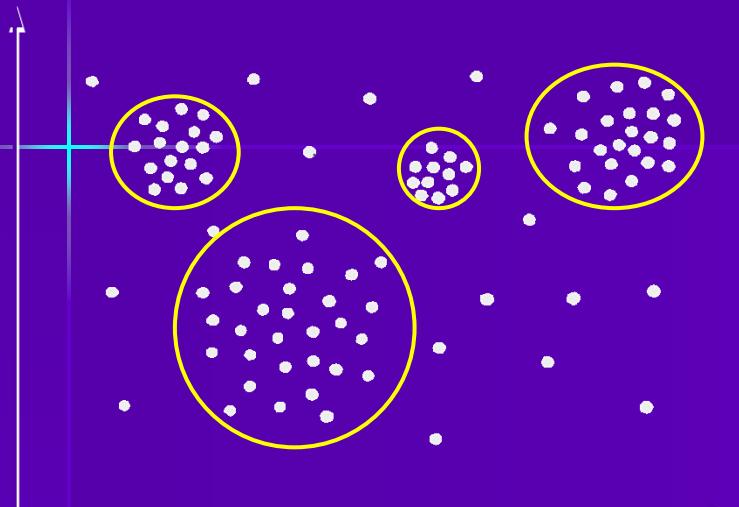


# Проекция трехмерной поверхности на плоскость

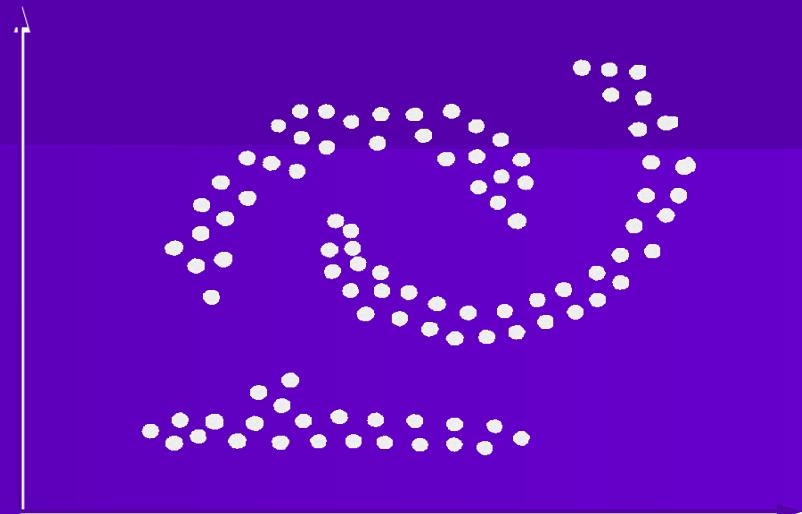
Область наибольшей кривизны



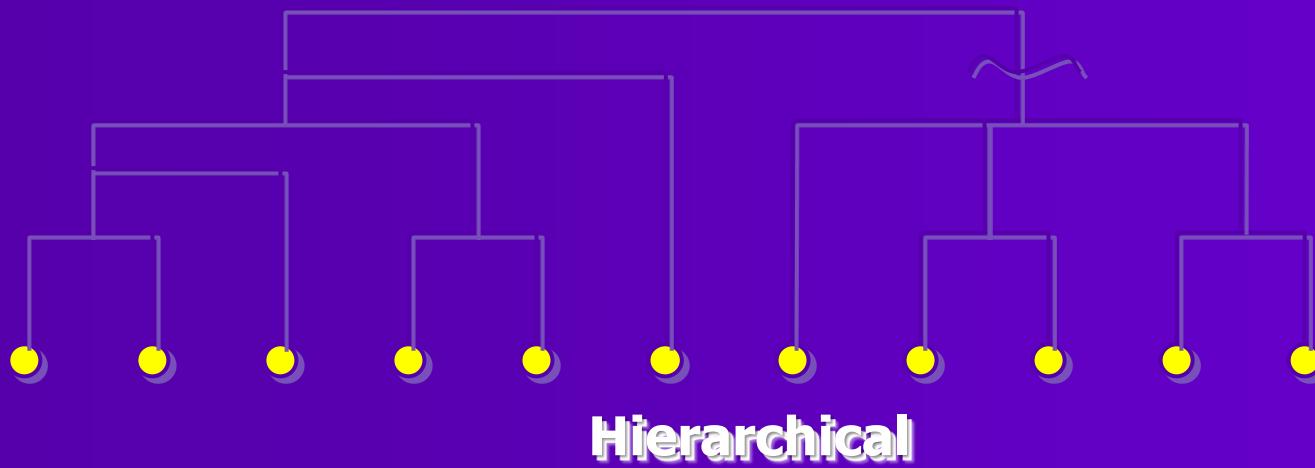
# Basic methods of clustering



**"K-means"**



**KRAB**



**Hierarchical**

# Fibonacci-like algorithm of partitioning of mixtures of multidimensional functional dependencies of complex systems

$$X = \{x_{\cdot,1}; x_{\cdot,2}; \dots; x_{\cdot,m} \}$$

$$x_{k+l,i} = \sum_{j=1}^m (x_{kj,i}) + x_{li} - S_{ki};$$

$$S_{ki} = x_{ki} \times p; p = [0; m-l];$$

$$x_{ki} = \frac{1}{m} \sum_{j=1}^m (x_{kj,i});$$

“Golden” series:

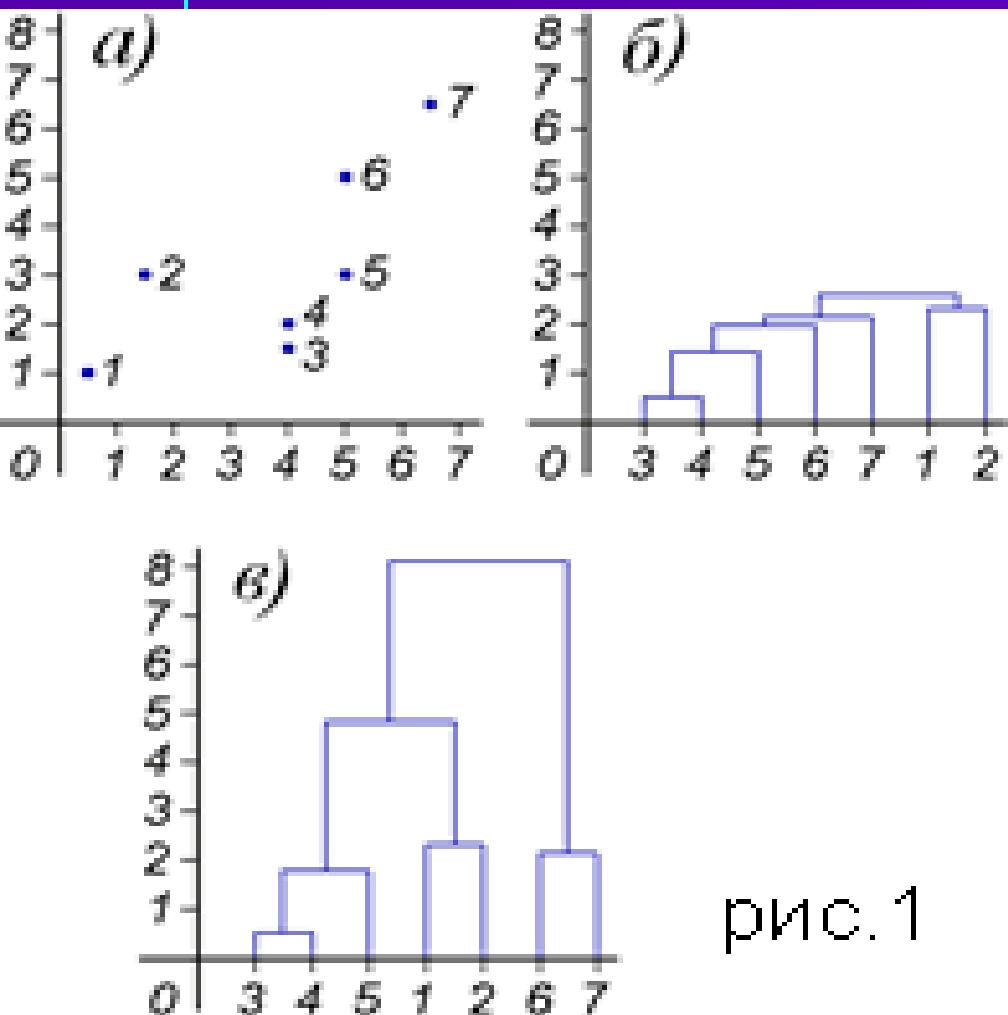
$$x^p = 1 - x; p = [1, 2, 3, \dots];$$

$$x_i = x_{i-p} + x_{i-1};$$

**На основе специальных алгоритмов  
кластер-анализа создана система  
распознавания многомерных  
функциональных характеристик  
сложных (биомедицинских) систем.**

**Начальные центры  
выбираются как в методе  
«k - средних»**

# Цепочечный эффект метода ближнего соседа



**Цепочечный эффект:**  
независимо от формы  
кластера к нему  
присоединяются  
ближайшие к границе  
объекты (б).

Метод дальнего соседа не  
приводит к подобному  
эффекту (в)

рис. 1

# Вычисление расстояний

- $d_{kl} = (\sum_j (x_{kj} - x_{lj})^2)^{1/2}$  ;
- $d_{kl} = \sum |(x_{kj} - x_{lj})|$  ;
- $d_{kl} = \sum |(x_{kj} - x_{lj}) / (x_{kj} + x_{lj})|$ ,

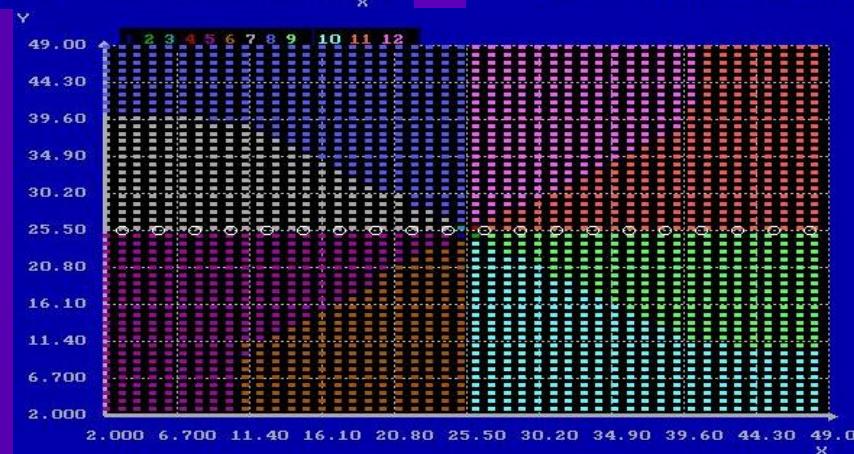
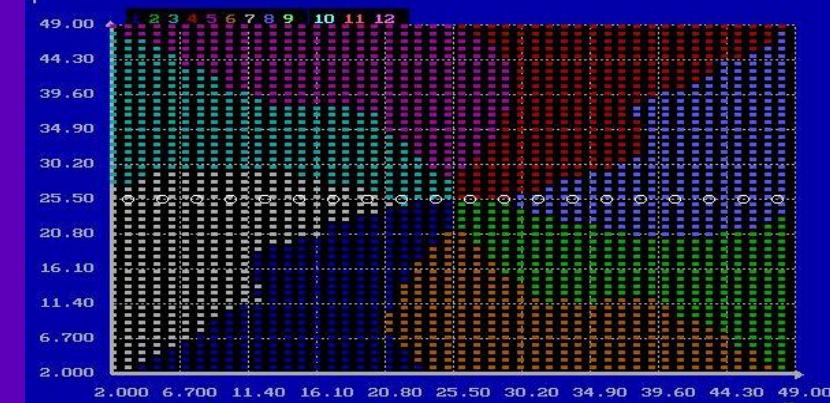
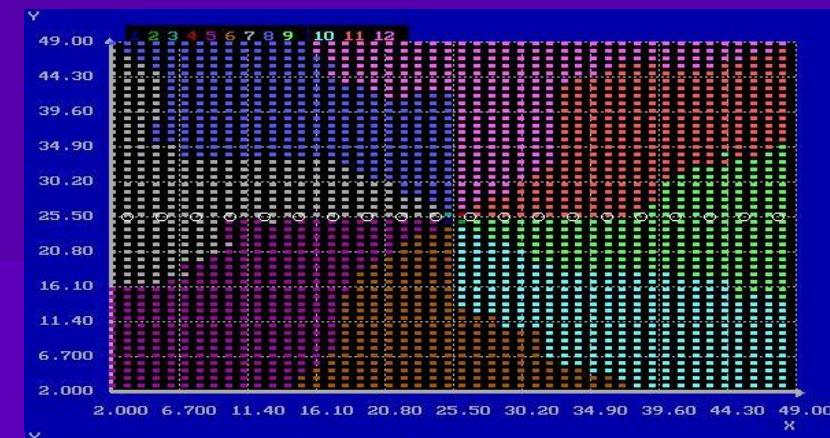
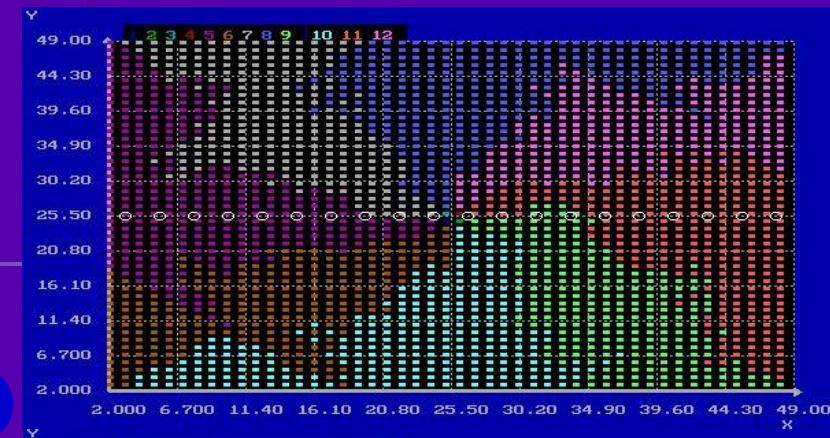
# Управление инерционностью кластера

- при  $n_k < I_{max} \Rightarrow x_{k+l,j} = (x_{kj}) + x_{lj},$
- при  $n_k = I_{max} \Rightarrow x_{k+l,j} = (x_{kj}) + x_{lj} - c \cdot m_j$
- И  $n_k = n_k - c, n_k = n_k - c, c = 0, 1, 2, \dots, I_{max};$

$I_{max}$  – максимальное кол. объектов в кластере;

$c$  - кол. “забываемых” условных объектов

# Влияние массы (инерционности) кластера на качество распознавания



5

# Последовательность действий

1. Выбрать начальные центры.
2. Провести кластеризацию в один проход в пространстве признаков сниженной размерности.
3. Выполнить аппроксимацию семейства восстановленных зависимостей.
4. Найти параметр (группу параметров), генерирующих данное семейство (по массиву данных, либо с помощью математического моделирования)

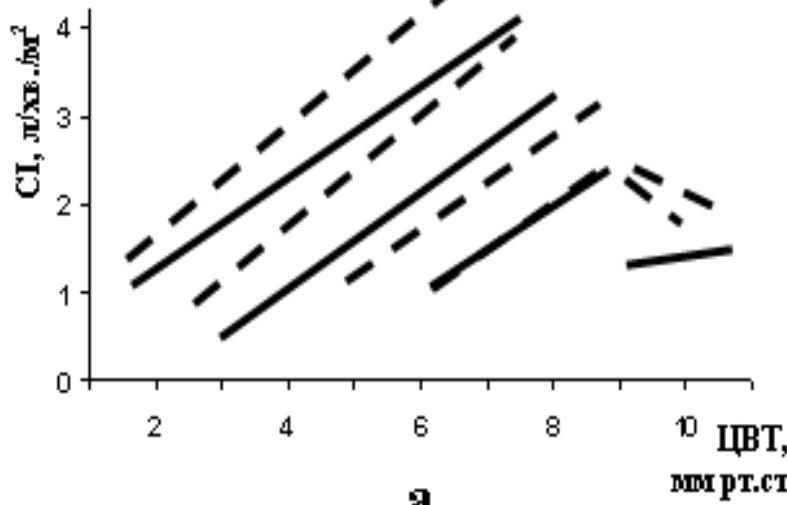
# **Качество распознавания:**

Максимизируется  
коэффициент  
детерминации

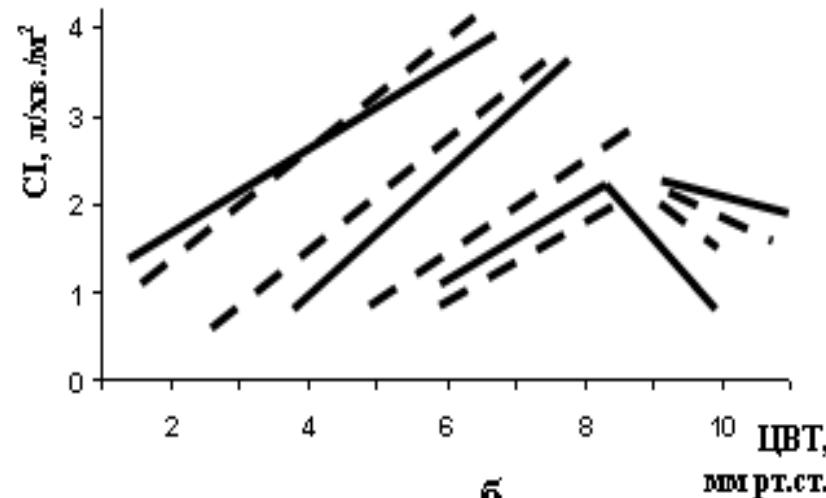
Алгоритм сразу стал давать хорошие результаты на реальных массивах наблюдений и работал гораздо хуже на семействах тестовых, искусственно зашумленных зависимостей.

Возник вопрос, почему?

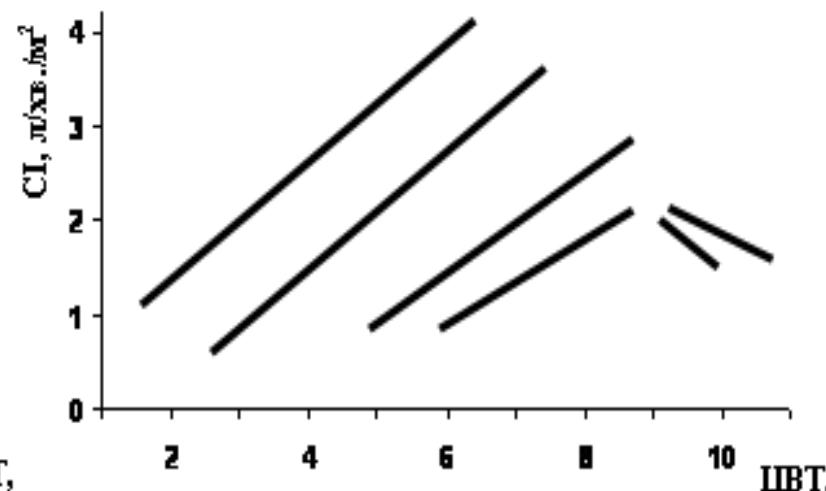
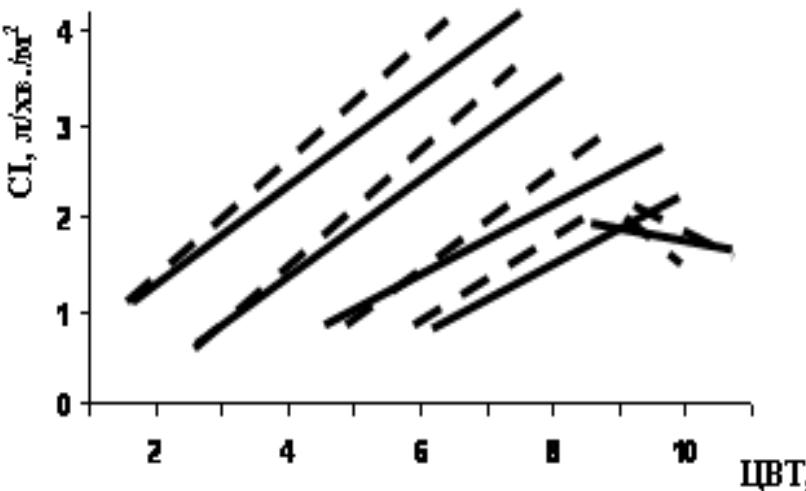
# Формирование семейства зависимостей с возрастанием объема выборки $(N/3; N/2; 2N/2; N=602$ набл.)



а

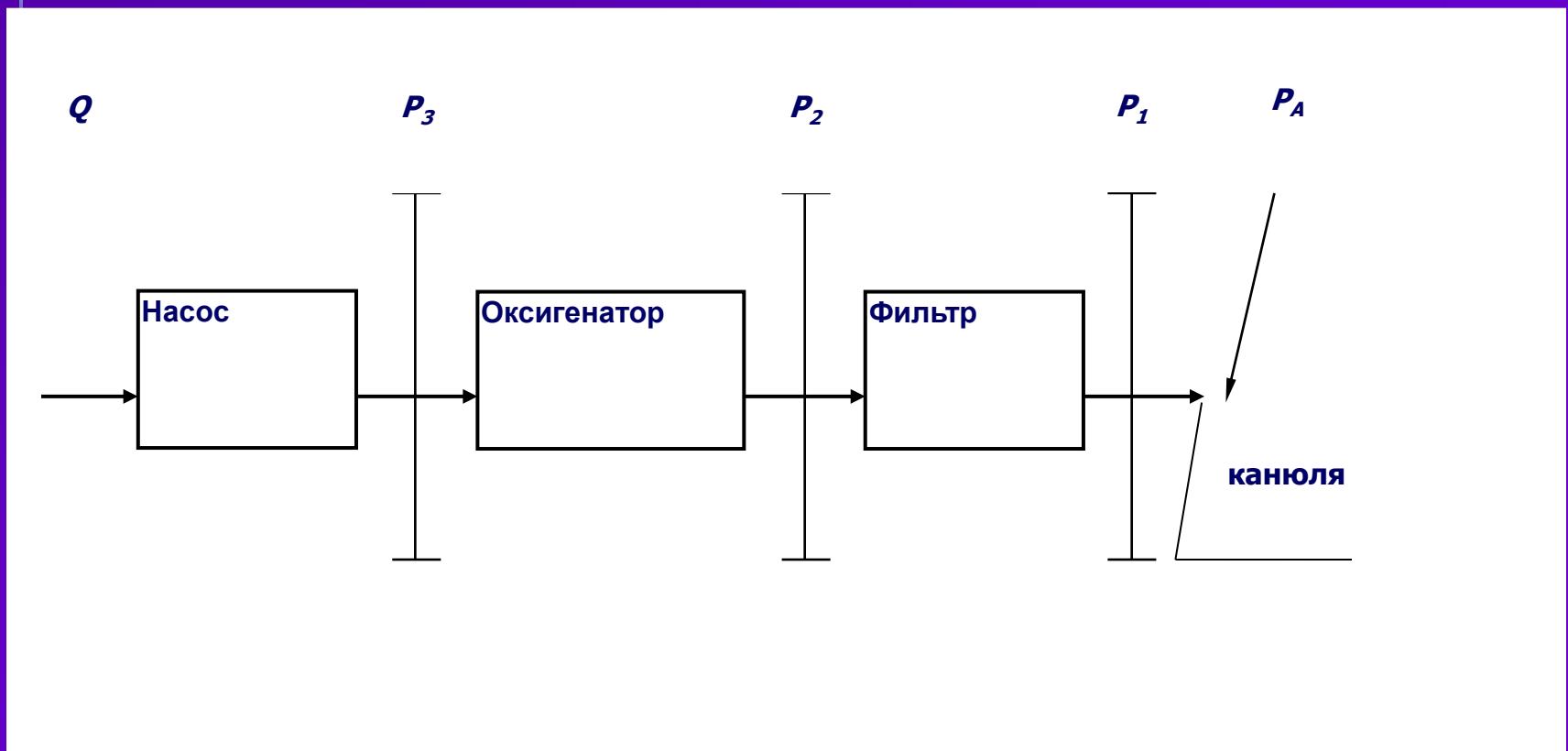


б

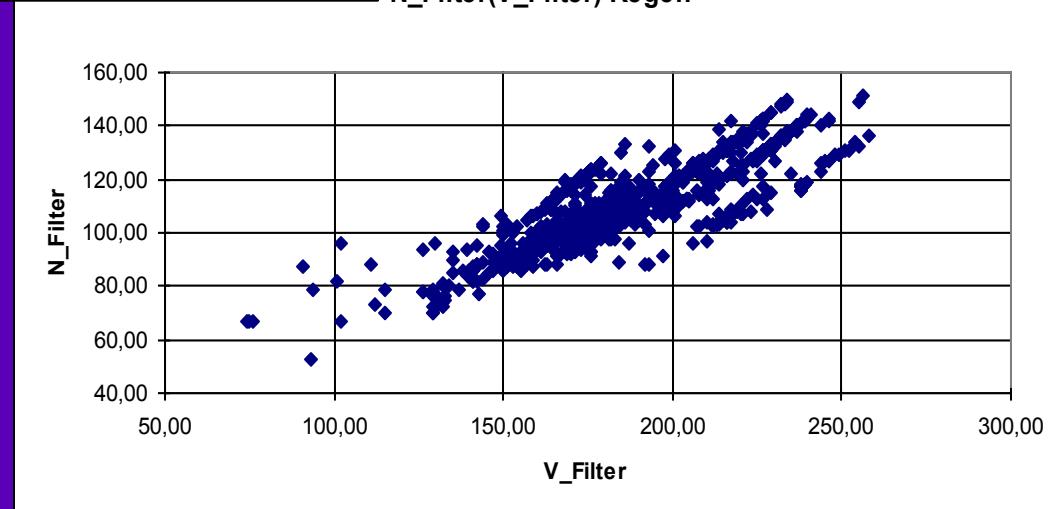
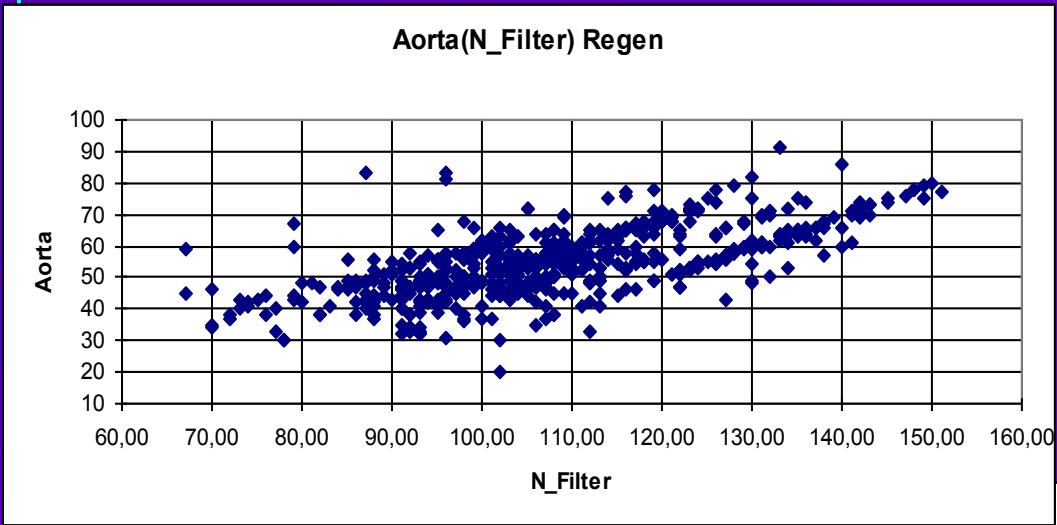


г

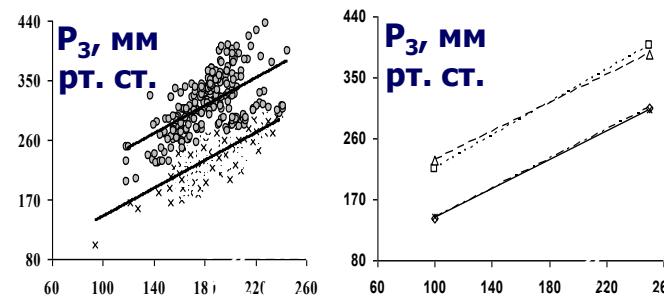
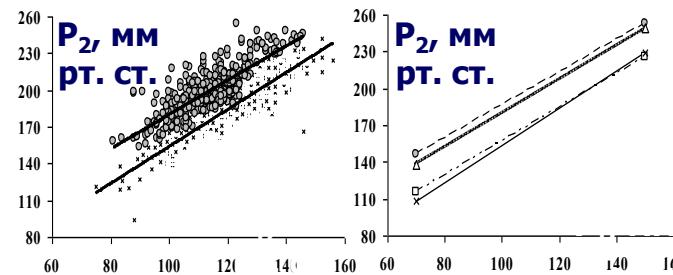
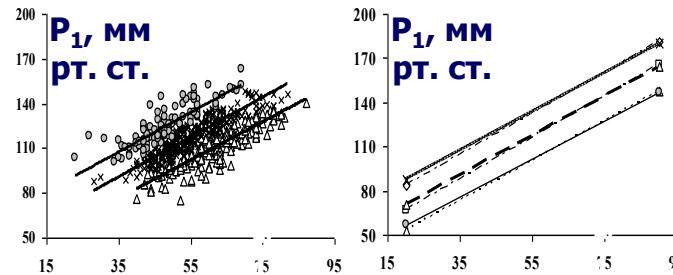
# Схема установки с элементами различного сопротивления



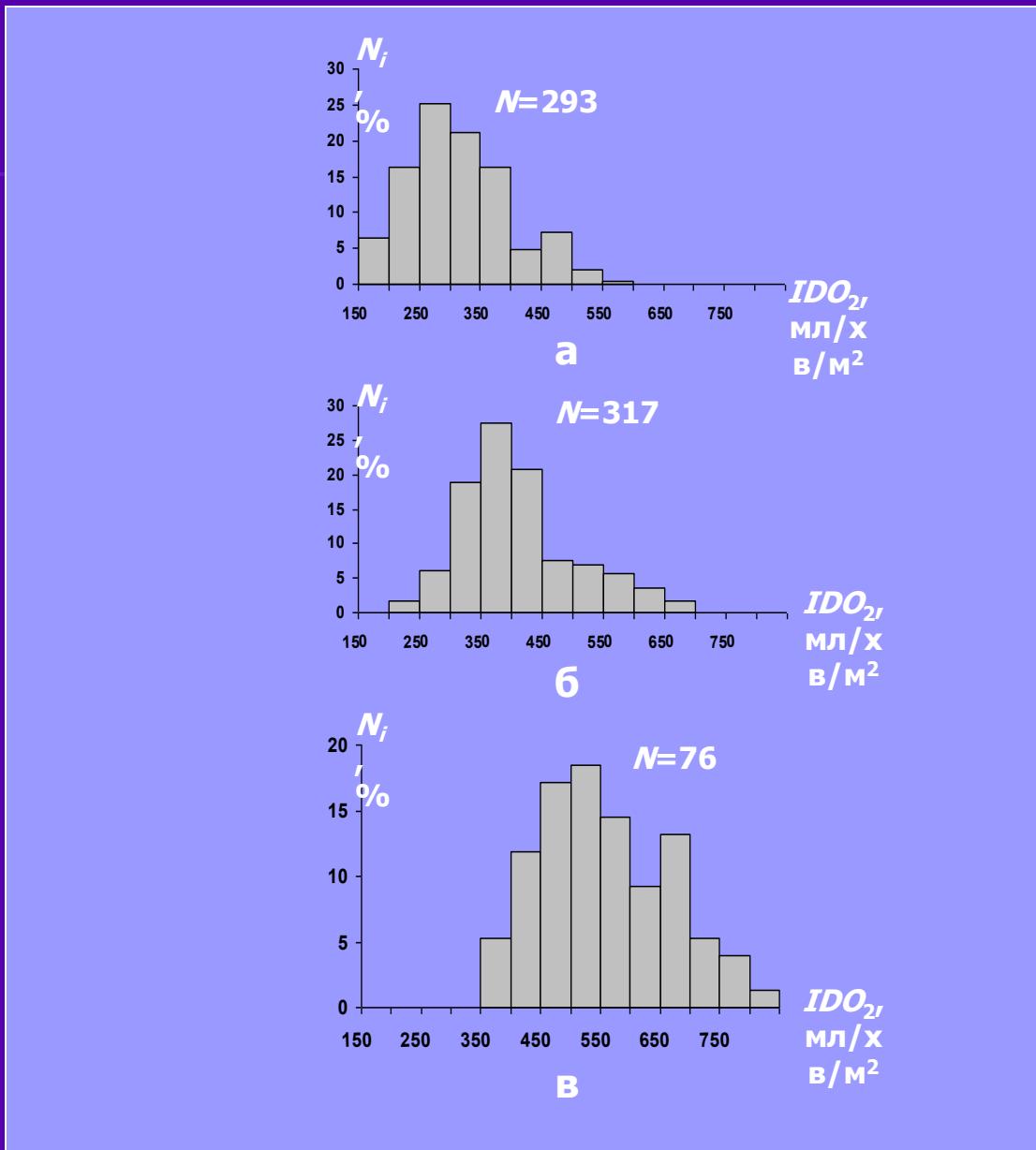
# Семейства зависимостей давлений при различных установленных элементах

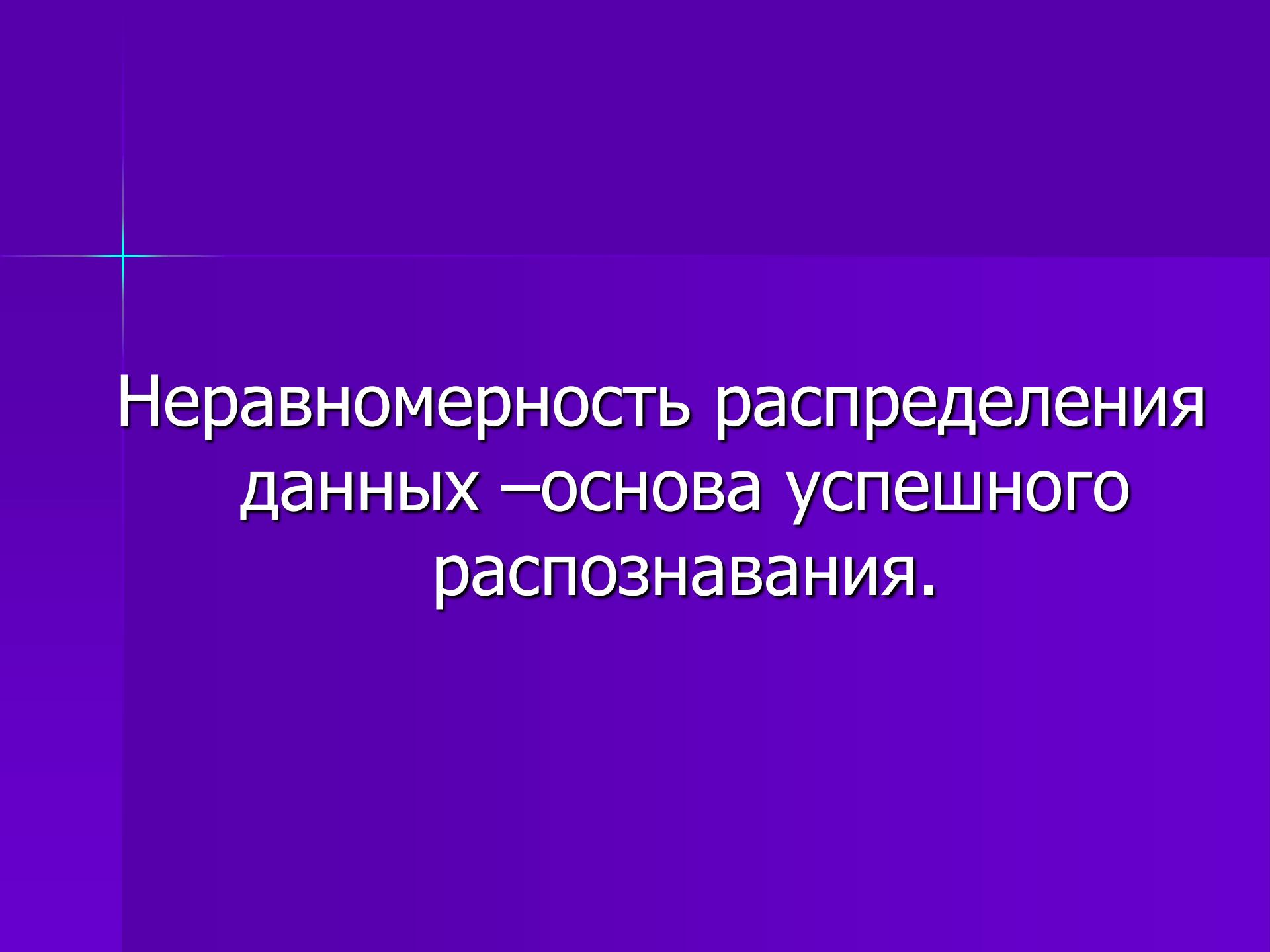


# Сопоставление известных зависимостей с результатами кластеризации



# Полигоны частот показателей после кластеризации





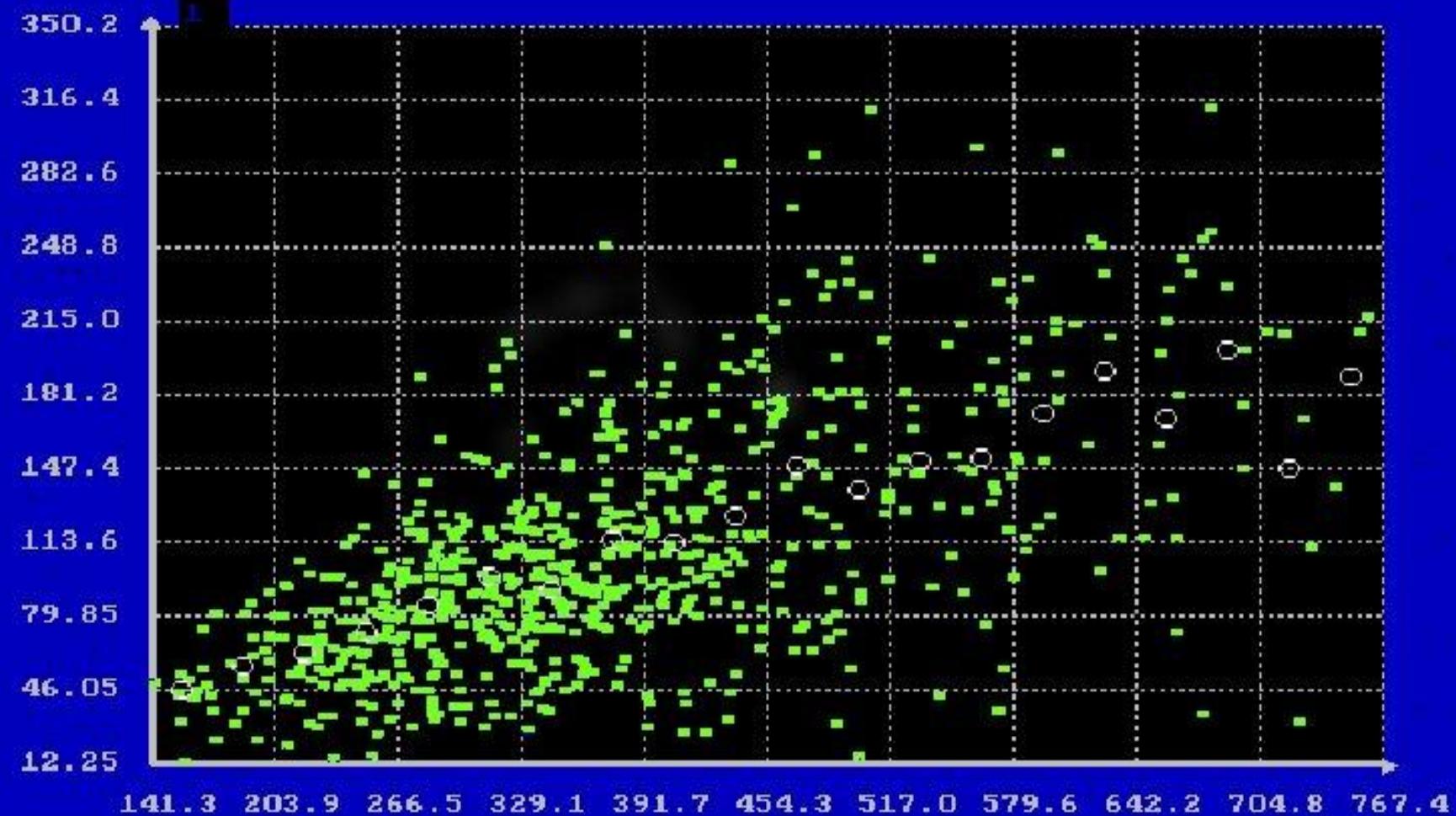
**Неравномерность распределения  
данных – основа успешного  
распознавания.**



# Структура пространства состояний системной доставки и потребления кислорода в интактной системе кровообращения человека

# Исходная скатерограмма $\text{IDO}_2$ - $\text{IVO}_2$

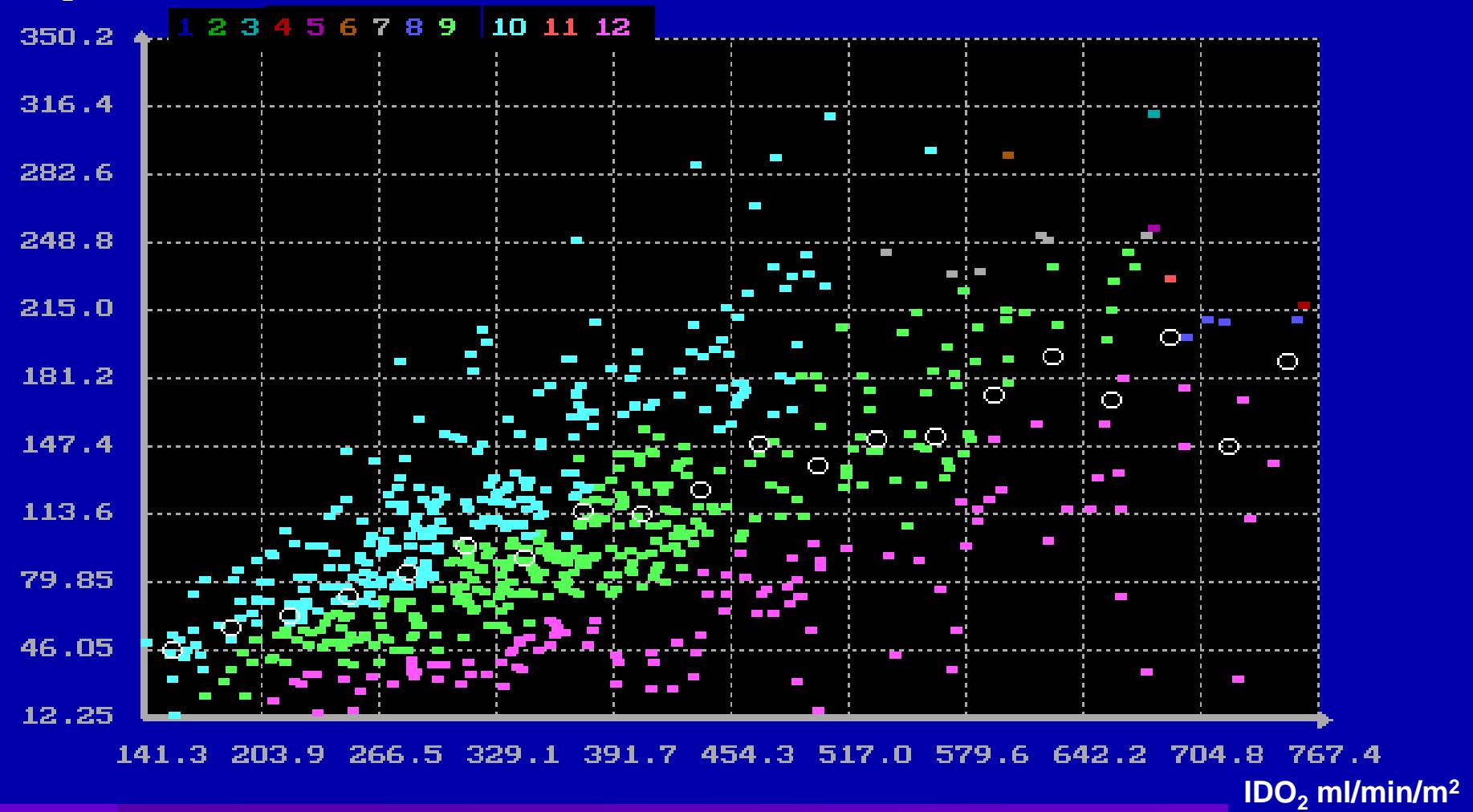
$\text{IVO}_2$ , ml/min/m<sup>2</sup>



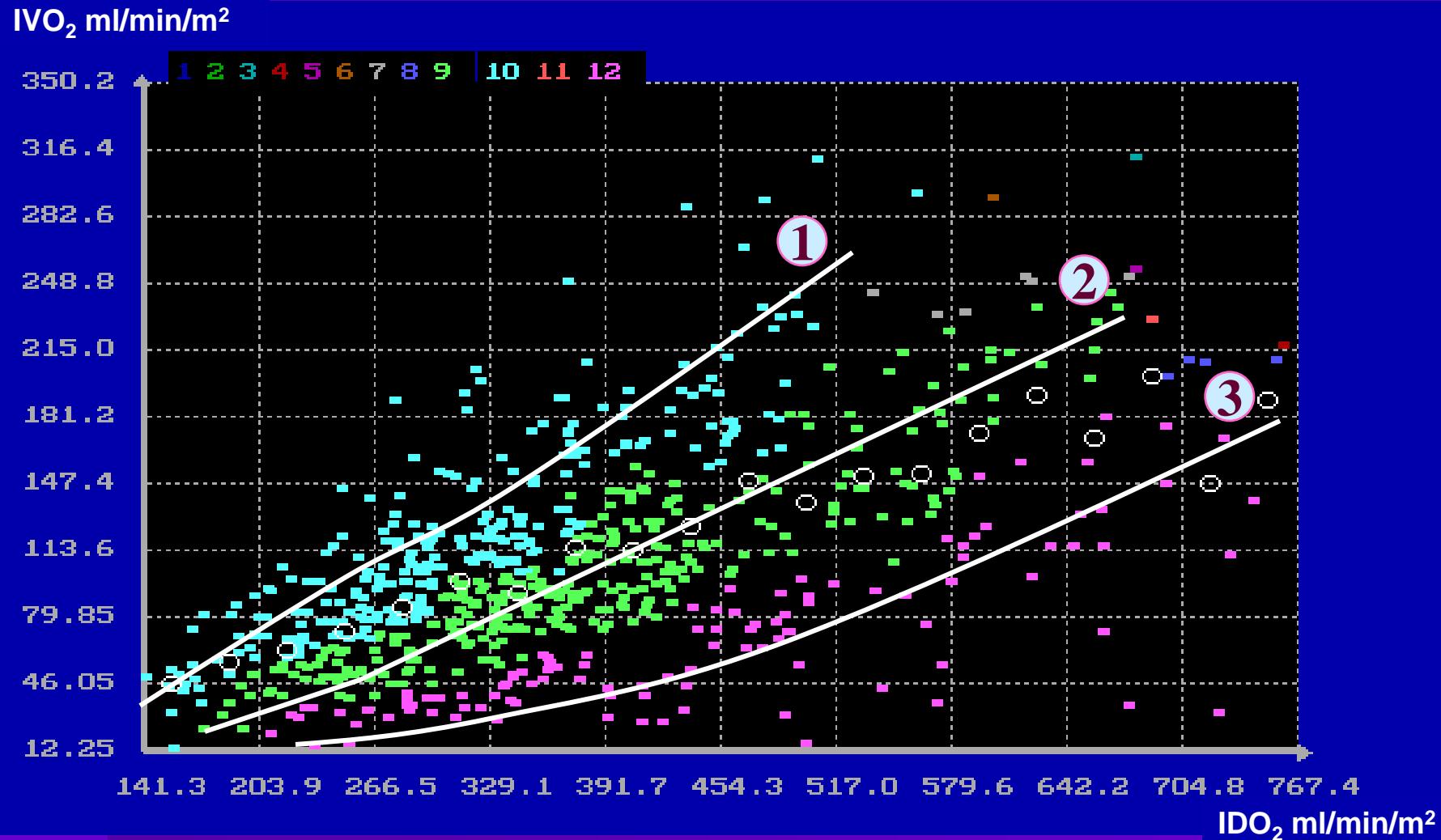
$\text{IDO}_2$ , ml/min/m<sup>2</sup>

# Структура пространства состояний доставки и потребления $O_2$ . Интактная ССС. Шаг 1.

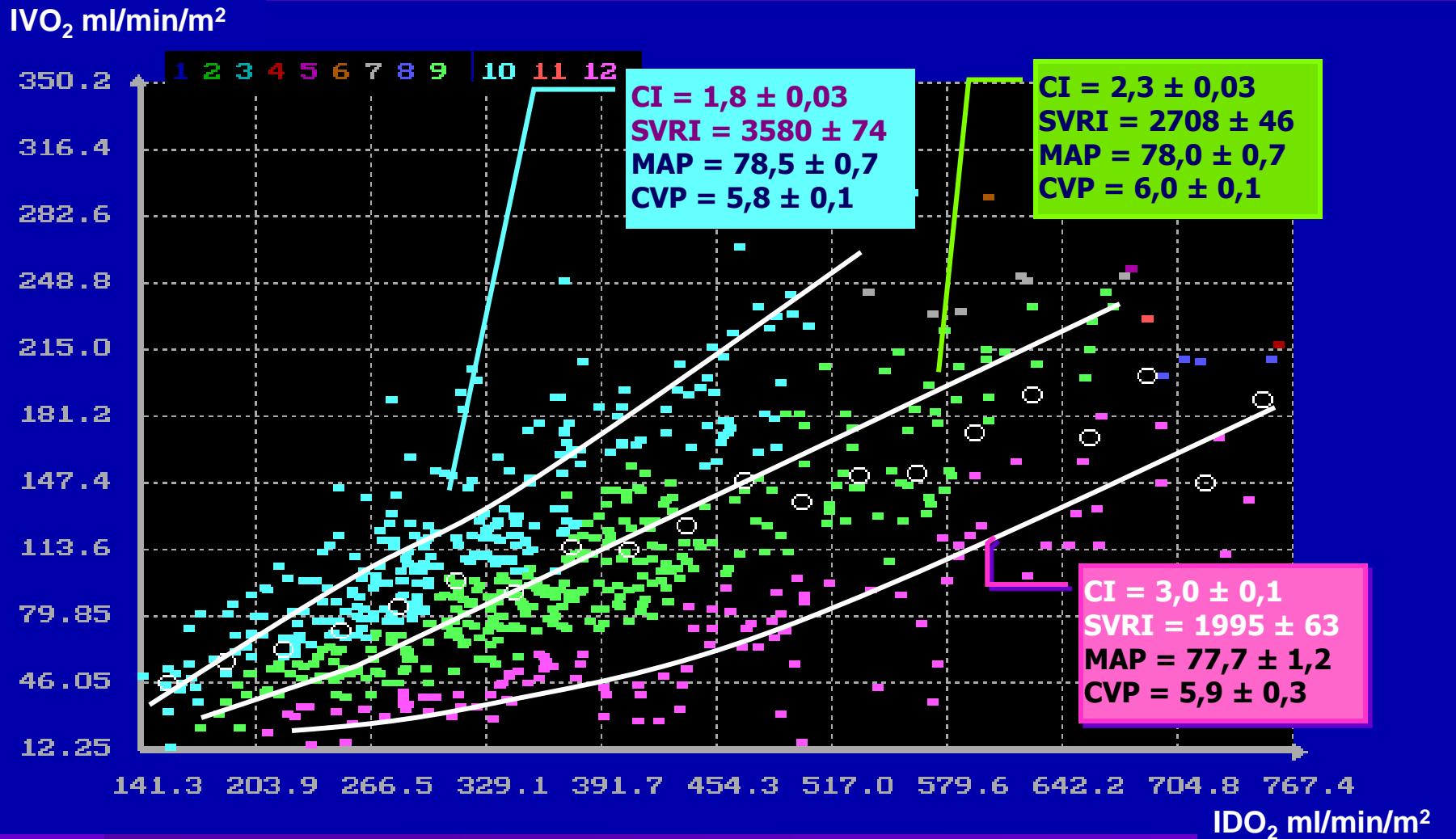
$IVO_2$  ml/min/m<sup>2</sup>



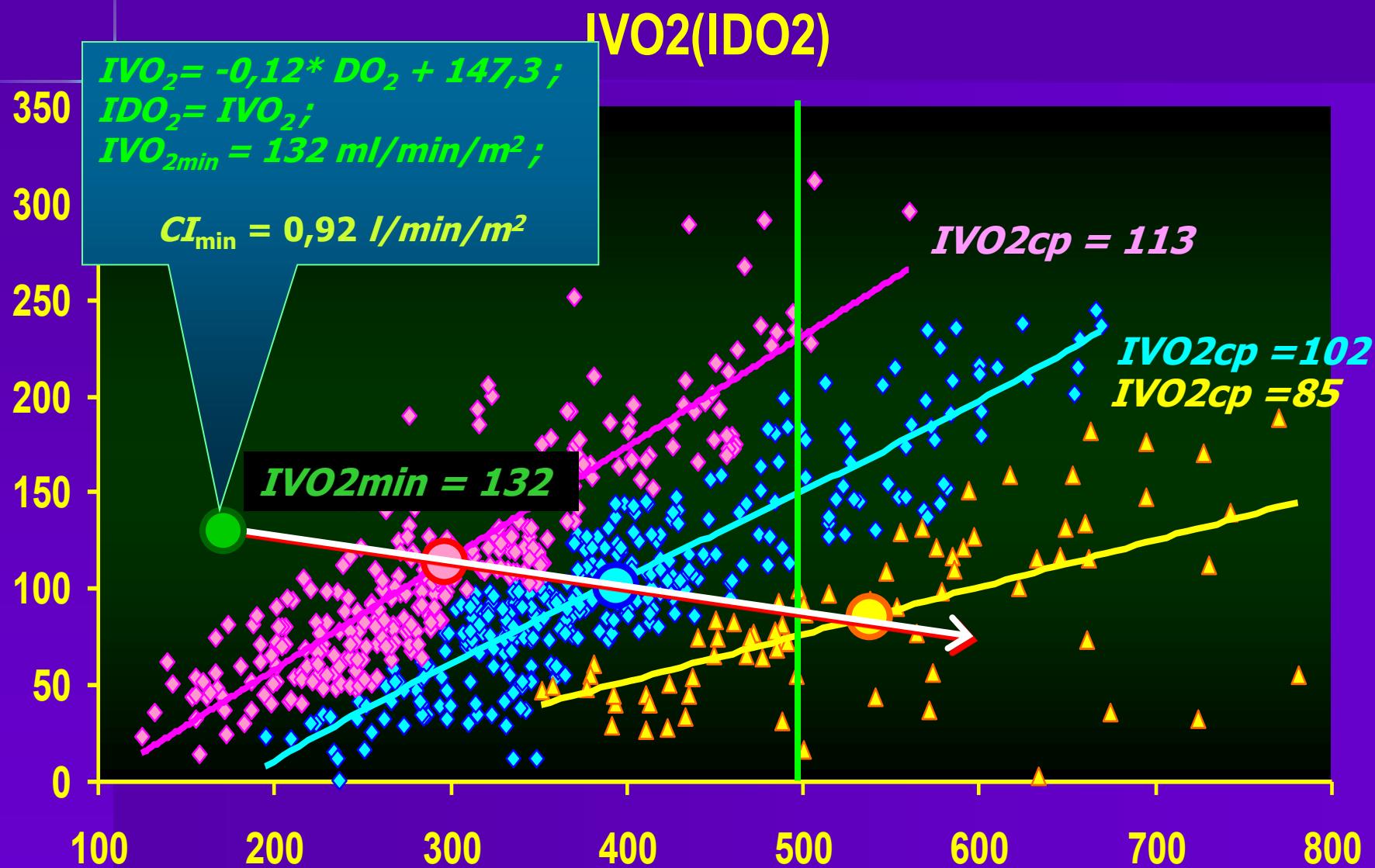
# Структура пространства состояний доставки и потребления $O_2$ . Интактная ССС. Шаг 2.



# *Структура пространства состояний доставки и потребления $O_2$ . Интактная ССС. Шаг 3.*



# Структура пространства состояний доставки и потребления $O_2$ . Интактная CCC.

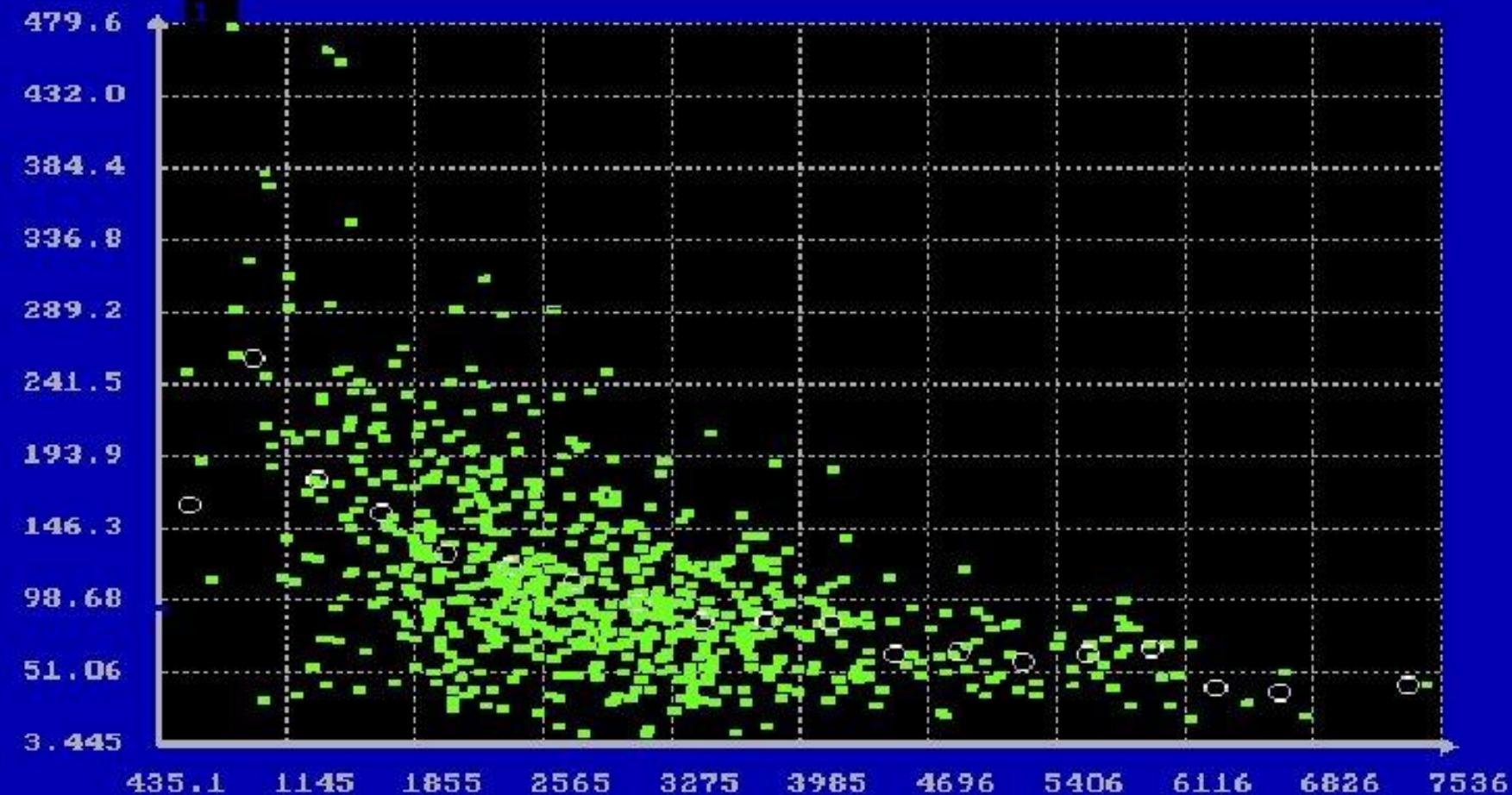




# *Роль периферического сосудистого тонуса в оптимизации системного потребления $O_2$*

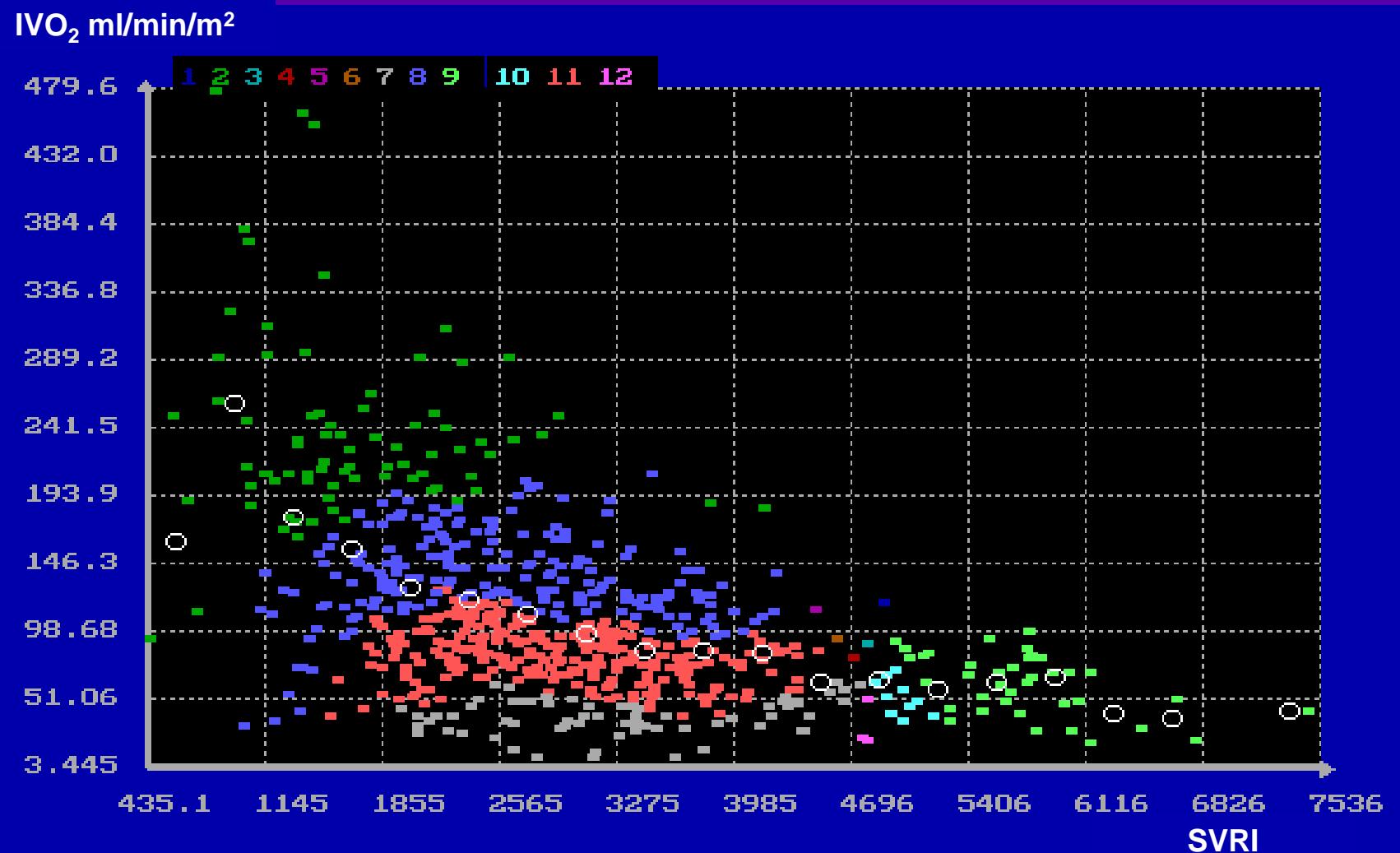
# Исходная $IVO_2$ - $SVRI$ скатерограмма. Интактная ССС.

$IVO_2$ , ml/min/m<sup>2</sup>

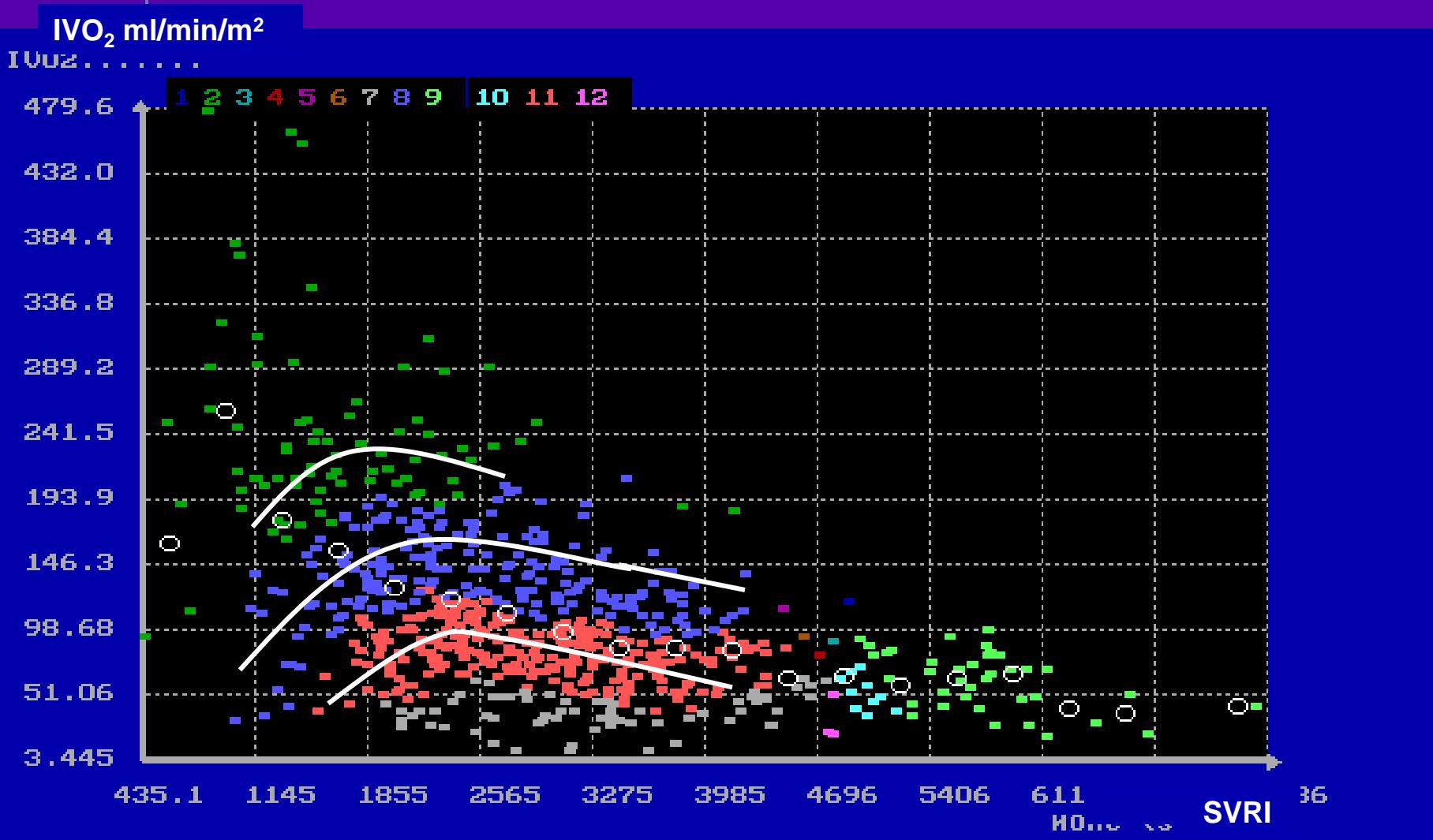


$SVRI$ , dyn\*s\*cm<sup>-5</sup>\*m<sup>2</sup>

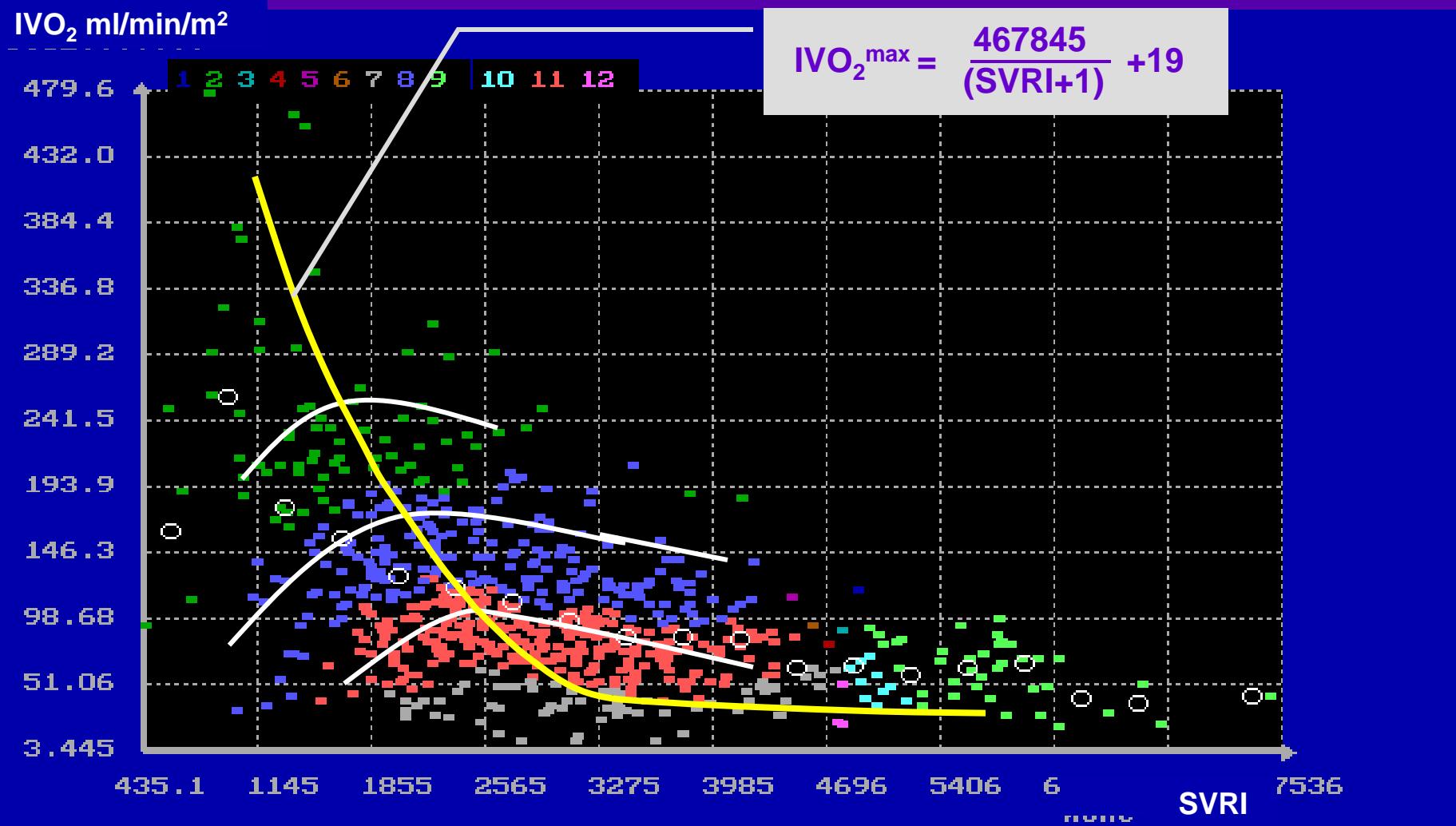
# Структура пространства состояний «системное потребление $O_2$ - периферическое сосудистое сопротивление» $IVO_2(SVRI)$ .



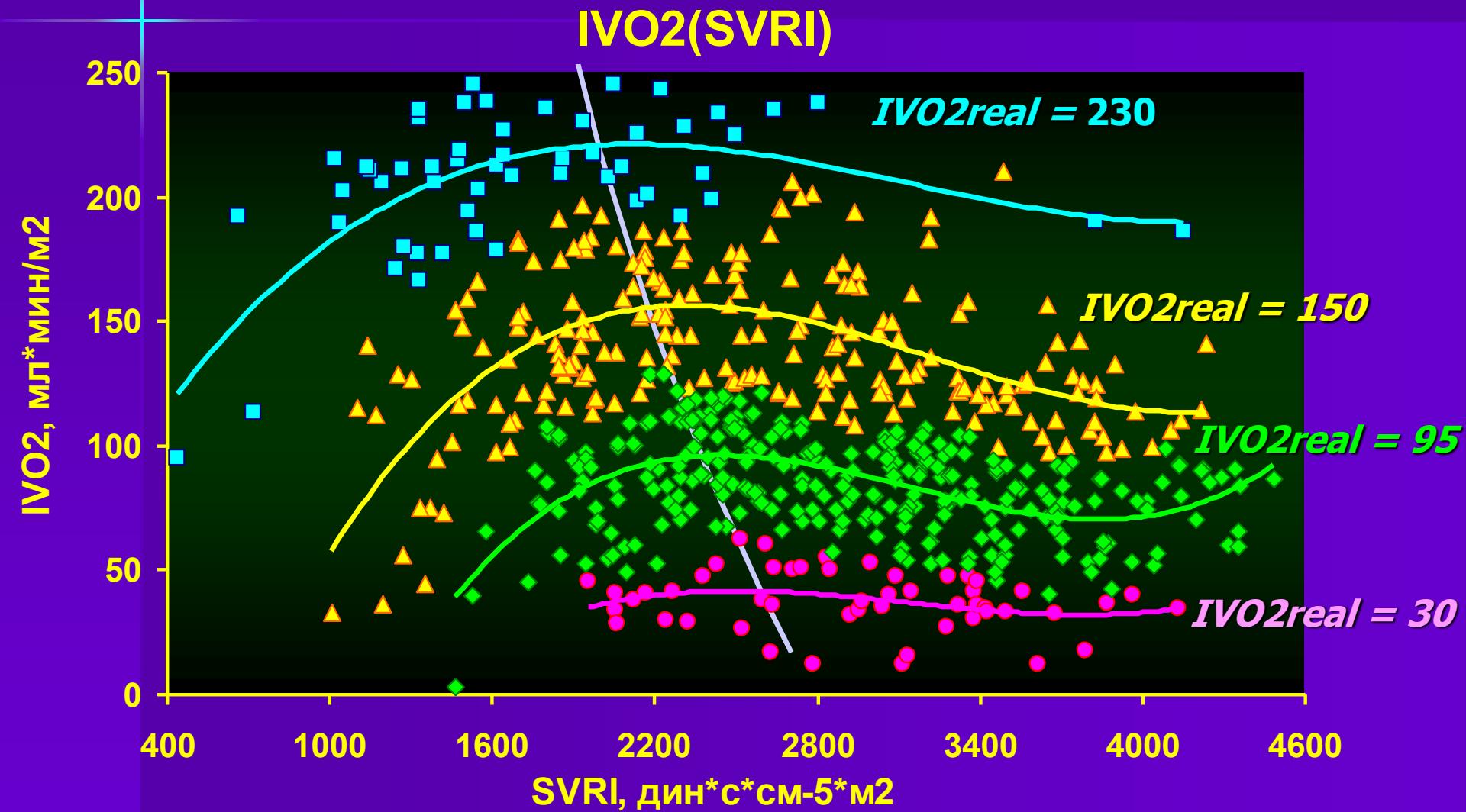
**Структура пространства состояний  
«системное потребление  $O_2$  - периферическое  
сосудистое сопротивление»  $IVO_2(SVRI)$ .**



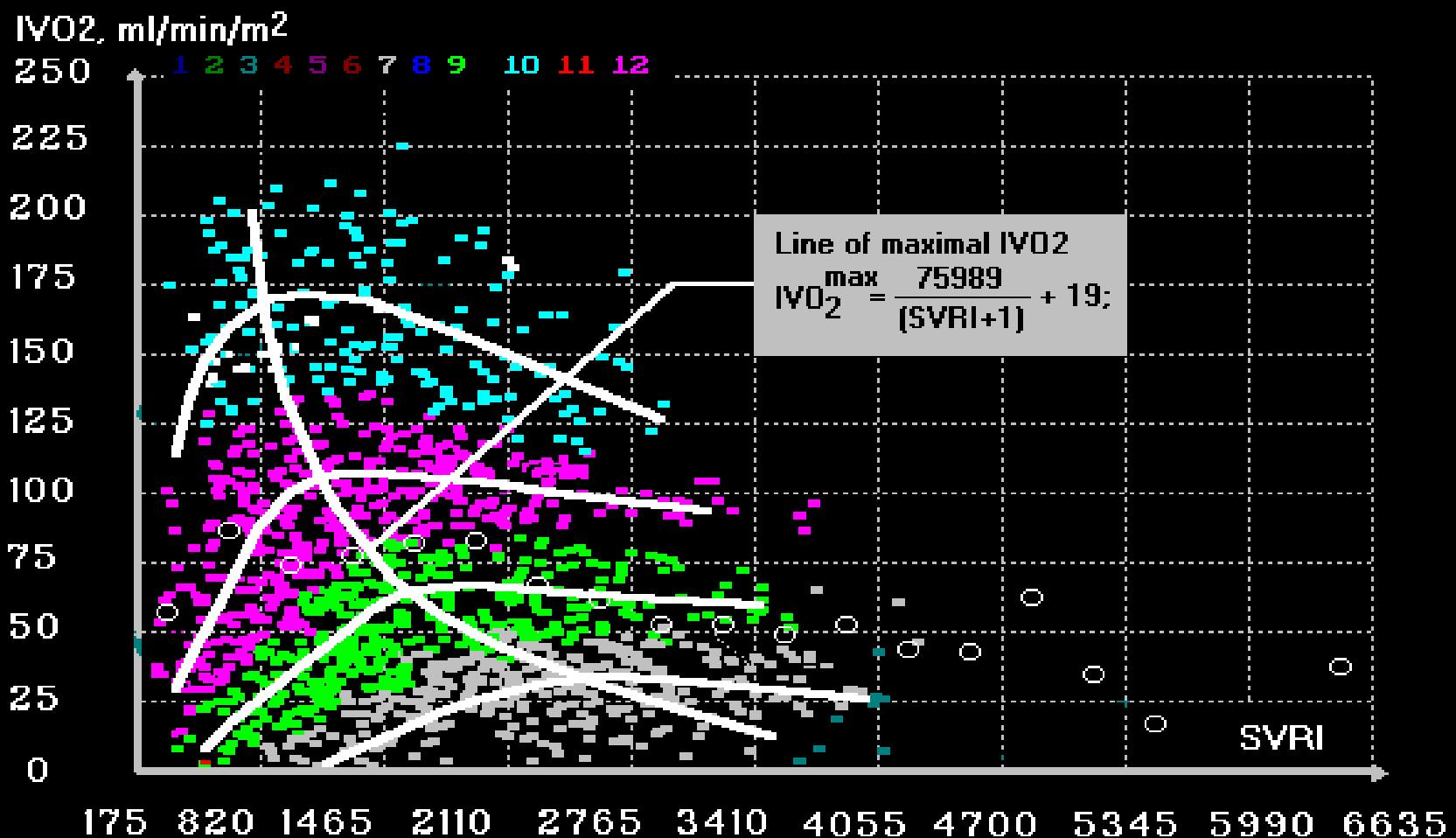
# *Структура пространства состояний «системное потребление $O_2$ - периферическое сосудистое сопротивление» $IVO_2(SVRI)$ .*



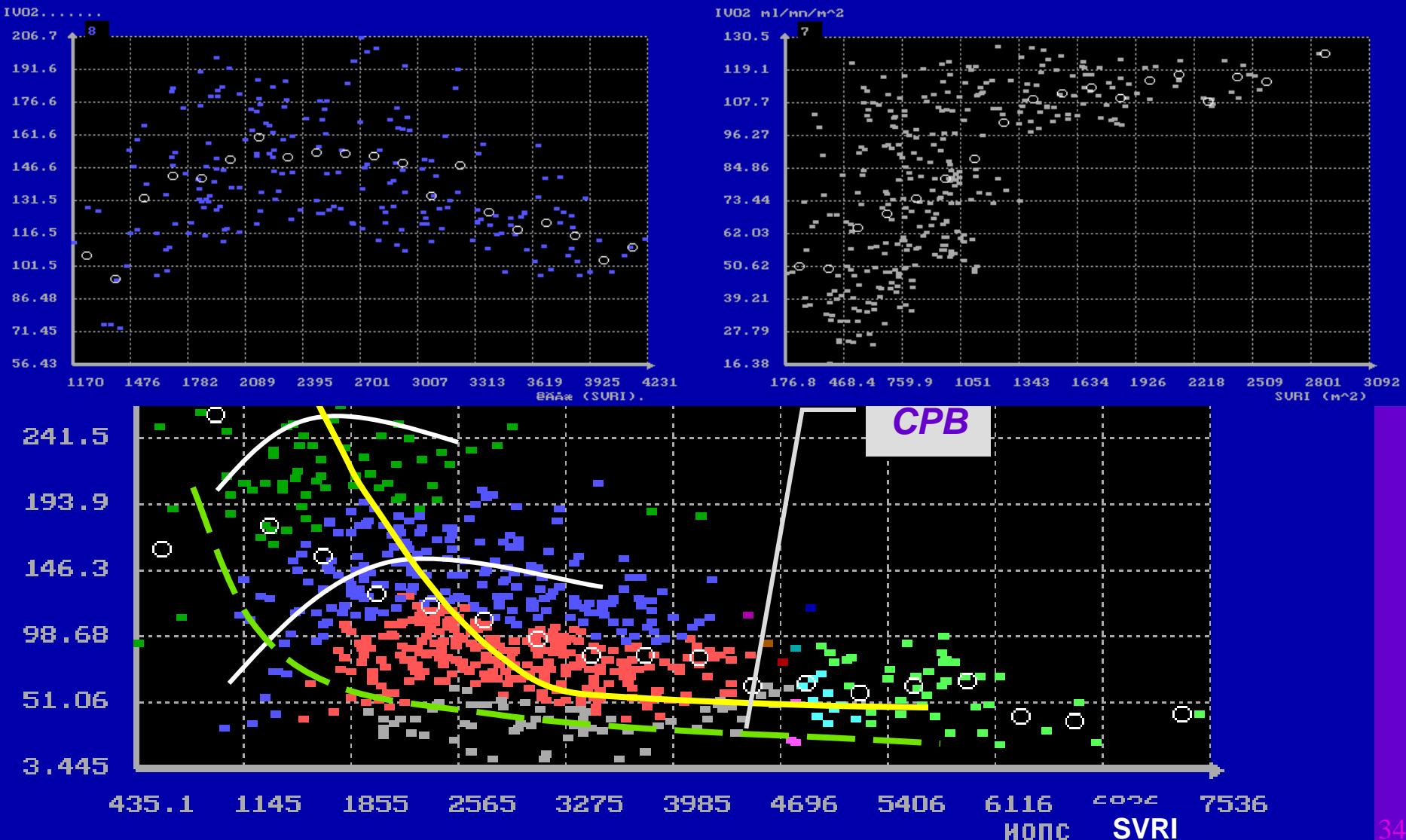
# Соотношение шунтового и нутритивного кровотока при различном уровне SVRI (сердечного выброса)



# **Потребление $O_2$ ( $IVO_2$ ) - сосудистое сопротивление ( $SVRI$ ) при ИК**



# Типичные зависимости потребления $O_2$ от сосудистого сопротивления в интактной ССС и при ИК

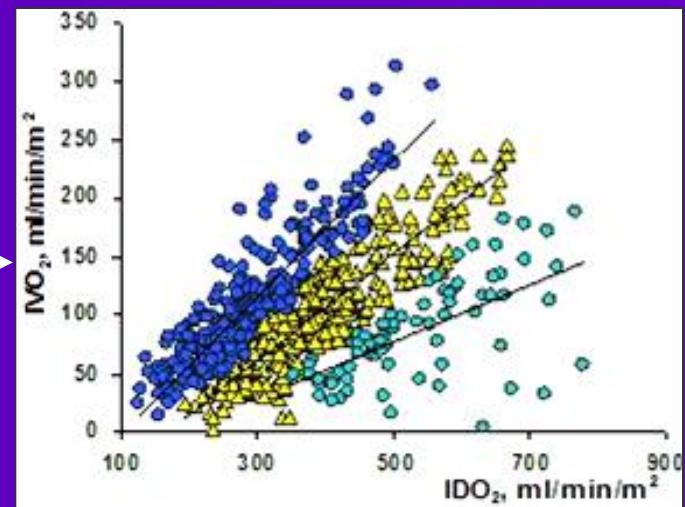
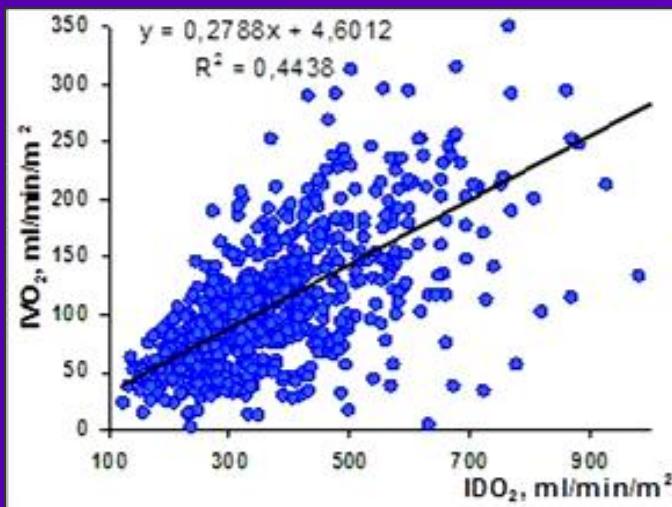


# **Результаты распознавания зависимостей**

# Доставка и потребление кислорода

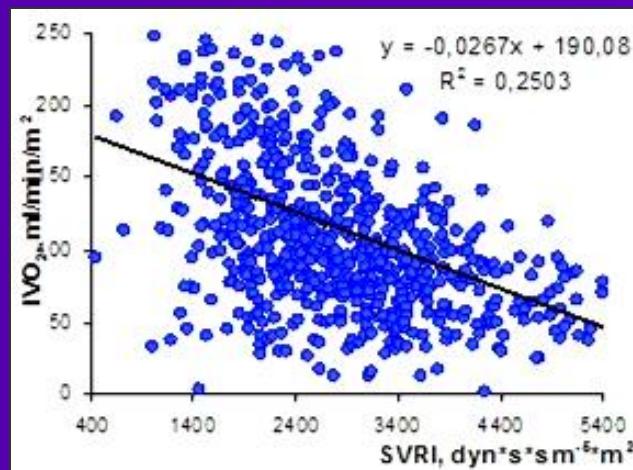
*Исходное*

*Структурированное*

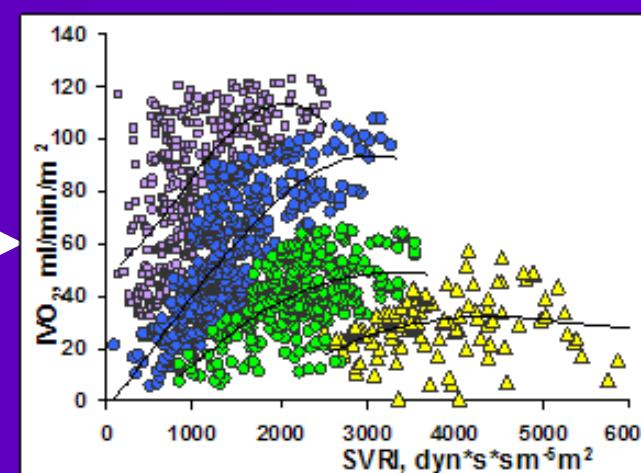
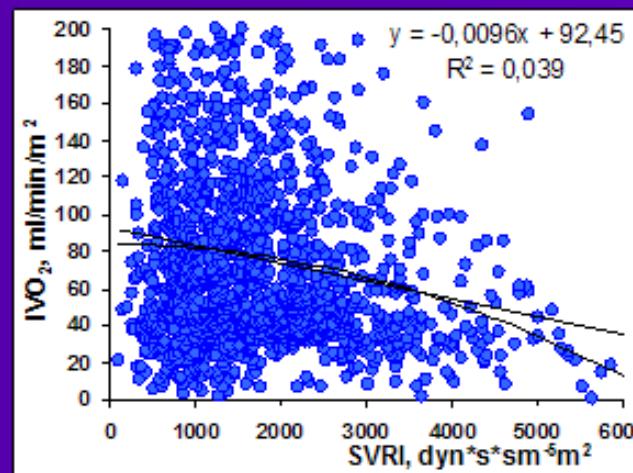
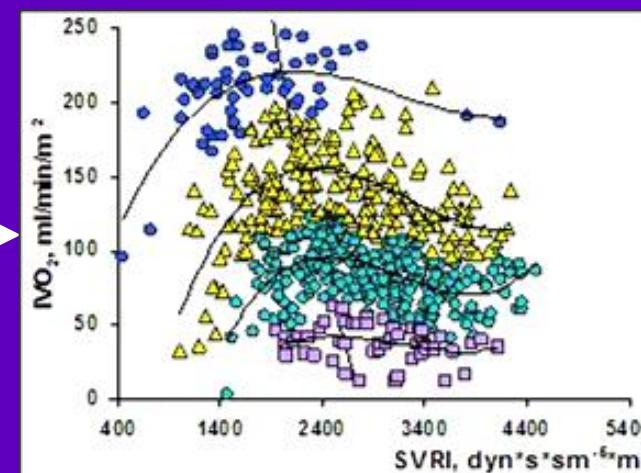


# Потребление кислорода – периферическое сопротивление

*Исходные*

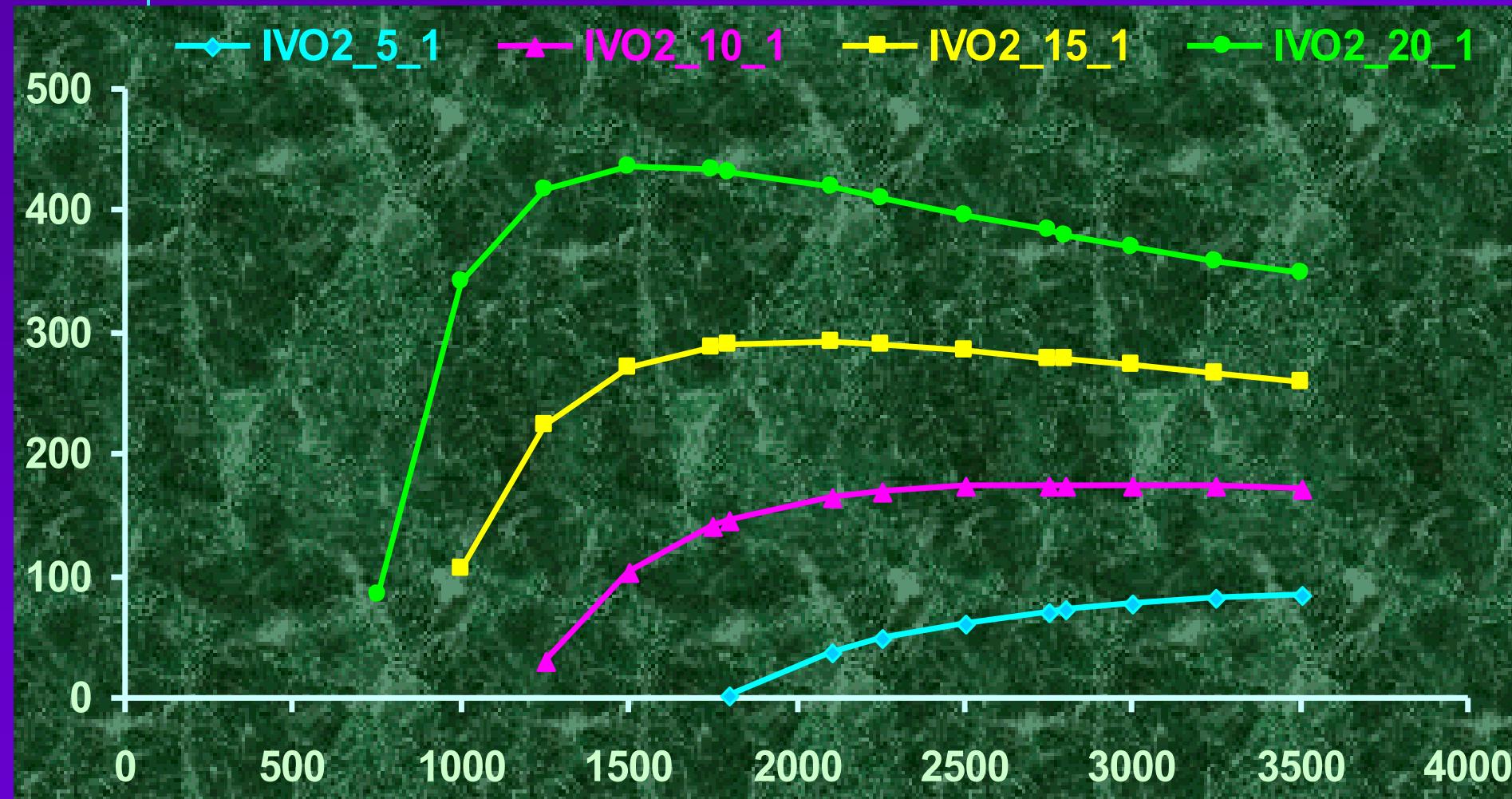


*Структурированные*



# Статистическое моделирование зависимостей потребления $O_2$ от периферического сопротивления

(Регрессионная модель, нелинейная относительно параметров)



Можно ли построить регрессионную модель, не выбирая ее из заданного класса функций?

Модели:

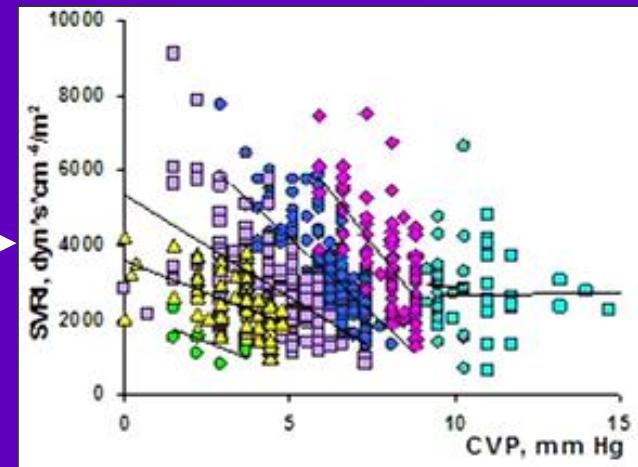
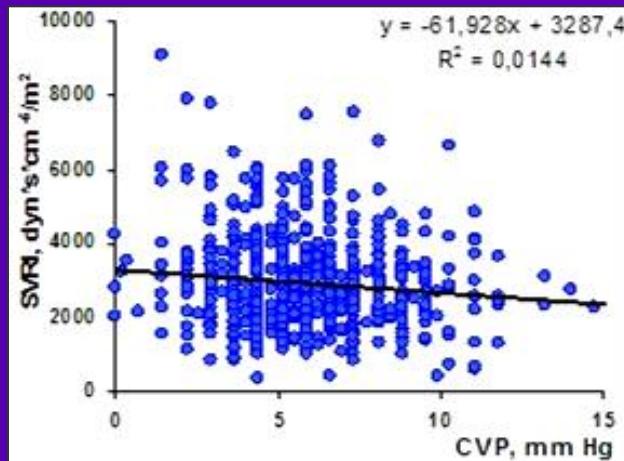
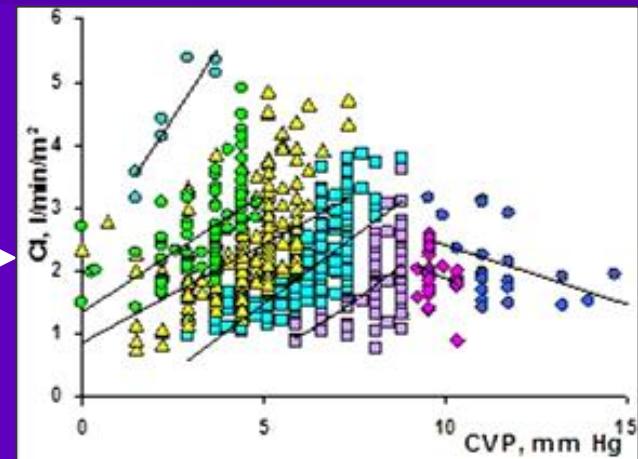
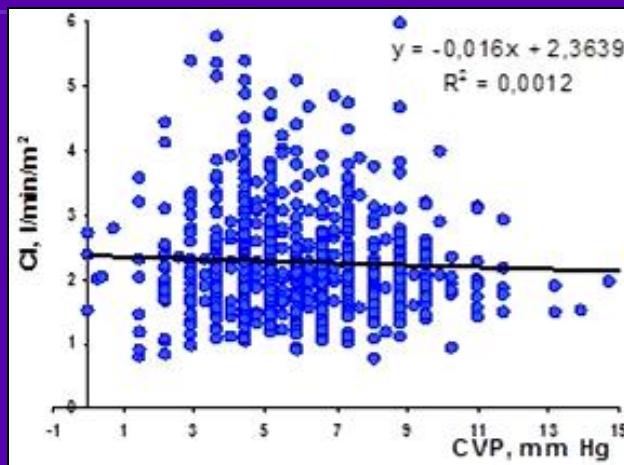
- Аддитивные
- Мультипликативные
- Другие...

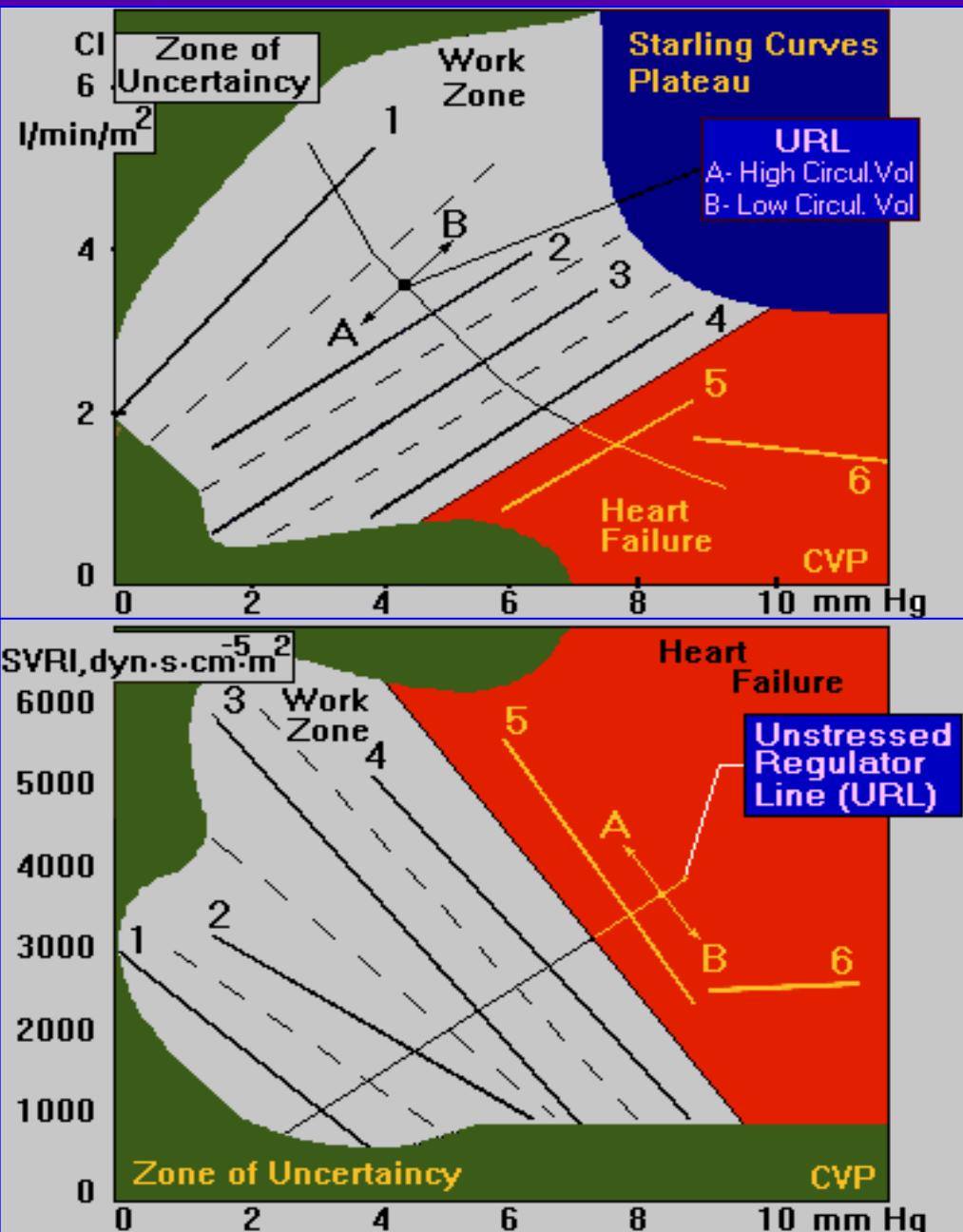
Можно ли построить «класс-независимую» модель, используя «пи-теорему»?

# Сердечный выброс и периферическое сопротивление - преднагрузка

Исходные

Структурированные



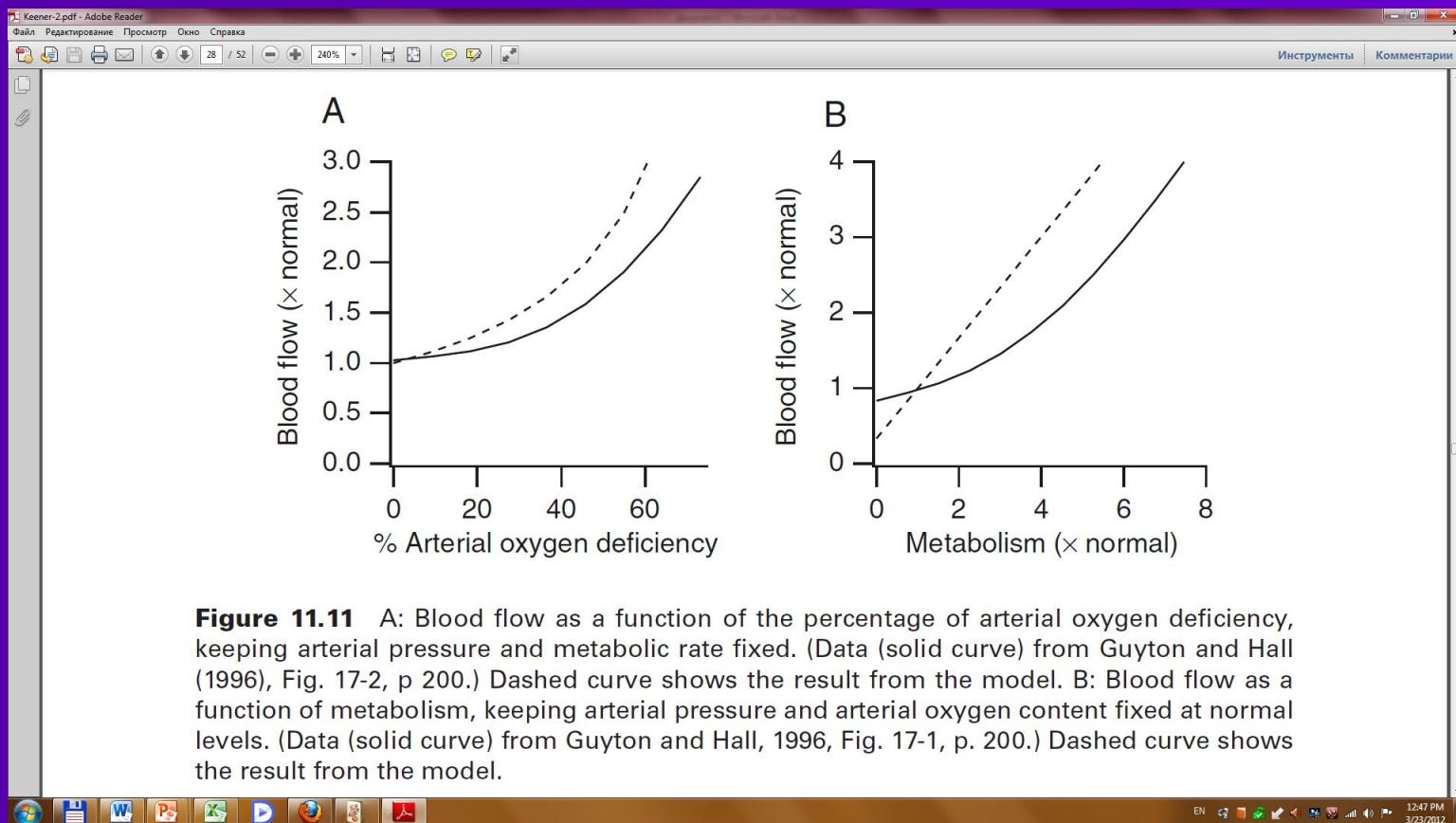


- **ФХ1 – гиперфункция сердца**
- **ФХ2 – нормальная насосная функция сердца (НФС)**
- **ФХ3 – умеренно сниженная НФС**
- **ФХ4 – значительно сниженная НФС**
- **ФХ5 – критически сниженная НФС**

# **Автоматизированное управление кровообращением**

Используются номограммы на основе распознанных по клиническим данным функциональных характеристик кровообращения и результатов математического моделирования.

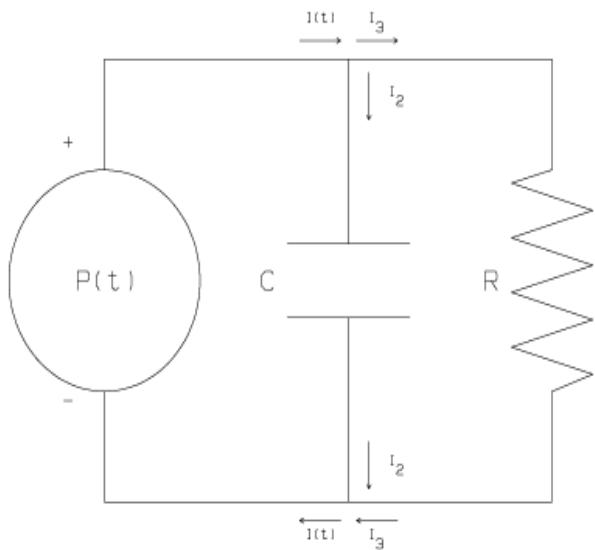
# Связь доставки и потребления $O_2$ с сердечным выбросом



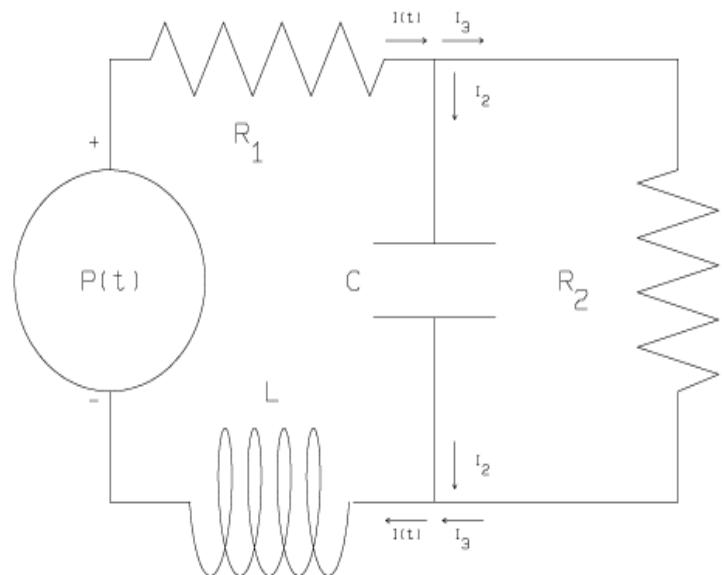
**Figure 11.11** A: Blood flow as a function of the percentage of arterial oxygen deficiency, keeping arterial pressure and metabolic rate fixed. (Data (solid curve) from Guyton and Hall (1996), Fig. 17-2, p 200.) Dashed curve shows the result from the model. B: Blood flow as a function of metabolism, keeping arterial pressure and arterial oxygen content fixed at normal levels. (Data (solid curve) from Guyton and Hall, 1996, Fig. 17-1, p. 200.) Dashed curve shows the result from the model.

# Модели «воздушного мешка»

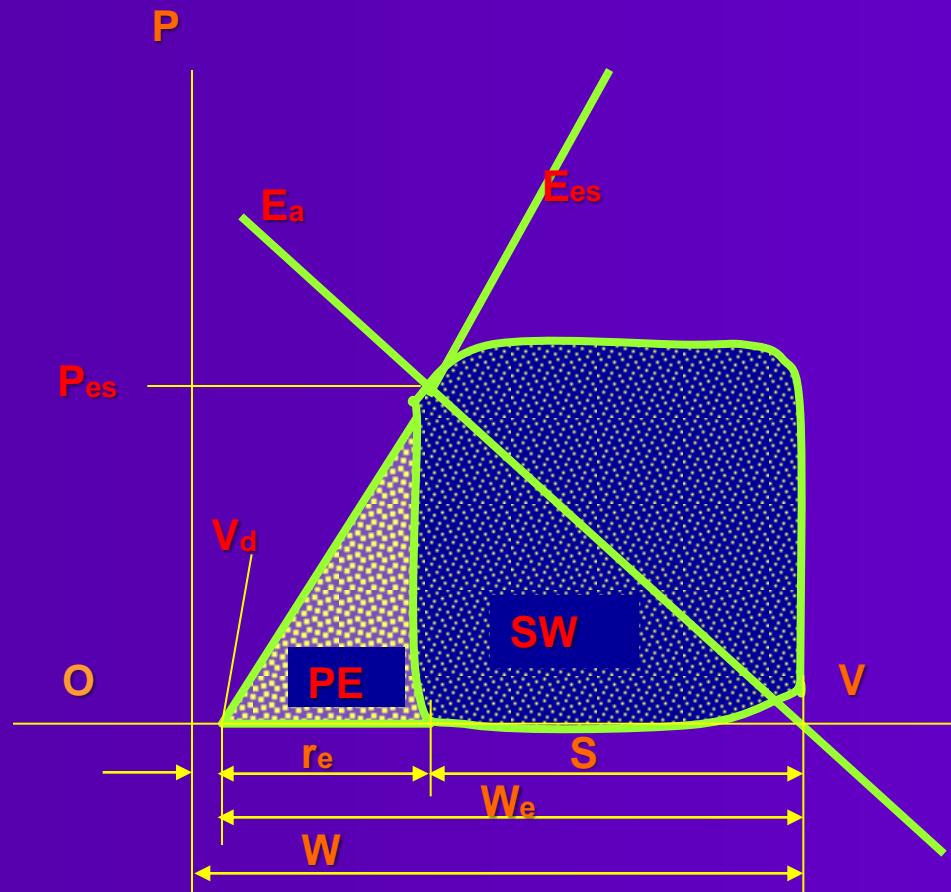
2-element Windkessel model

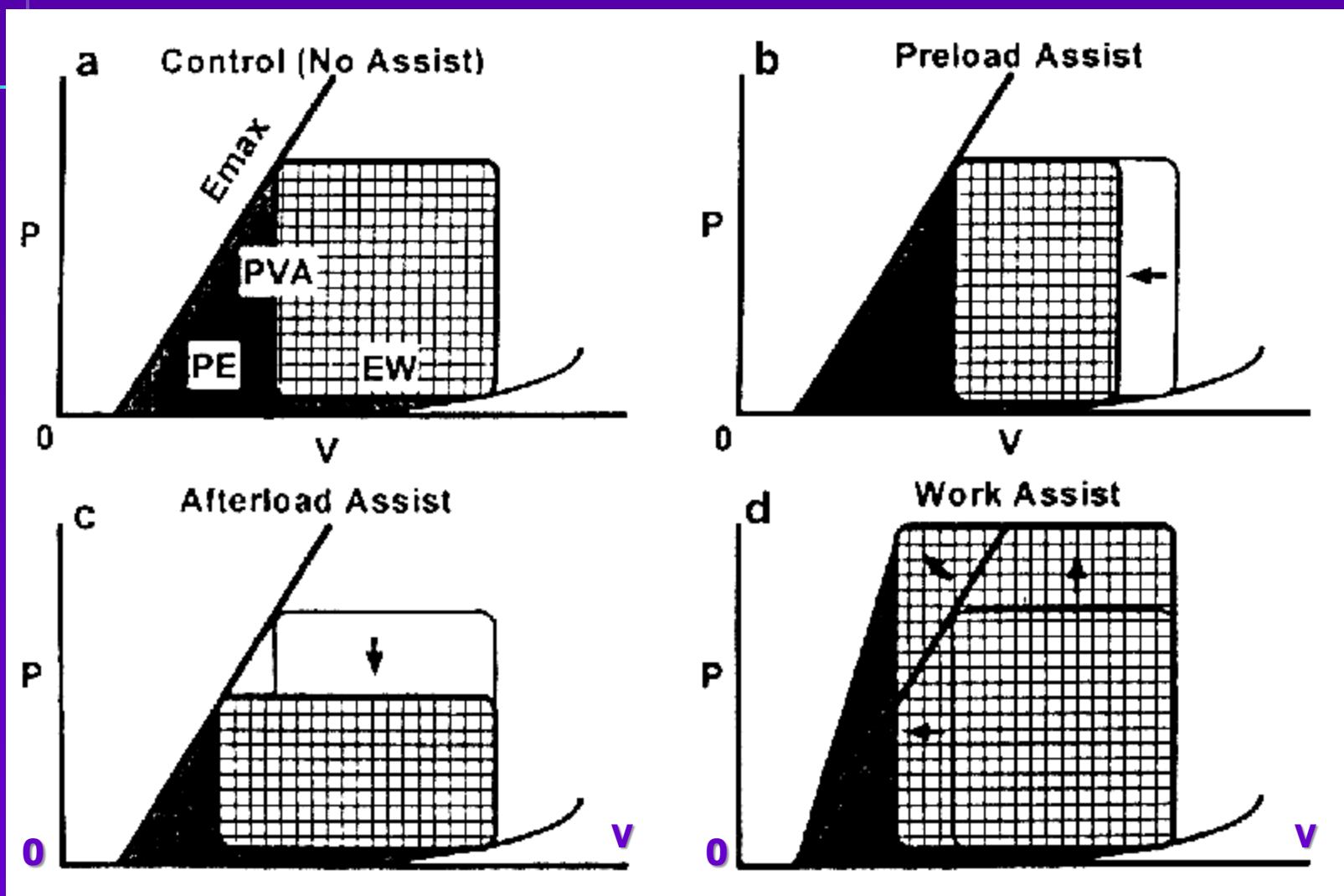


4-element Windkessel model

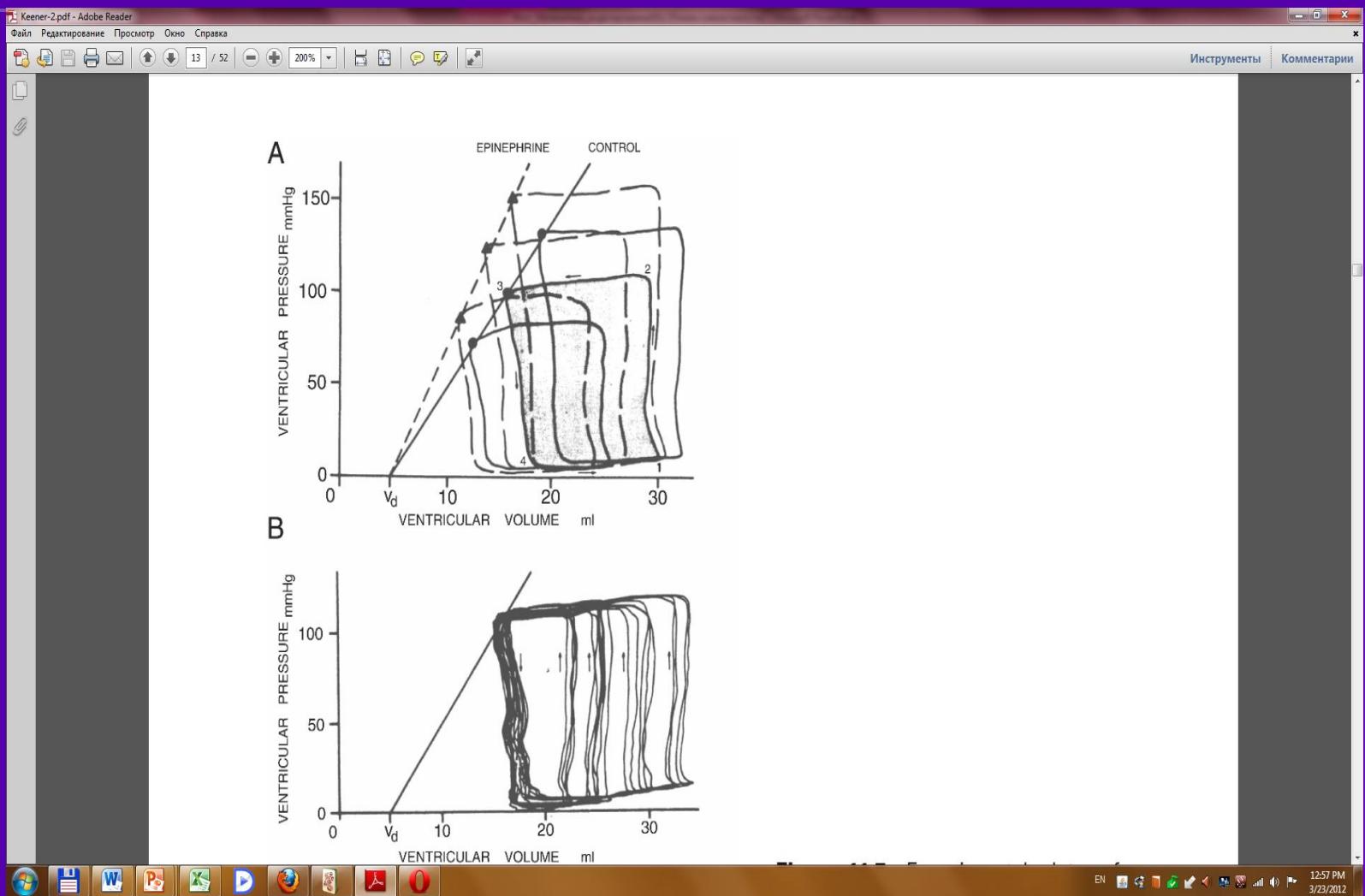


# Scheme of contractile activity of LV on the base of diagram pressure (p) - volume (v)

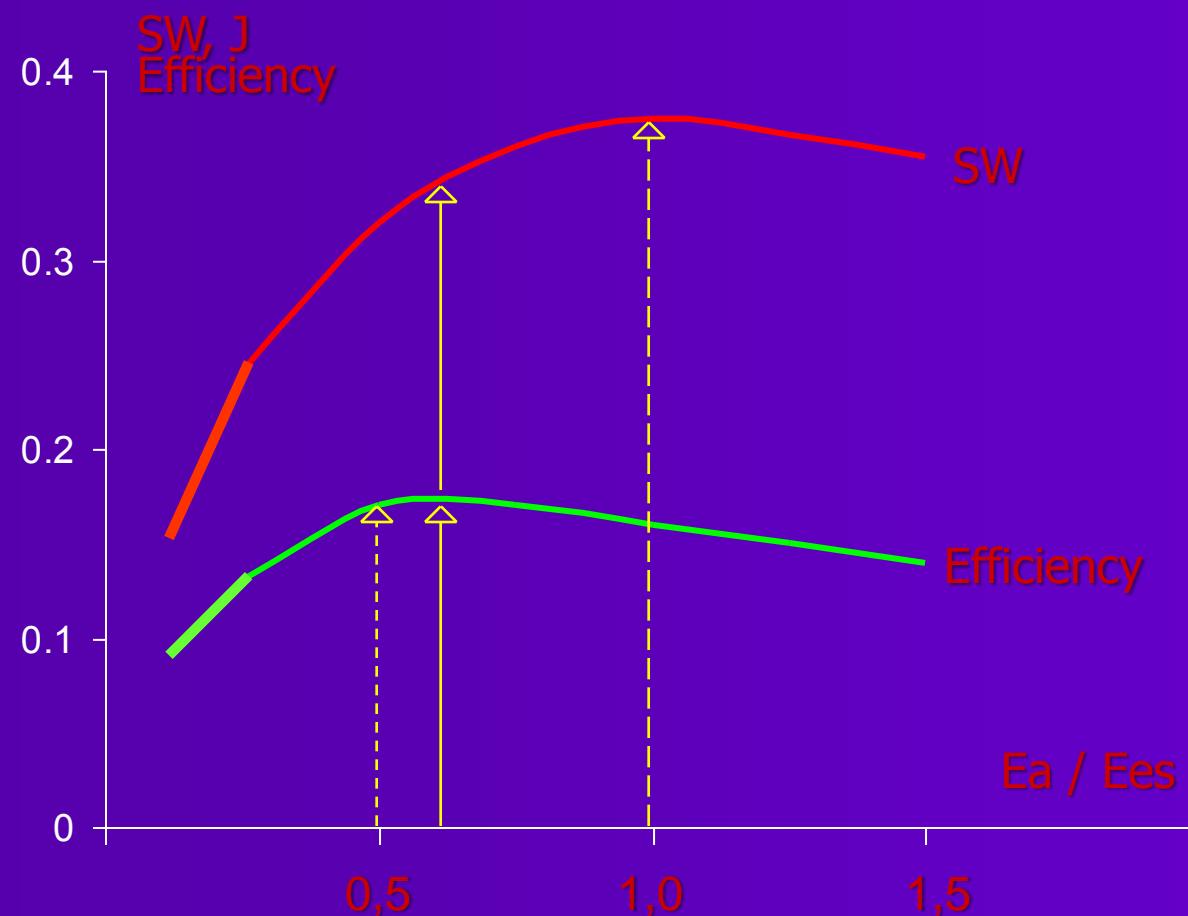
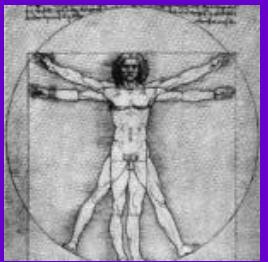




# Sagava, Suga. P-V диаграмма левого желудочка сердца



# Dependency of LV stroke work (SW) and its efficiency on ratio of aorta effective elastance (Ea) and LV contractility (Ees)



**Использование распознанных семейств зависимостей в качестве номограмм дает представление о характере протекающих процессов и о текущем состоянии объекта в этом процессе.**

**Появляется возможность контролировать параметры, не поддающиеся непосредственному измерению.**

# Мышечные артериальные сосуды – эволюция представлений

Артерии мышечного типа – квазипассивные сосуды с медленно изменяющимся просветом.

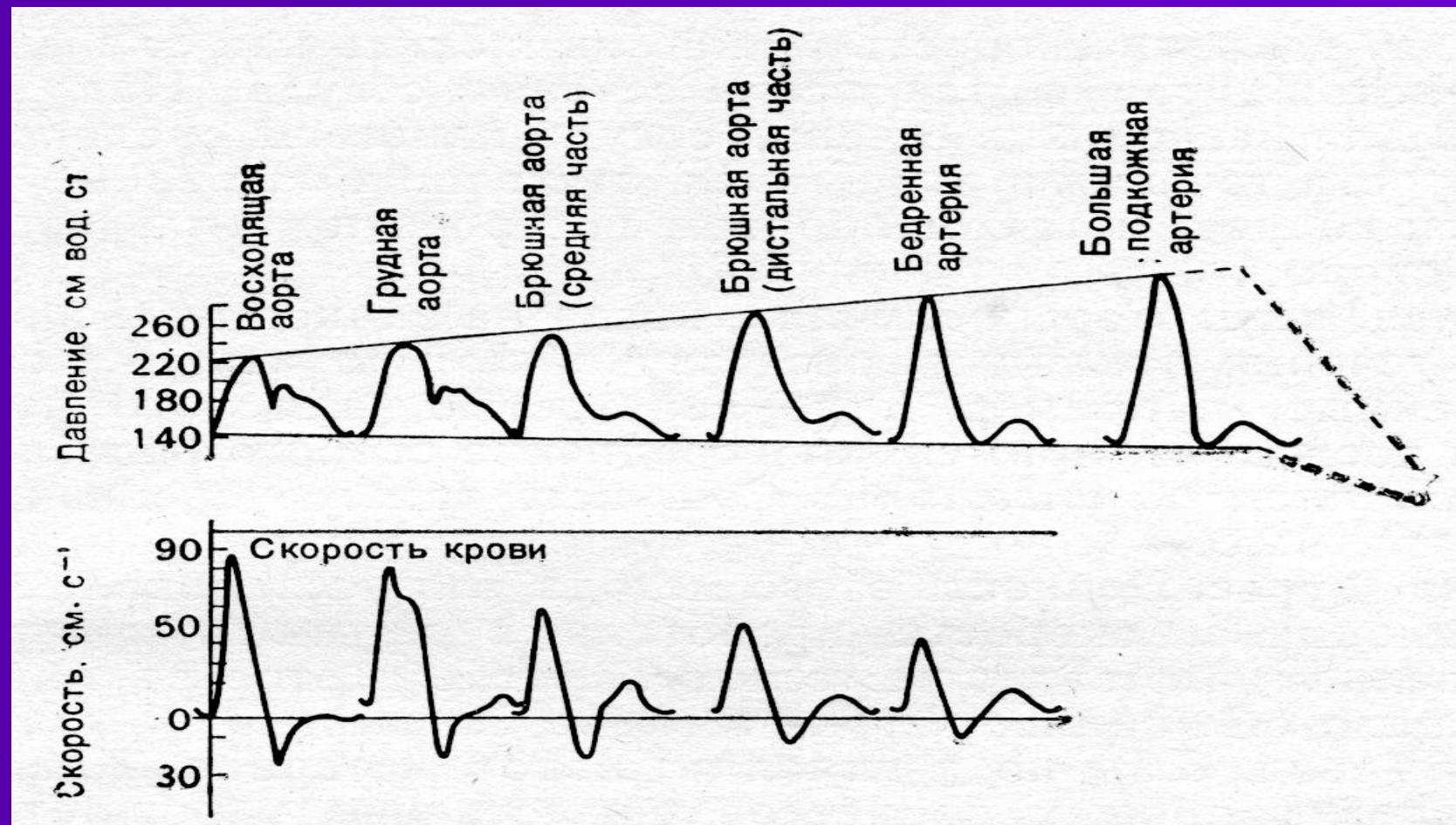
Основное падение давления от сердца к периферии обусловлено их высоким сопротивлением (А.А. Богомолец).



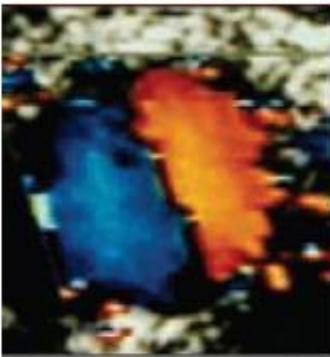
Артерии мышечного типа - активные сосуды, увеличение просвета которых приводит к падению сопротивления и падению давления соответственно, что приводит к снижению энергозатрат на перемещение крови.

Функционируют активно, совершают перистальтические движения синхронизированно с пульсацией кровотока (корректируются локальной метаболической ситуацией и(или) РАС).

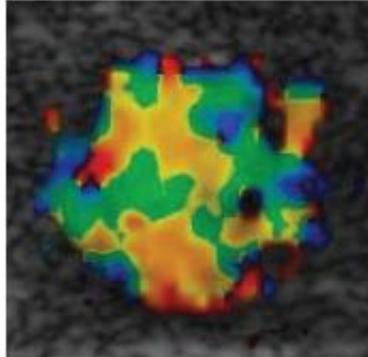
# Сопоставление кривых давления и скорости крови в разных участках артериального русла собаки



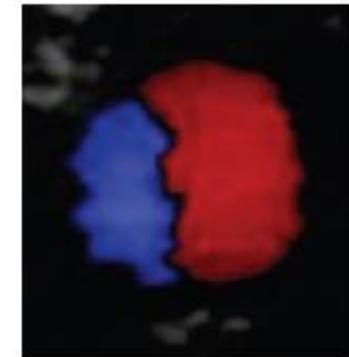
*Natural Blood Flow  
Through a Healthy  
Artery*



*Kinetic Disturbance  
– Blood Flow  
Through a Standard  
Graft*

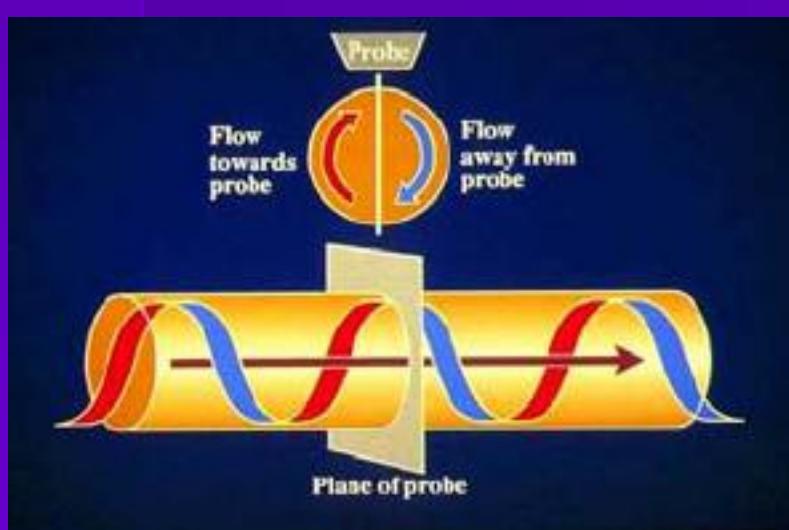


*Spiral Laminar Flow  
(SLF™) of Blood  
Through the Spiral  
Flow Graft*



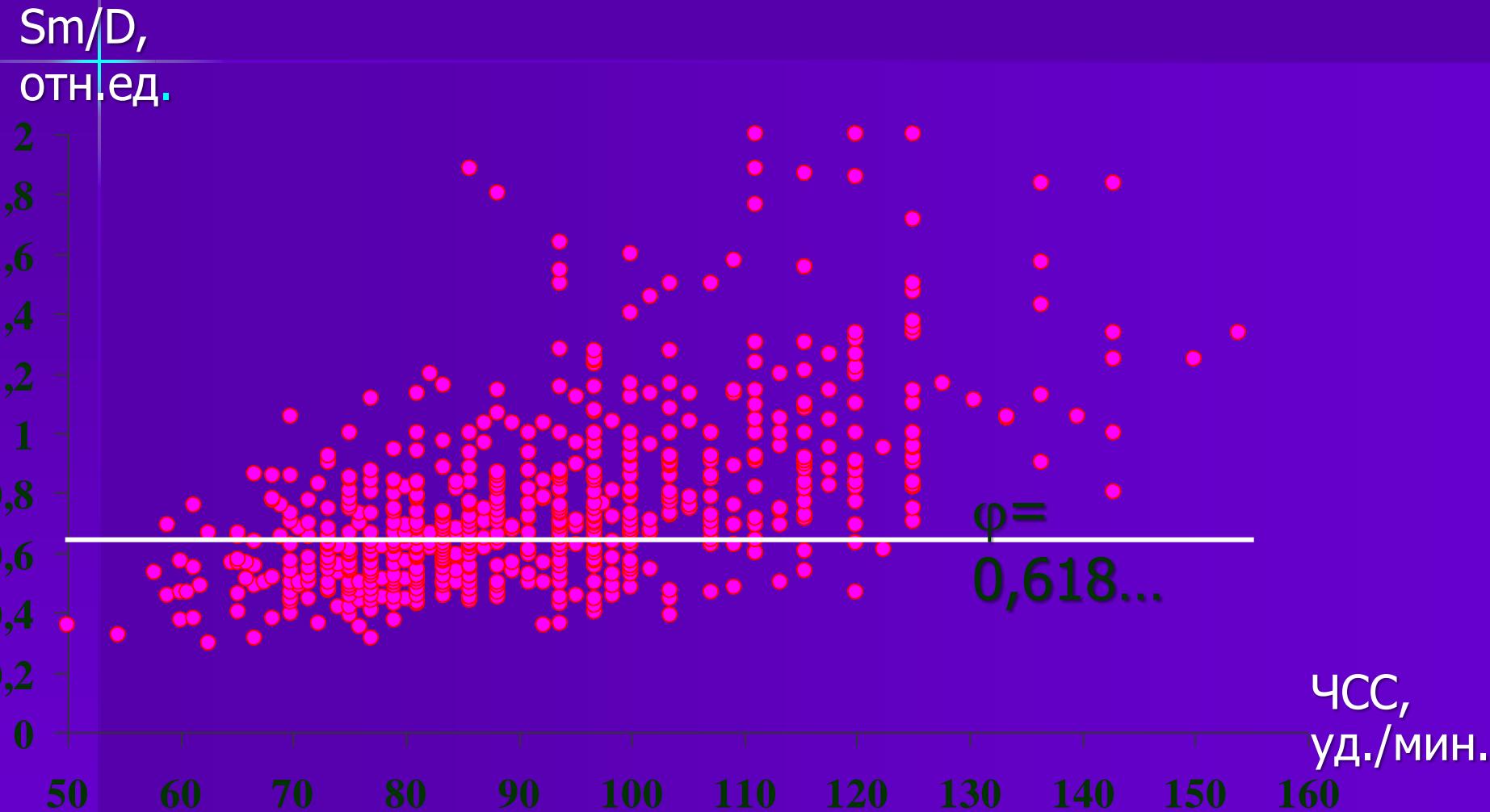
## **Spiral Laminar Flow (SLF™)**

### **A Revolution in the Treatment of Peripheral Artery Disease**



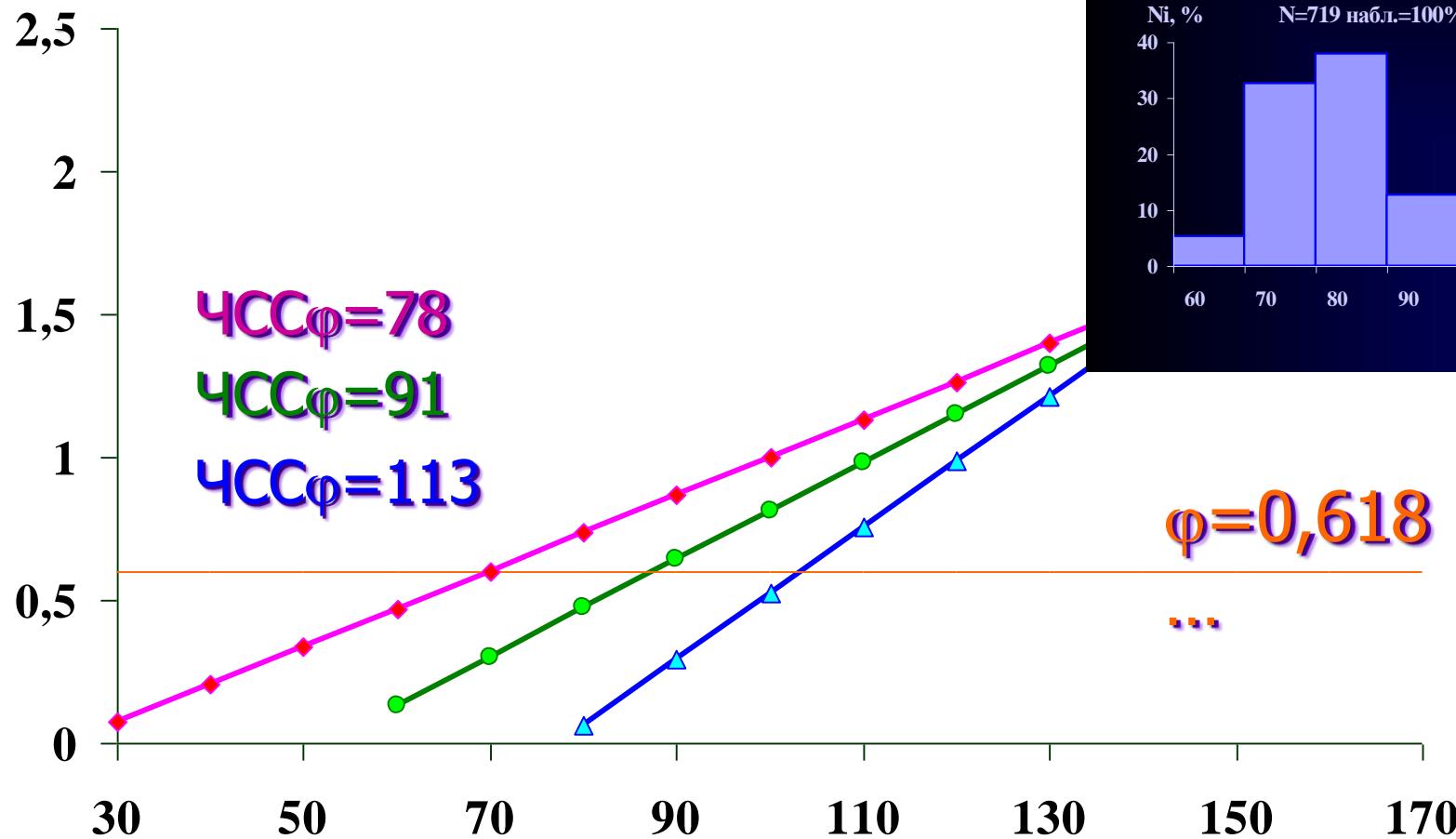
- A real alternative to vein harvesting
- Reduces operating times and infection risk
- Potential to reduce downstream disease progression
- Improved patient quality of life
- Very High Patency (up to 90%)

# Скатерограмма отношения Sm/D(ЧСС) («работа-отдых ЛЖ»)

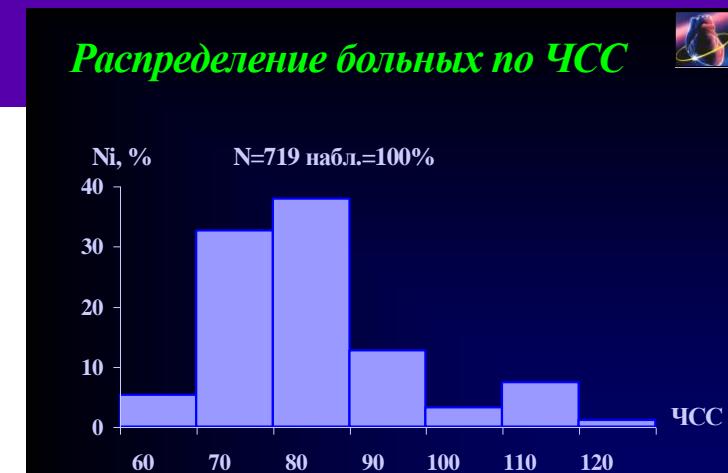


# Семейство регрессионных зависимостей $Sm/D(\text{ЧСС})$

$Sm/D$ ,  
отн.ед



Распределение больных по ЧСС

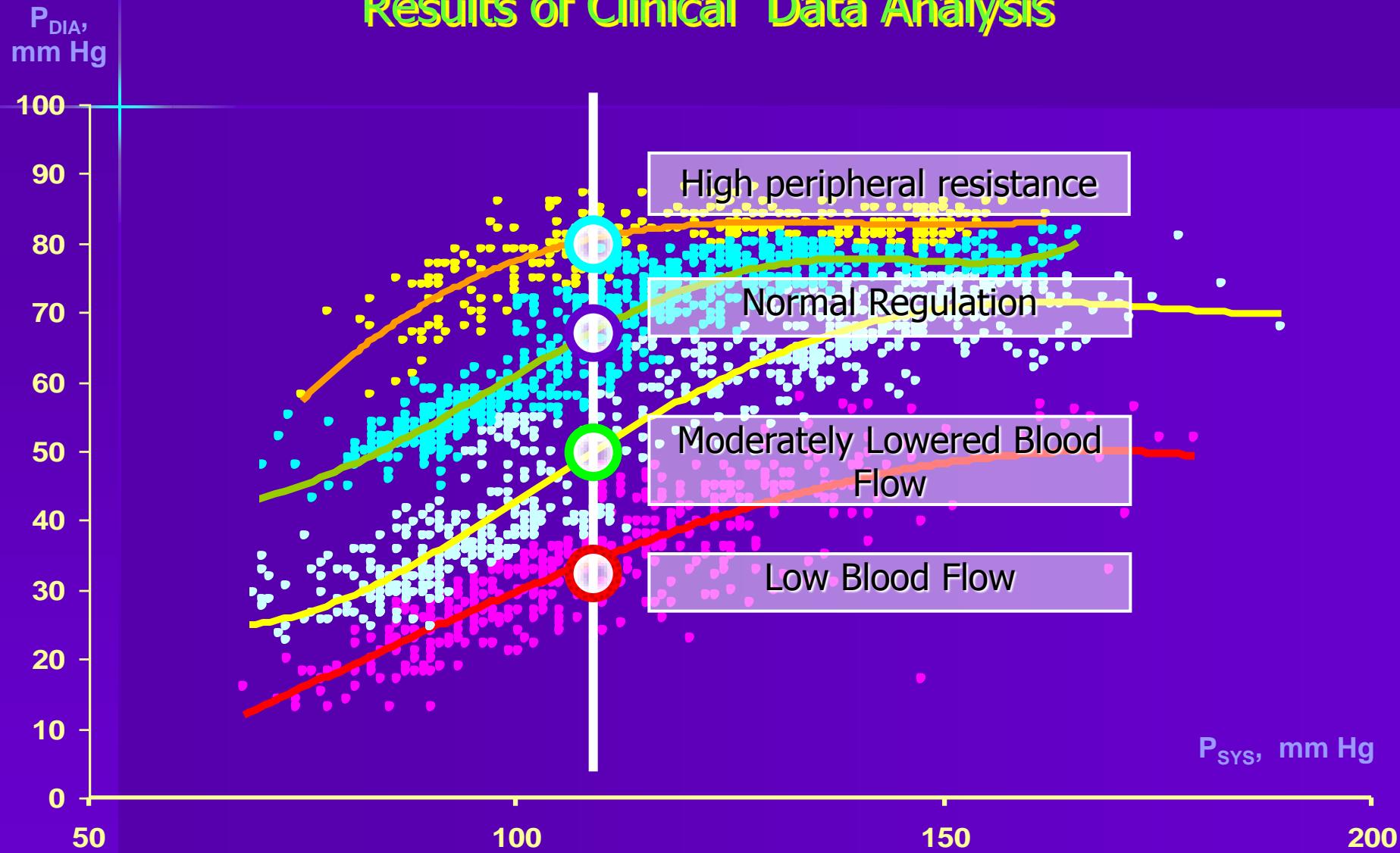


$\varphi=0,618$

ЧСС,  
уд./мин.

# The Types of Systolic & Diastolic Arterial Pressure Interrelationships

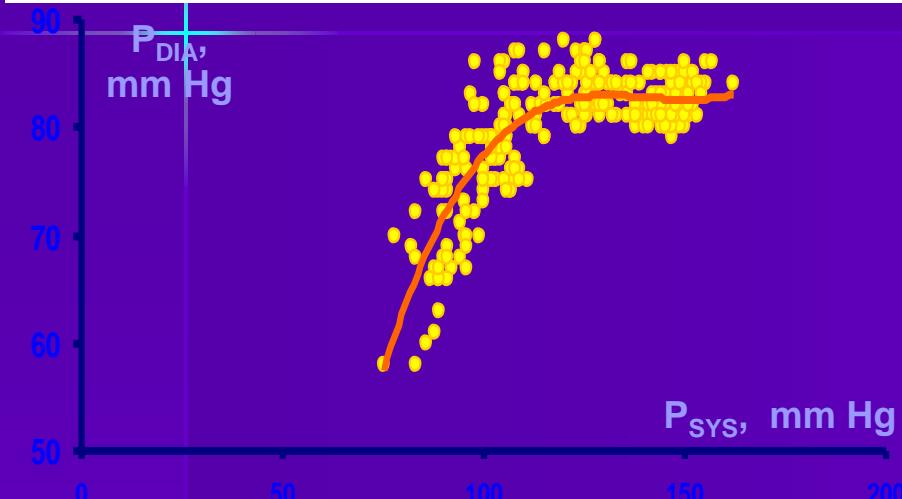
## Results of Clinical Data Analysis



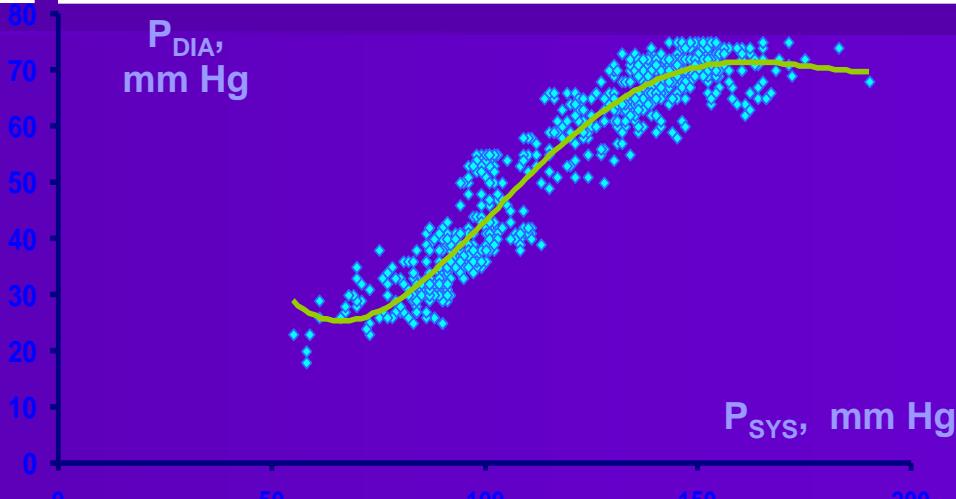
# The Types of Systolic & Diastolic Arterial Pressure Interrelationships

## Results of Clinical Data Analysis

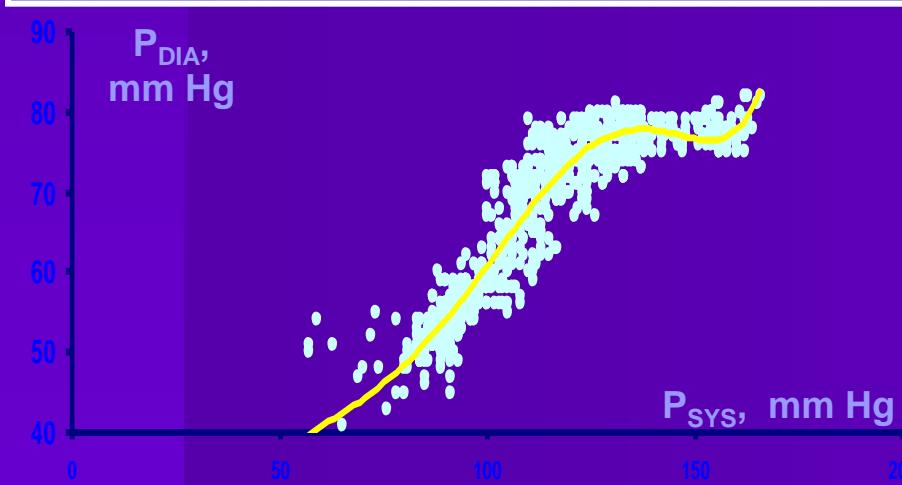
High peripheral resistance



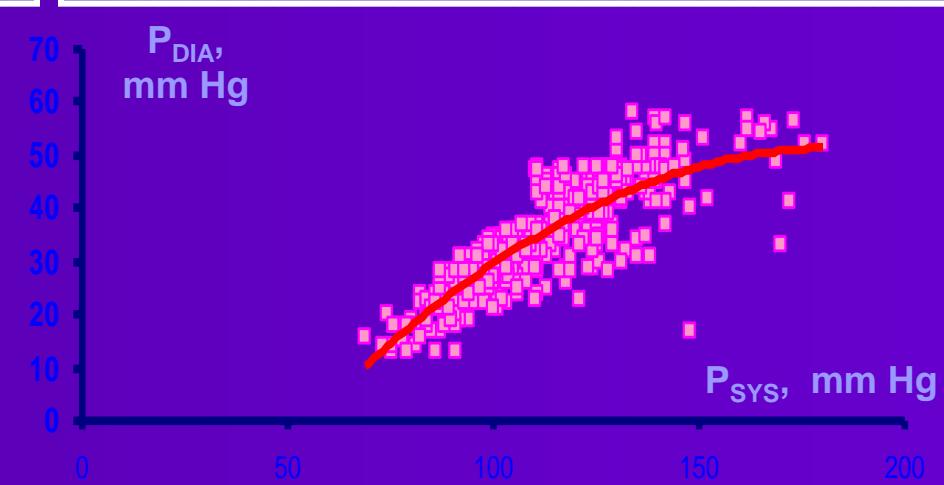
Normal Regulation



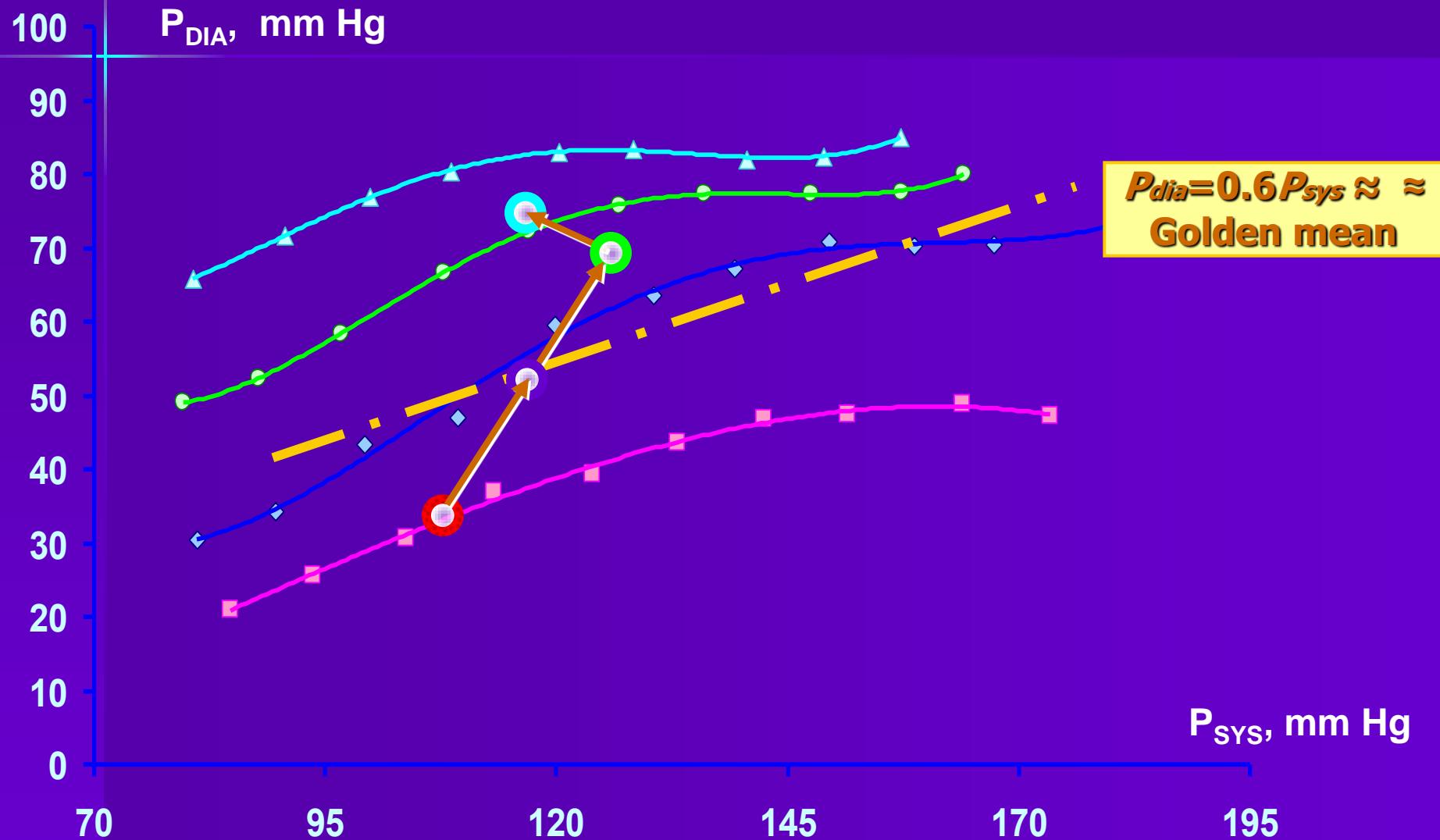
Moderately Lowered Blood Flow



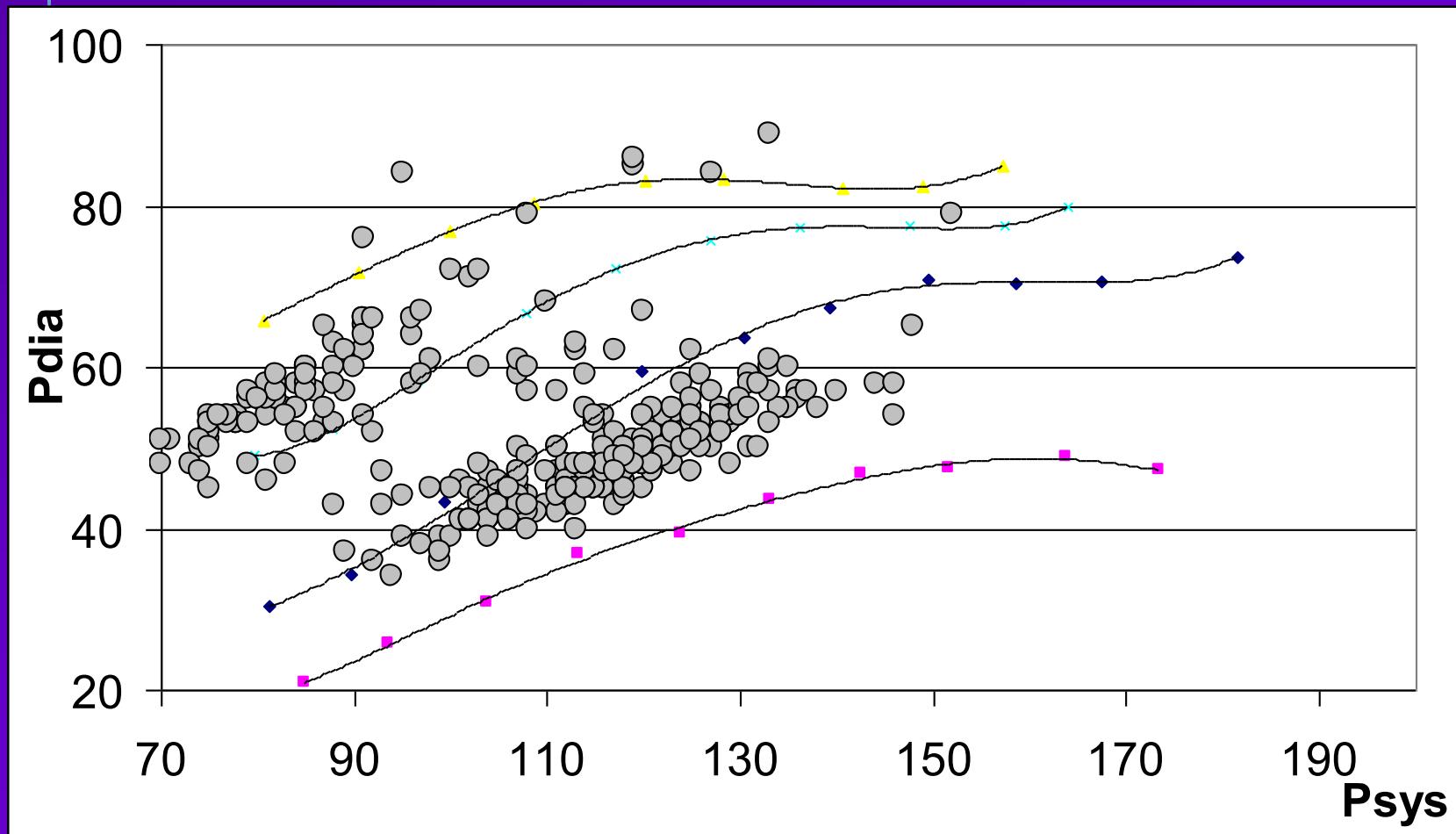
Low Blood Flow



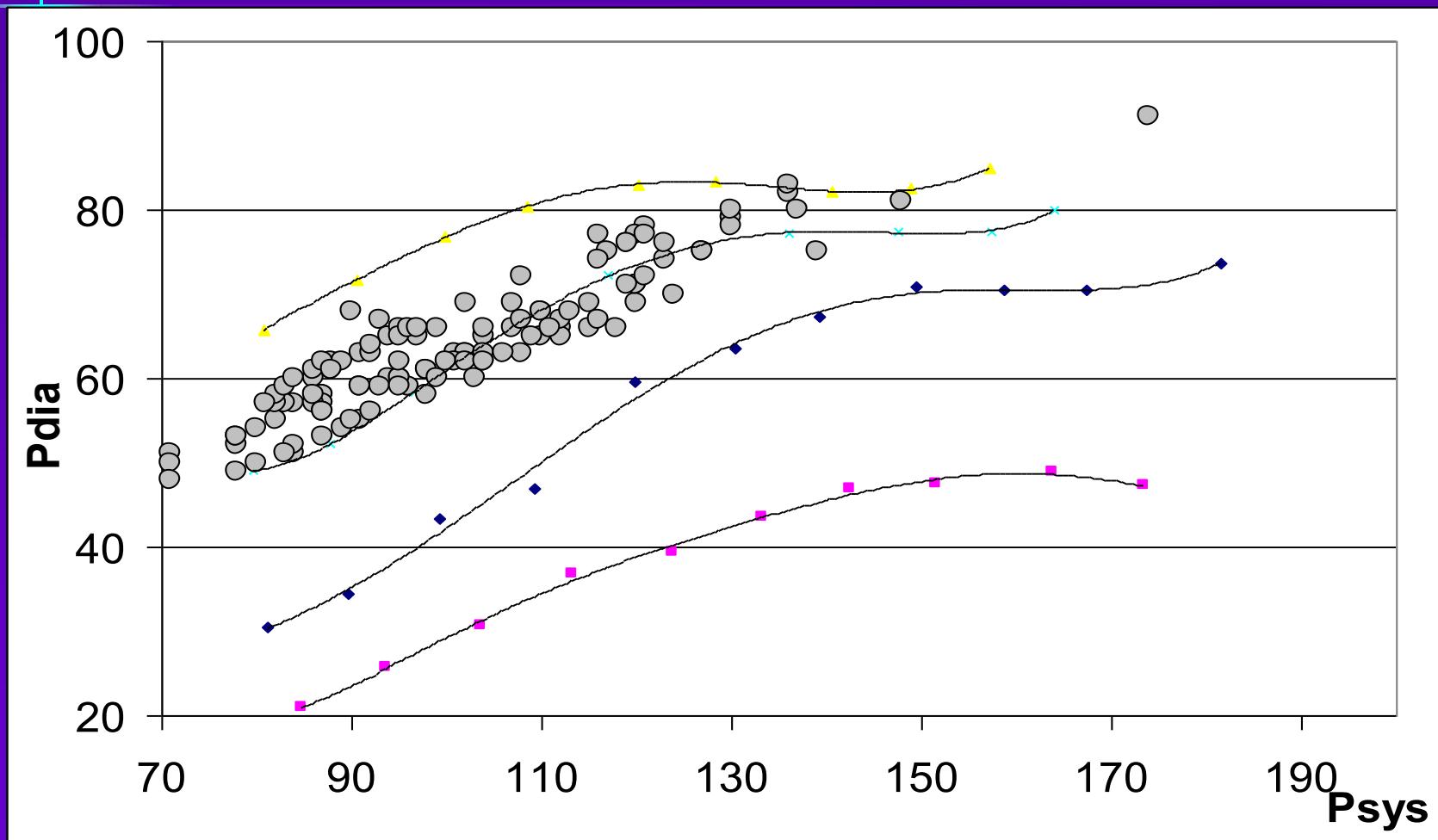
# Nomogram for Estimation of Blood Flow Adequacy



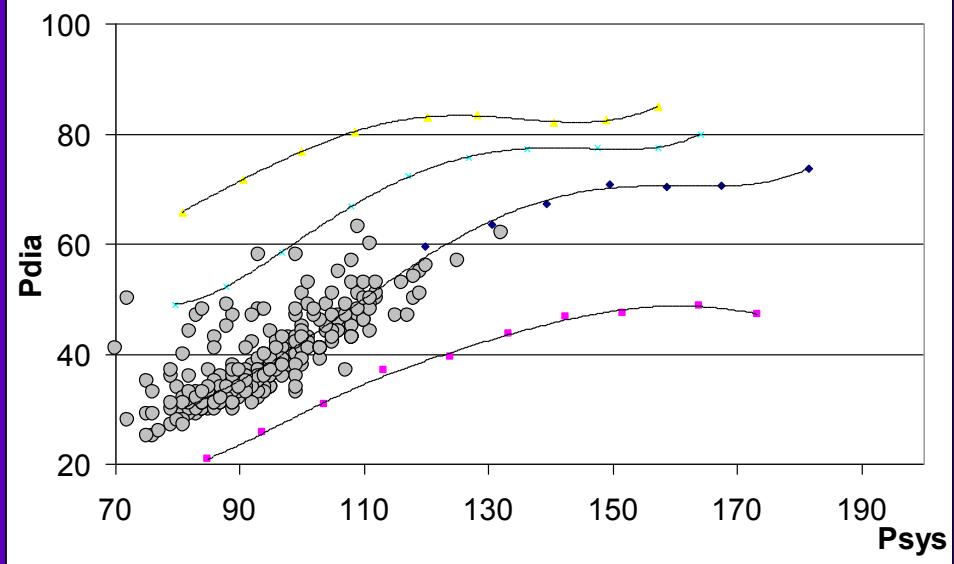
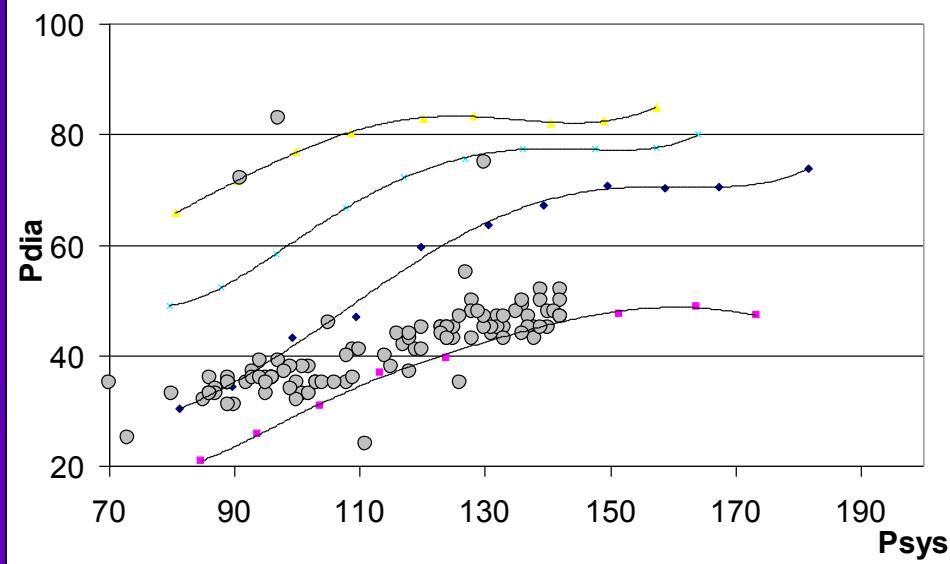
# ICU. Restoration of hemodynamics from low cardiac output syndrome to normal regulation level



# NONINVASIVE MONITORING OF HEALTHY PERSON

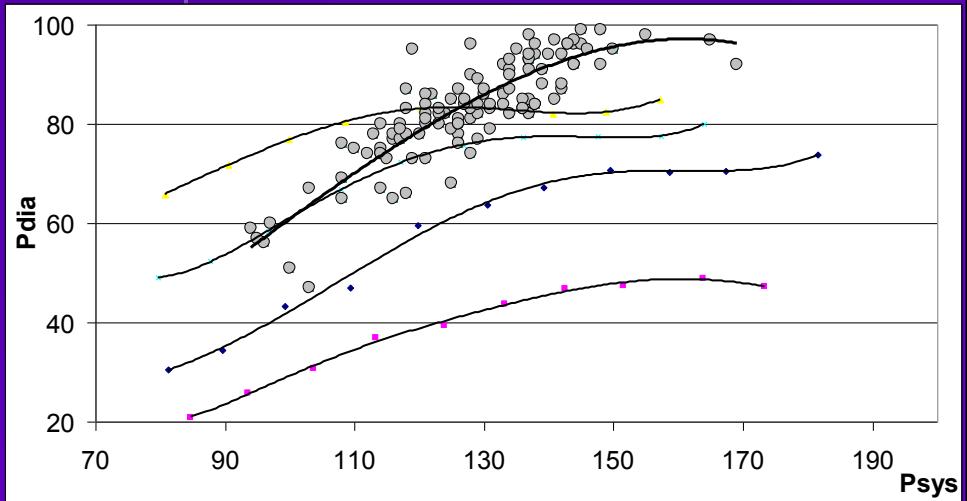


# ICU. HEART FAILURE

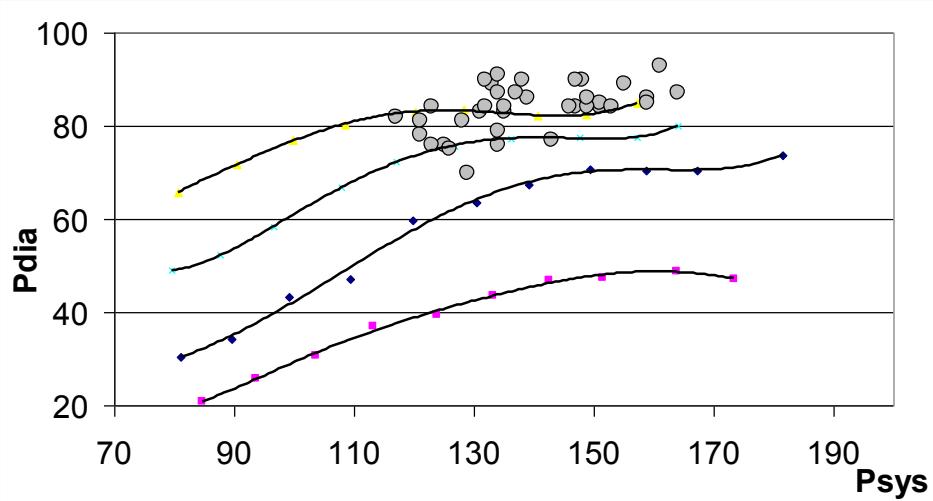


# INDIVIDUAL NONINVASIVE MONITORING OF

1 – healthy person; 2 – arterial hypertension



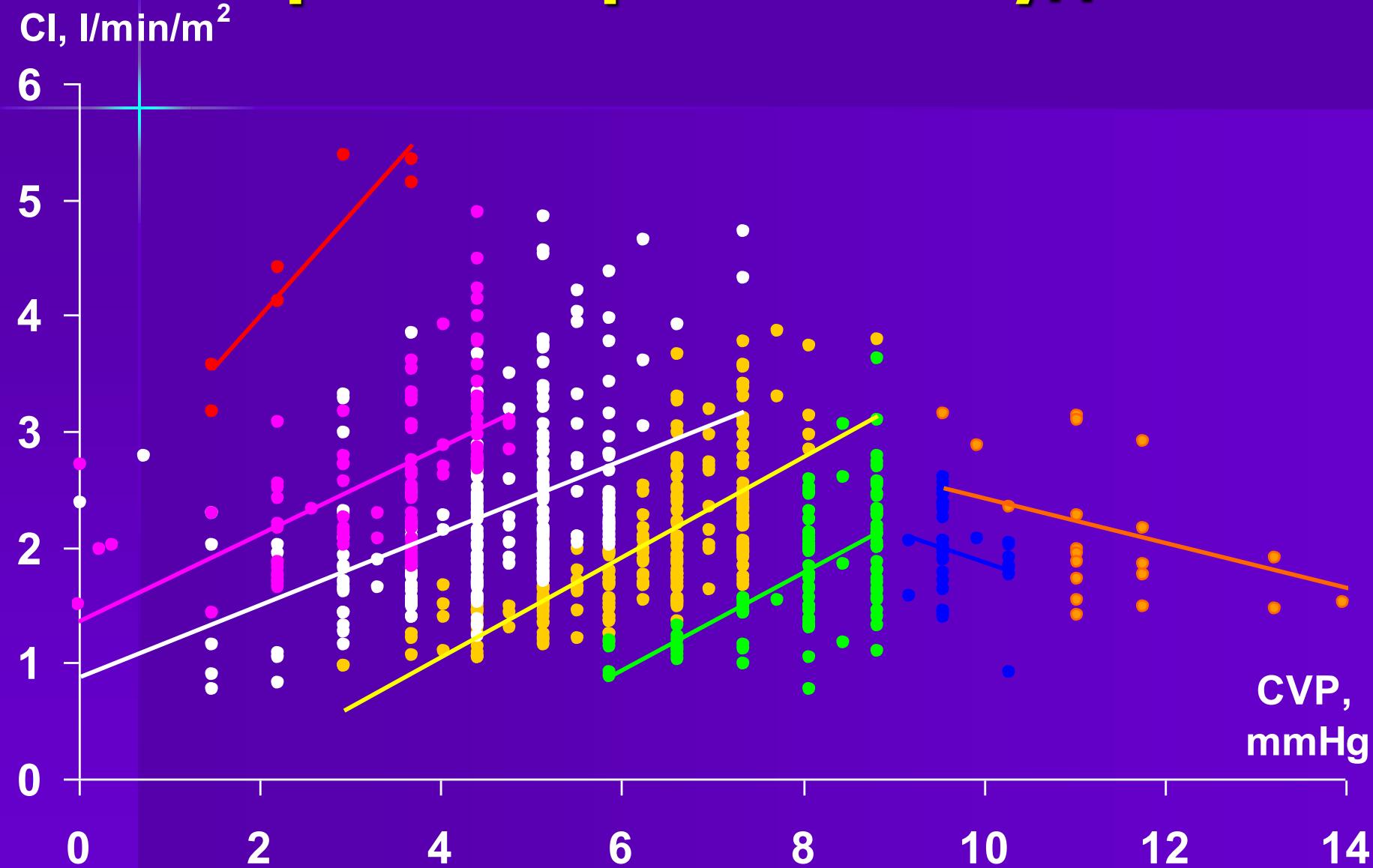
1



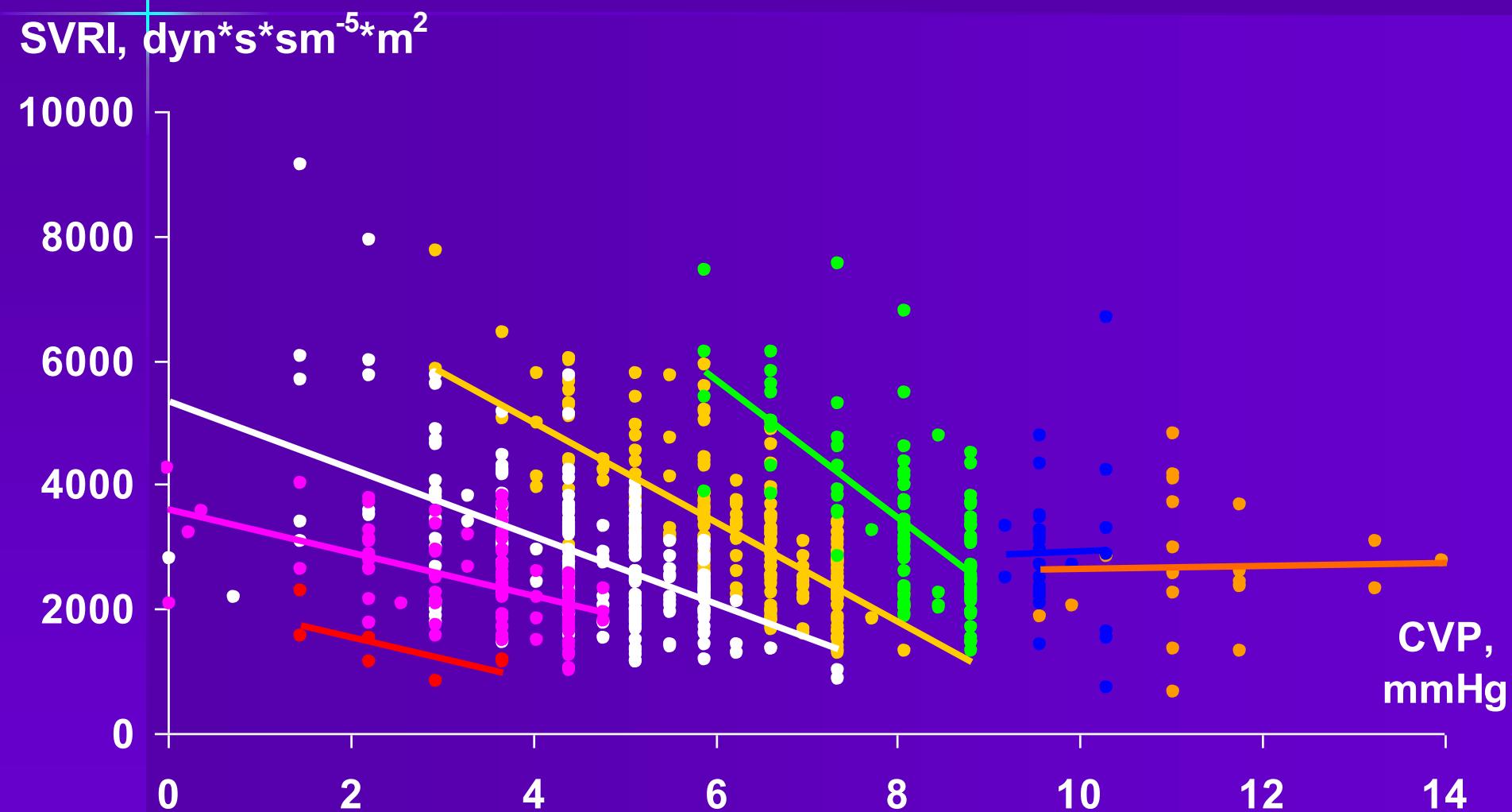
2

**Восстановление семейства  
зависимостей Франка-  
Старлинга правого  
желудочка сердца человека  
Интактная ССС.**

# Семейство зависимостей Франка-Старлинга правого желудочка



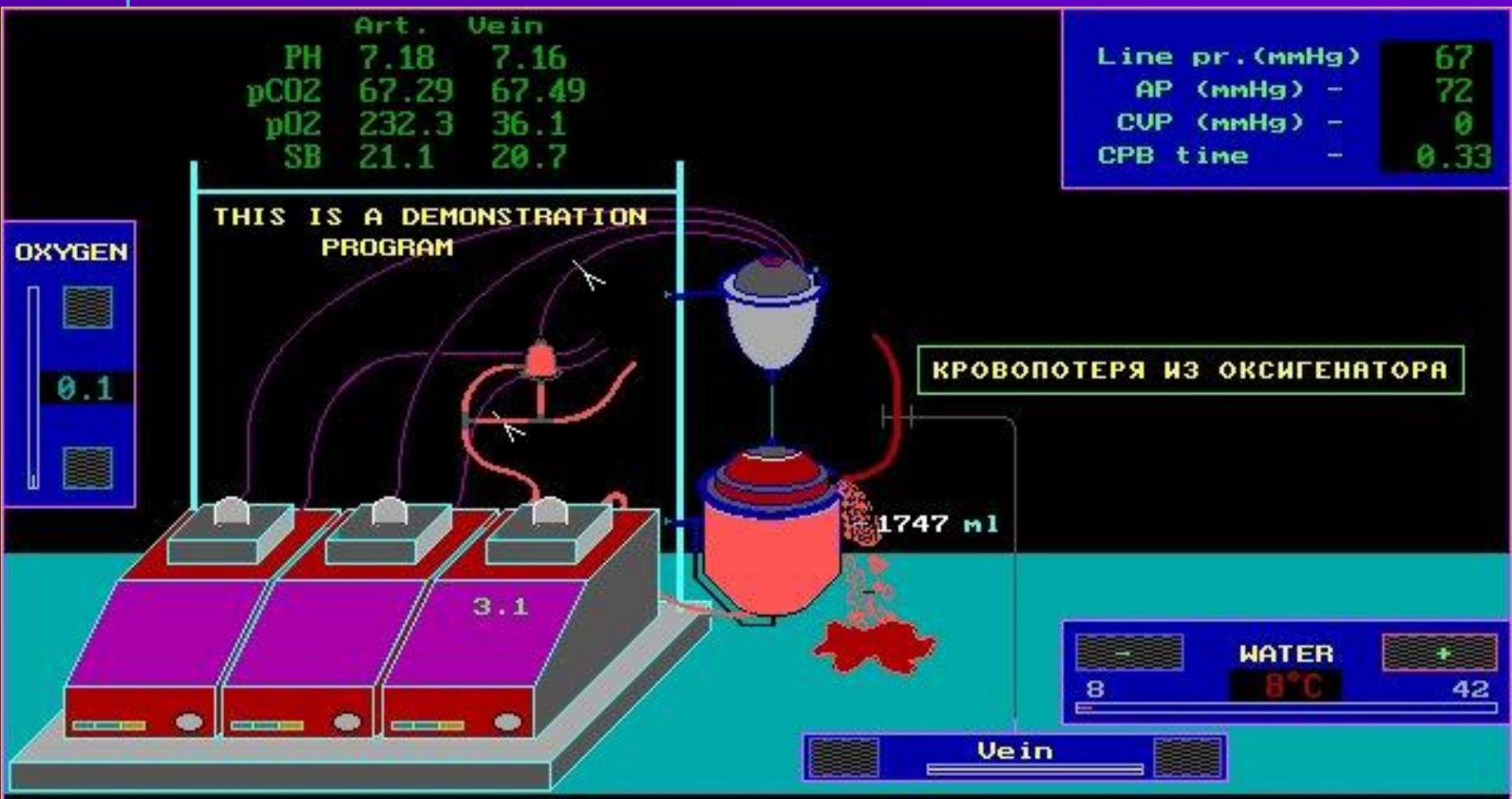
# Семейство зависимостей венозное давление (CVP) - периферическое сосудистое сопротивление (SVRI) в интактной ССС человека



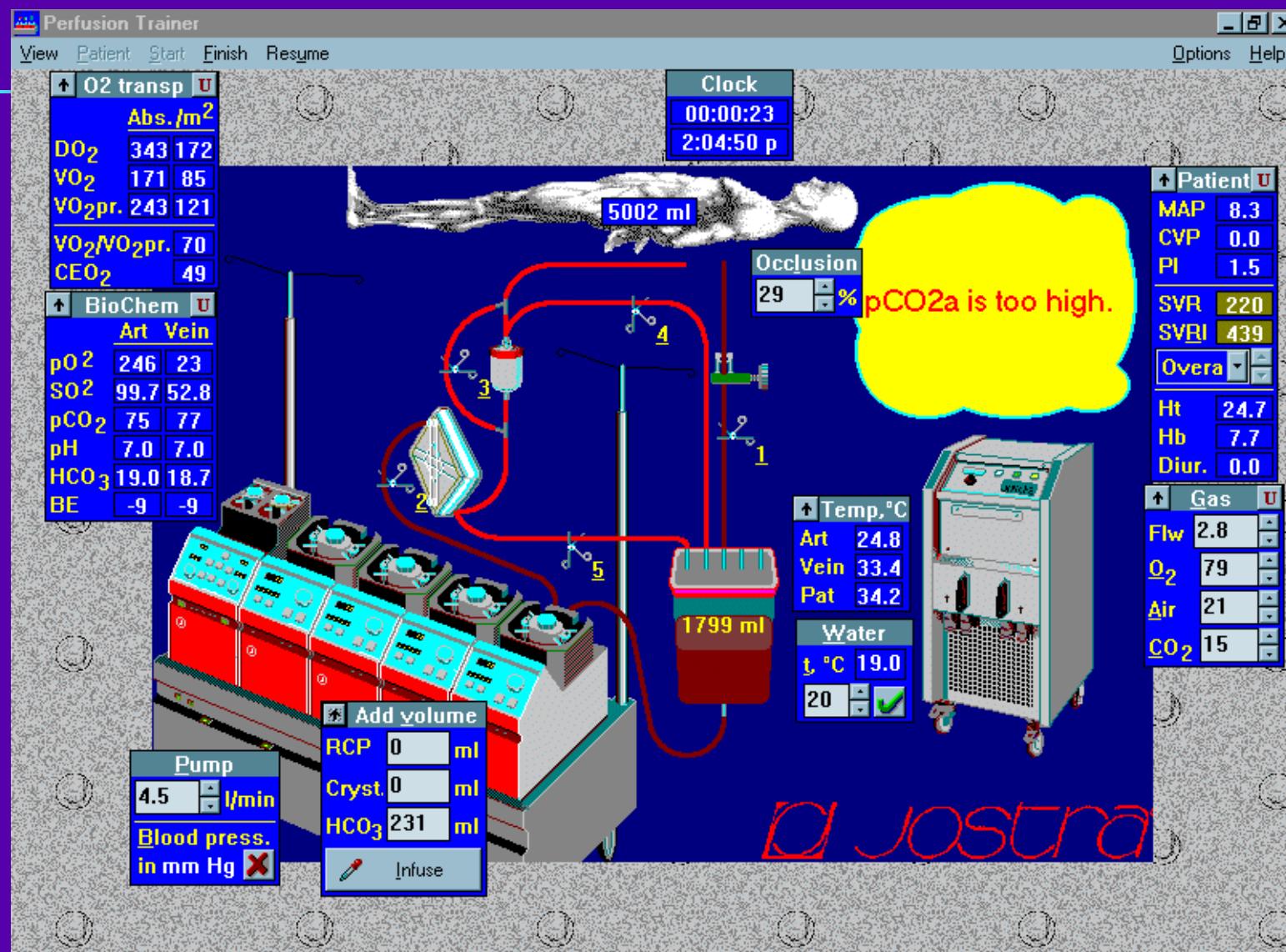
# **Моделирование и виртуальная реальность**

**Создание обучающих и  
диагностических систем  
нового поколения.**

# Симулятор искусственного кровообращения, 1993 г.



# Симулятор искусственного кровообращения, 1997 г.



# Симулятор искусственного кровообращения, 1999 - 2001 (www.perfusion.com)



**Изменение подходов к накоплению и обработке биомедицинской информации позволяет значительно расширить пределы возможного в области фундаментальных биомедицинских исследований и клинических испытаний, не выходя при этом за пределы допустимого и при существенной экономии материальных затрат**

**В чем же специфичность  
применения ИТ при  
исследованиях  
биомедицинских объектов?**

- 1. Что такое «однородная группа наблюдений» при наличии функциональной связи показателей?**
- 2. Нужно ли отфильтровывать наблюдения или стандартизировать условия в процессе сбора данных?**
- 3. Как создавать выборки, однородные в смысле функциональной связи между показателями?**

**Создание т.н. однородных массивов наблюдений не является абсолютно необходимым при применении современных информационных технологий.**

**Соответственно, объем необходимых контрольных групп наблюдений может быть существенно сокращен и оптимизированы процессы получения эмпирической и экспериментальной информации.**

# Способы накопления, обработки и использования биомедицинских данных

1. Создание законодательных механизмов анонимизированного накопления и последующего использования данных обследований, полученных в ходе рутинного лечения больных.

# *Способы накопления, обработки и использования биомедицинских данных*

2. Объединение ранее накопленных различными учреждениями и научными коллективами однородных биомедицинских данных, независимо от условий получения последних, формирование на их основе Интернет-ориентированных баз биомедицинских данных коллективного доступа.

# *Способы накопления, обработки и использования биомедицинских данных*

- 3. Предобработка данных биомедицинских систем командами специалистов в области биоинформационных технологий (напр. - выявление структуры пространства состояний и т.п.).**
- 4. Совместное использование и постоянное пополнение сформированных массивов наблюдений и результатов их обработки различными научными и лечебными учреждениями.**

# *Способы накопления, обработки и использования биомедицинских данных*

5. Постоянное сопровождение научных проектов профессиональными коллективами специалистов в области информационных технологий.

**Алгоритм обработки данных** – математический метод извлечения из массива наблюдений информации определенного содержания.

**Информационная технология** – комбинация алгоритмов (и методов) обработки данных.

**Математические модели позволяют  
исследовать то, что нельзя  
непосредственно измерить, выявить  
то, чего не может быть,  
сформулировать требования к  
этапным экспериментальным  
исследованиям**

# Благодарю за внимание!

