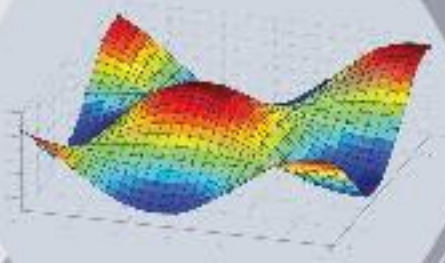
DESIGN SPACE EXPLORATION FOR EVERY EXPERTISE



Key technologies for design optimization, predictive modeling and data analysis

Dmitry Frolov,
Marketing Director
DATADVANCE

September, 2016

DATADVANCE

DATADVANCE

Sk Сколково

Design optimization, predictive modeling and data analysis based on pSeven technology support the implementation of the Digital Railway program initiated by RZD

DATADVANCE

Проектно-конструкторско-технологическое бюро по системам информатизации

Designing and Technological Office on Informatization Systems

Agenda

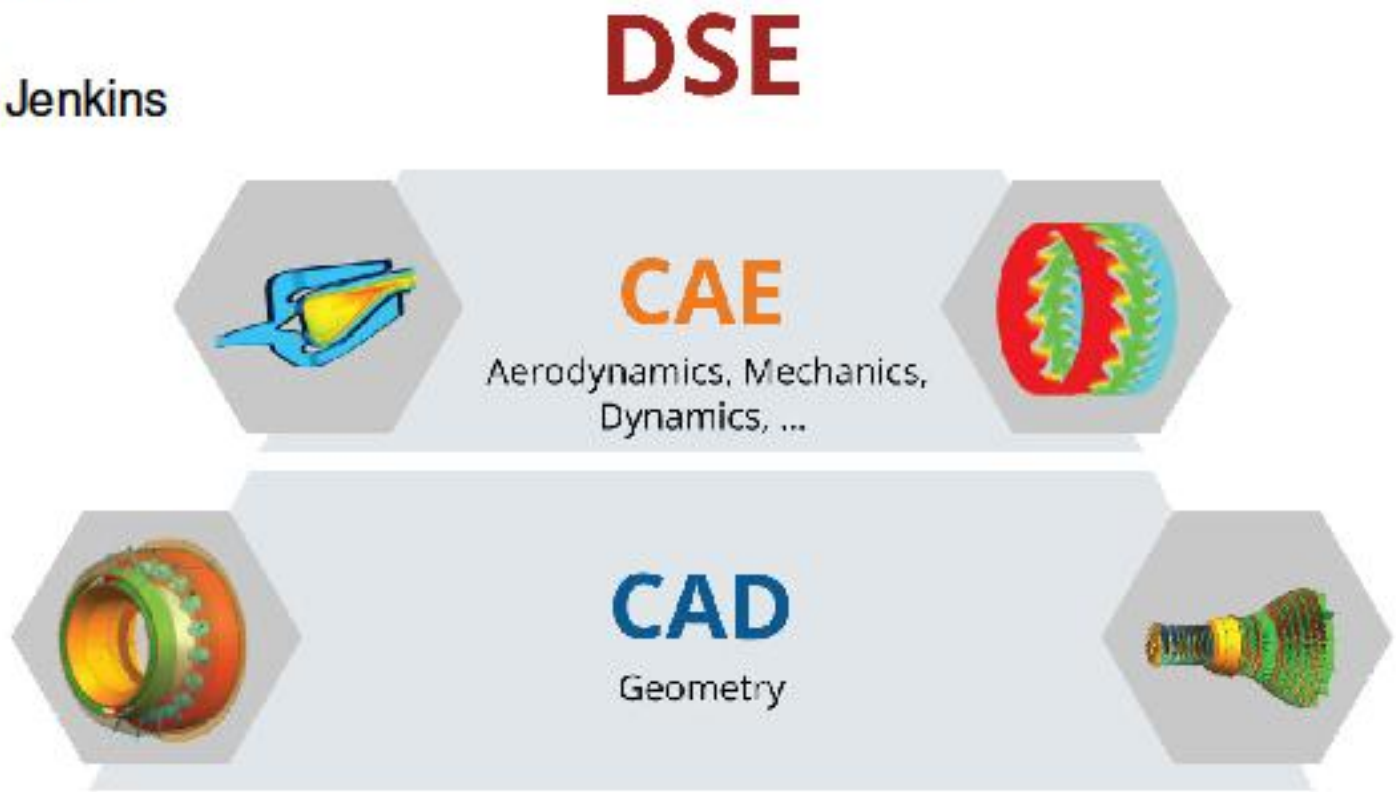
- **Design Space Exploration with pSeven**
  - Data & Model Analysis
  - Predictive Modeling
  - Design Optimization
- pSeven Platform
- Summary

DATADVANCE

Design Space Exploration

Design Space Exploration is both a class of quantitative methods and a category of software tools for **systematically and automatically** exploring very large numbers of design alternatives and identifying **optimal** performance parameters.

B. Jenkins



pSeven is Design Space Exploration Platform for Every Expertise

Design Space Exploration:

- Advanced mathematical algorithms and techniques

Platform:

- Powerful process integration environment
- Create, share and run simulation workflows
- Collect, manage and reuse engineering data

Every Expertise:

- Apply Simulation Driven Design methodology and design the best products even if you are not math expert and even not an engineer with SmartSelection™

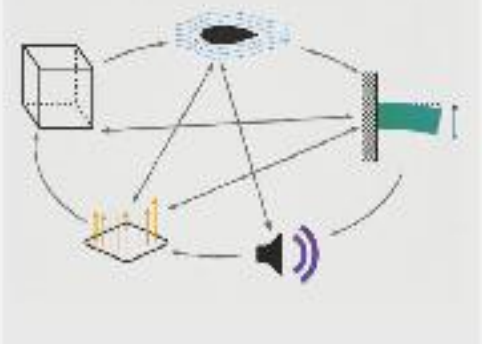
Advantages of pSeven application

- **Improve** your product performance, quality, reliability, safety.
- Significantly **reduce design lead time** and cost thanks to state-of-the-art algorithms.
- **Formalize** and preserve your knowledge, experience and design practices through automation.
- Improve **collaboration** between departments and engineers – one more step towards multidisciplinary design optimization.

DATADVANCE

pSeven components

Process Automation and Integration

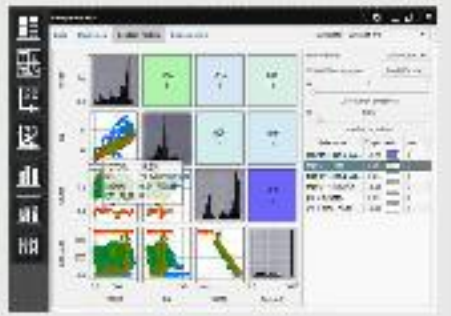


Powerful Workflow Execution Engine

pSeven Core

Efficient Data Mining and Optimization

Data Analysis and Visualization



pSeven Platform

DATADVANCE

Agenda

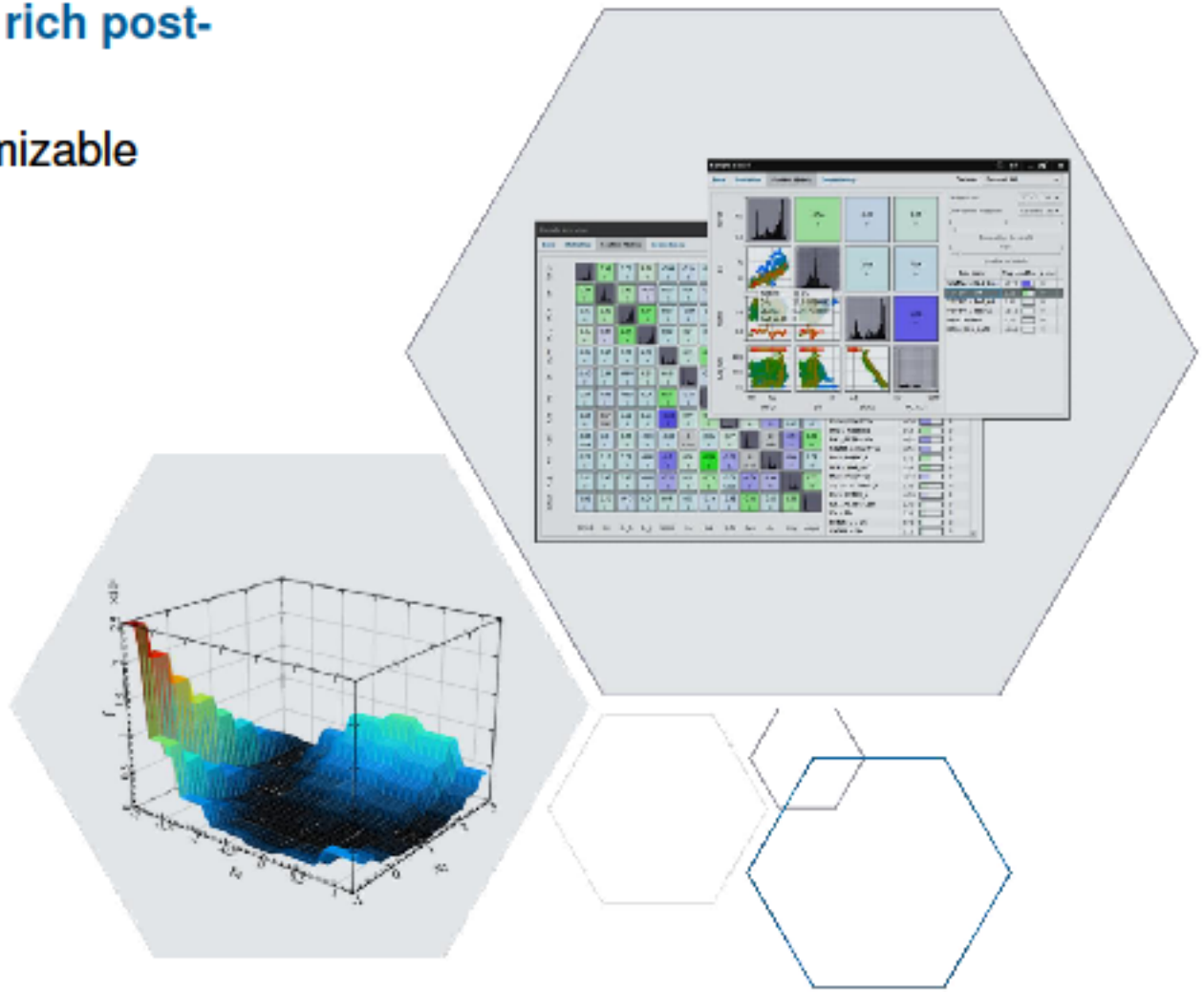
- Design Space Exploration with pSeven
  - **Data & Model Analysis**
  - Predictive Modeling
  - Design Optimization
- pSeven Platform
- Summary

DATADVANCE

Data analysis in pSeven

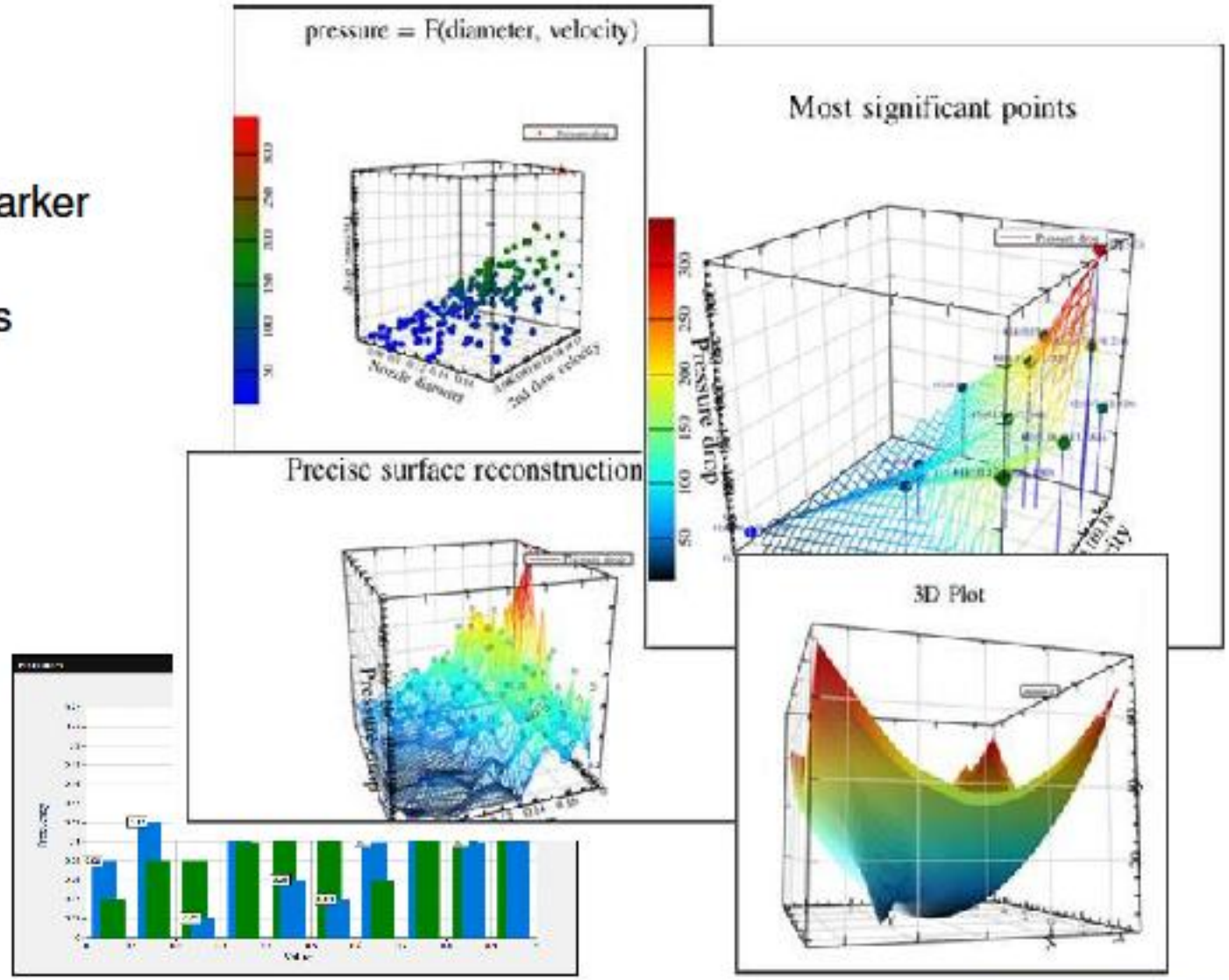
pSeven provides full control over external data and rich post-processing capabilities.

- Visualize results with rich set of interactive and customizable visualization tools:
  - 2D visualization
  - 3D visualization
  - Scatter Matrix
  - Tables and statistics
  - Dependency
  - Parallel coordinates
- Analyze results and other engineering data
- Visualization re-use



Visualize in 2D or 3D

- Interactively select and analyze data
- Put multiple samples on Histogram charts to compare and analyze frequency distributions
- Draw 2D point and line plots with rich set of marker and color styles
- Draw point clouds from 3-dimensional samples
- Use 4th dimension as color axis
- Reconstruct surfaces from unstructured data
- Zoom any area for details
- Customize visual styles and data filters

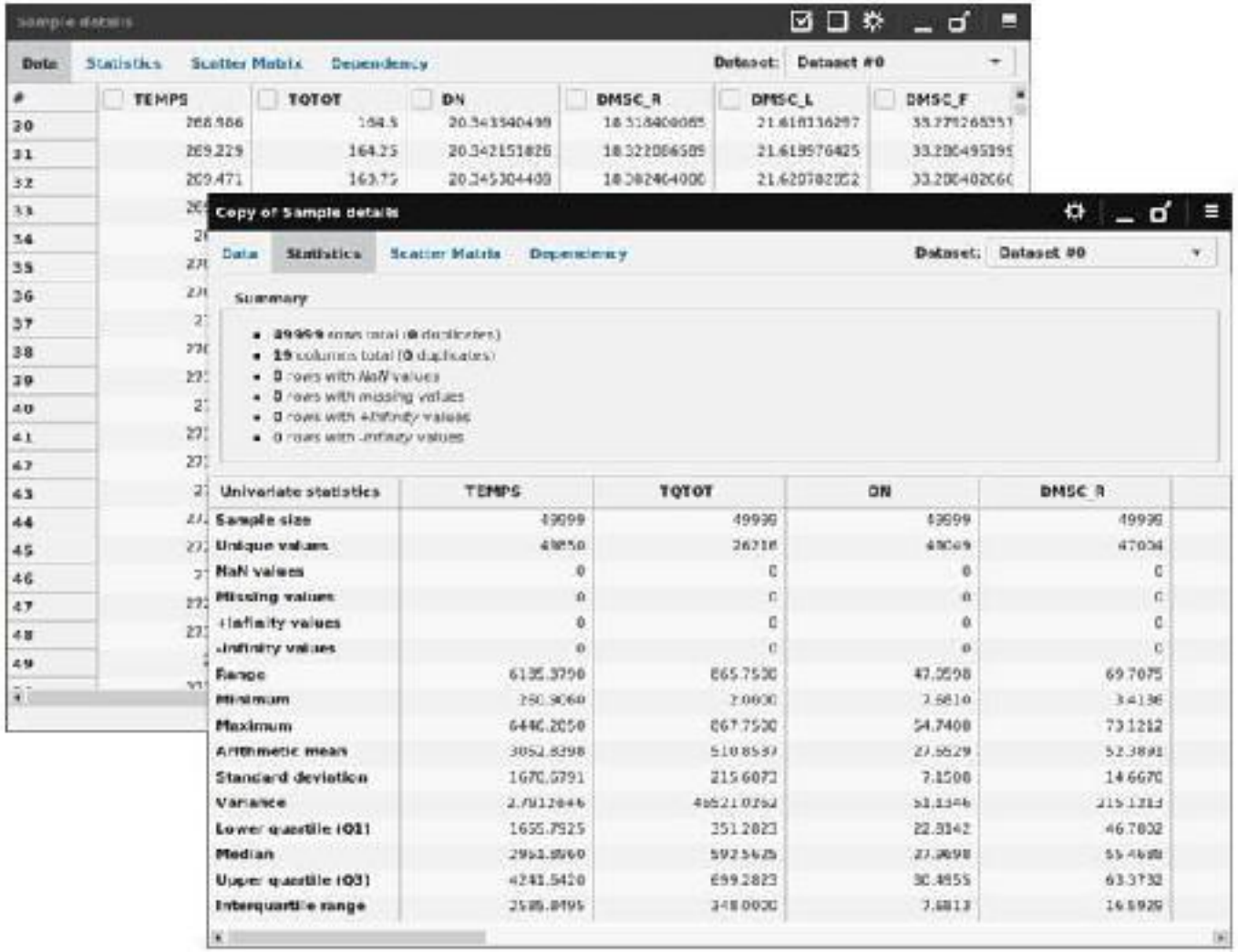


DATADVANCE

Tables and statistics

See raw data table and descriptive statistics for each data column:

- Sample size
- Unique values
- Variance
- Standard deviation
- Median
- Quartiles Q1 and Q3
- Interquartile range
- Range
- Minimum / Maximum
- NaN values
- Missing values
- +- Infinity values



DATADVANCE

Scatter matrix

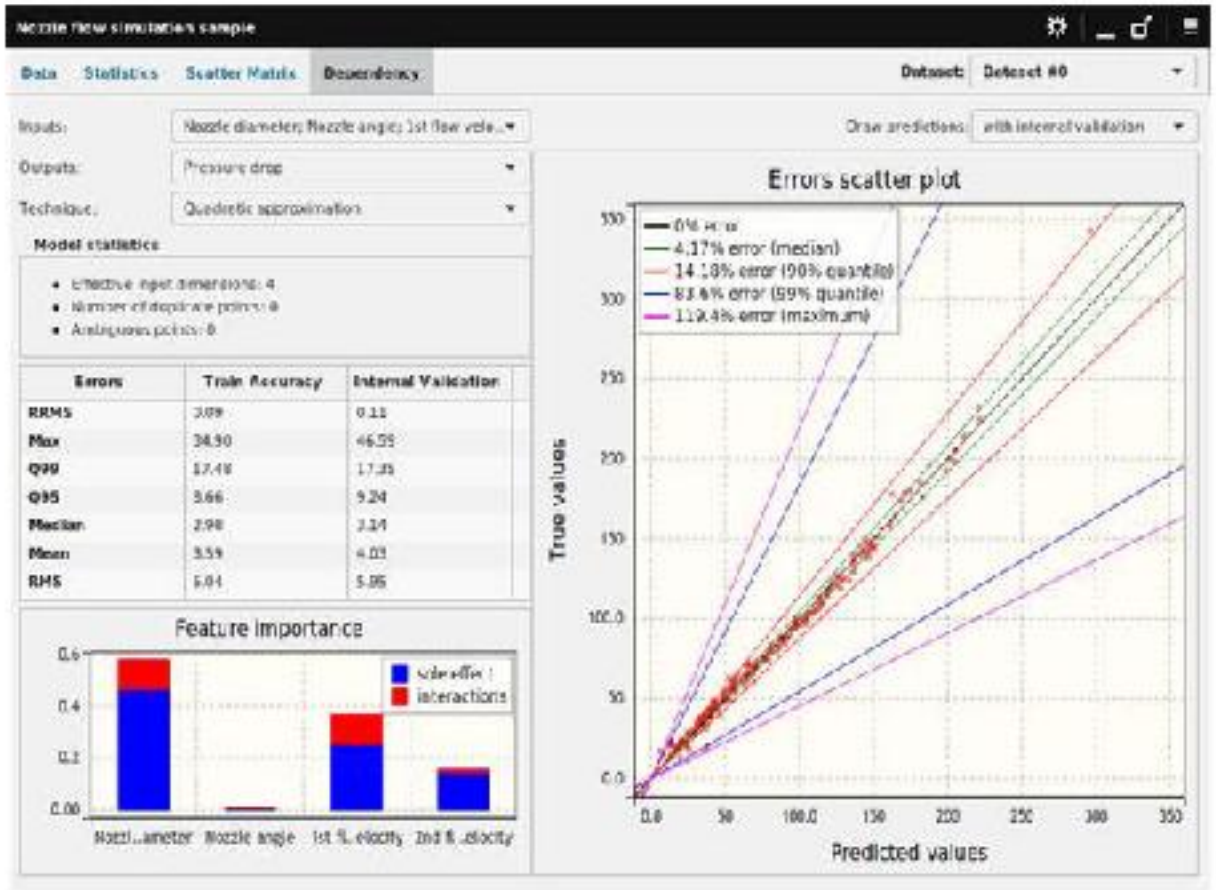
Analyze high-dimensional data and spot correlations with various measures of dependency:

- Pearson
- Spearman
- Kendall
- Mutual information
- Partial and distance correlations
- P-value estimation included

Dependency

Analyze functional dependencies in your data and models:

- Import data from Project Database or CSV/Excel
- Make Linear or Quadratic approximations
- Estimate and compare inputs features importance
- Assess quality of created dependency model

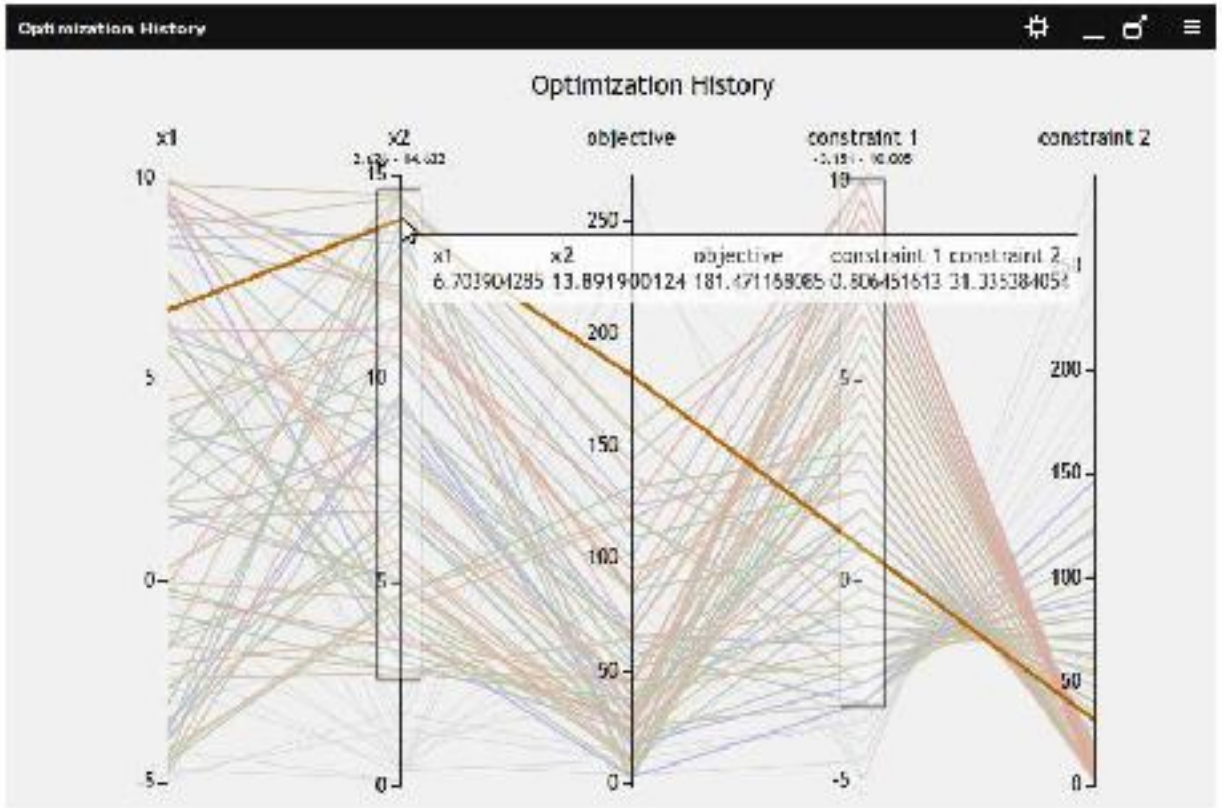


DATADVANCE

Parallel coordinates

Use parallel coordinates to visualize and analyze high-dimensional and multivariate data:

- Make slices with interactive range filter on each dimension's axis
- Highlight particular areas to discover individual points components values
- Experiment with interactive axes reordering to spot patterns and dependencies between components



DATADVANCE

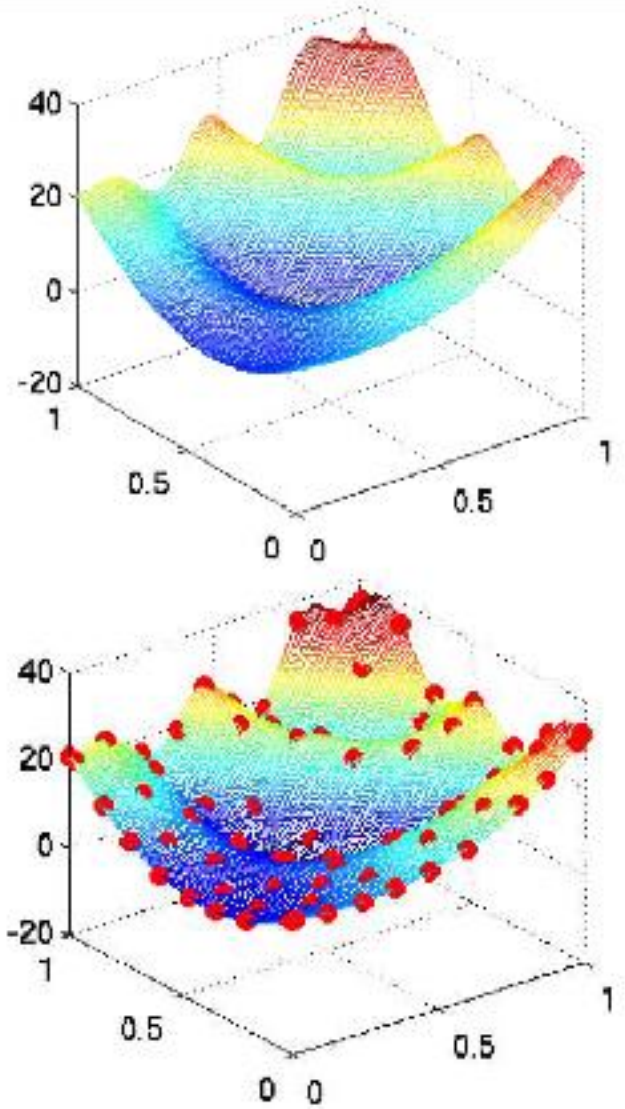
Design of Experiments (DoE)

Design of Experiments is a selection of inputs at which outputs are measured to achieve specific goals:

- **Explore design space** using as small number of observations as possible
- Get as much **information** as possible about the model behavior
- Measure output **sensitivity**, variability and other characteristics
- Perform reliable **surrogate-based optimization**
- Generate a training data sample for construction of an accurate **surrogate model**

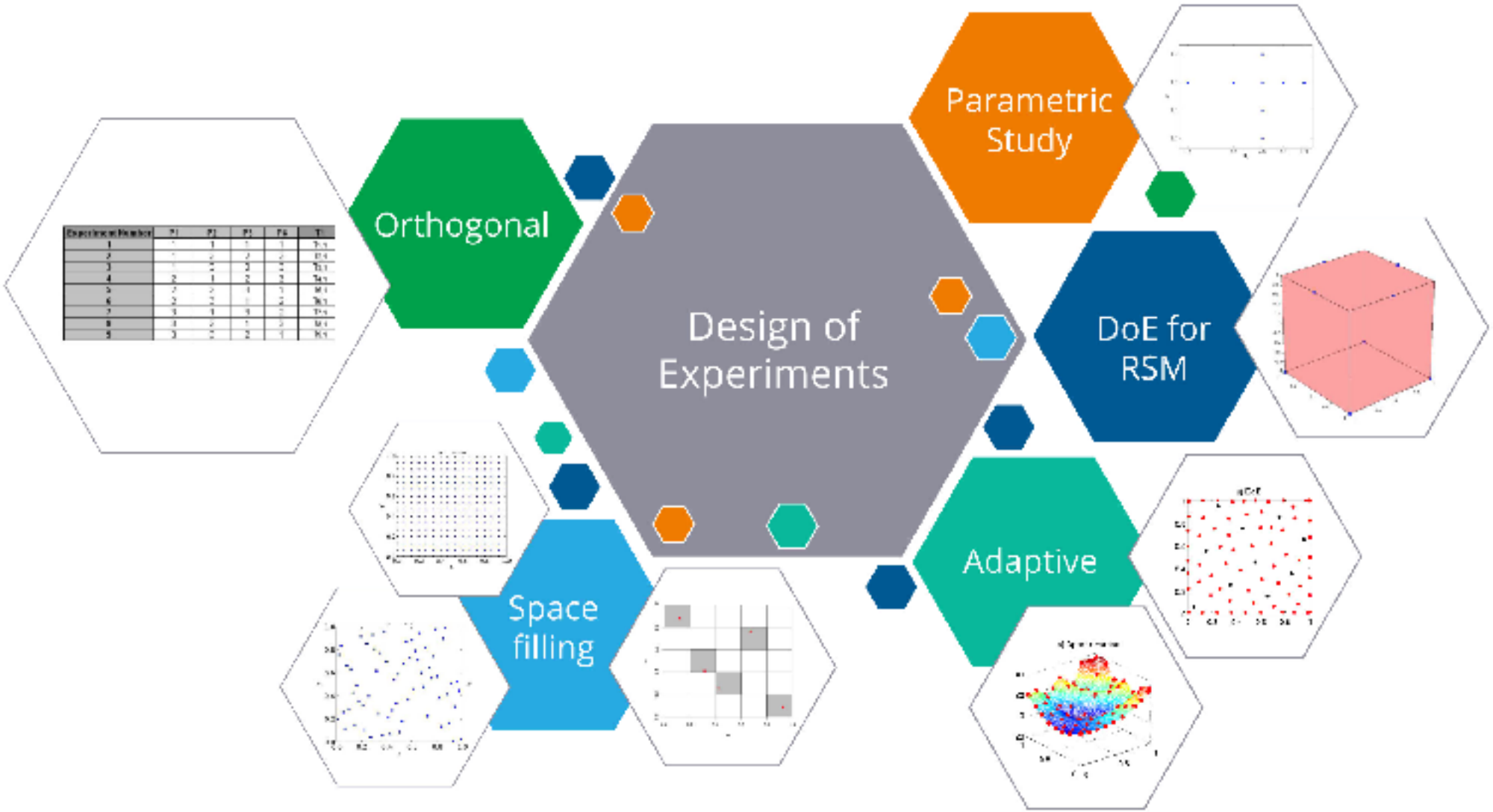
DoE challenges:

- DoEs behaviors can be very different in **dimensionality**, **size**, **smoothness**, **noisiness** etc.
- Often there are also special requirements to DoE like **anisotropy** and **factorization**
- Available number of **calculations** are often **limited**



DATADVANCE

DoE techniques in pSeven



DATADVANCE

DoE techniques in pSeven

pSeven provides a wide range of techniques to construct DoE:

Batch Space-filling DoE:

- Random sampling with given distribution
- Latin hypercube sampling (LHS)
- Optimized LHS (OLHS)
- Full Factorial

Sequential Space-filling DoE:

- Halton sequence
- Sobol sequence
- Faure sequence

Model-based Adaptive DoE:

- Maximum Variance criterion
- IMSEGain-Maximum Variance criterion
- Probability of improvement (used in SBO)

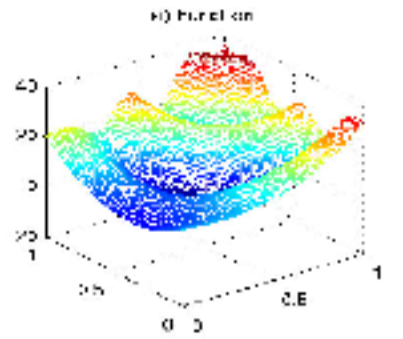
Uniformity-based Adaptive DoE:

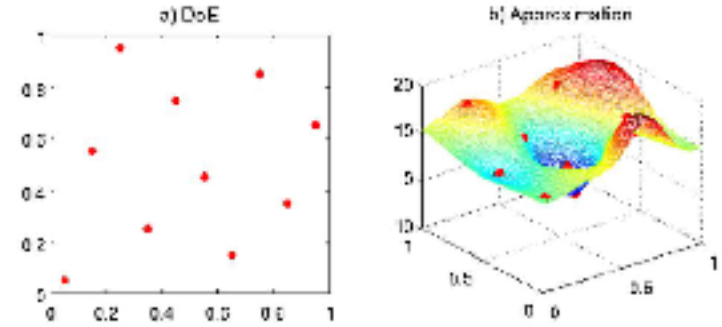
- Parametric Study
- Orthogonal Array
- Fractional Factorial designs

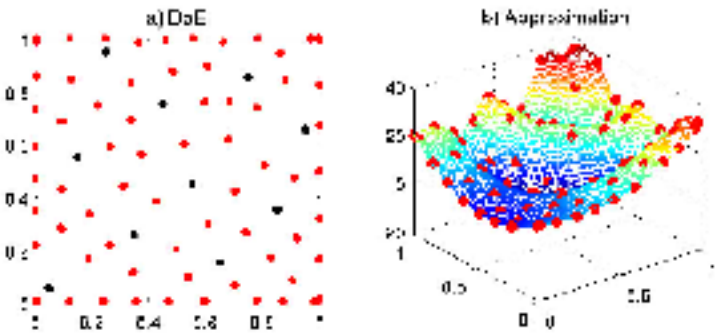
Optimal Designs for RSM:

- D-optimality, I-optimality
- Box-Behnken technique

15
Techniques







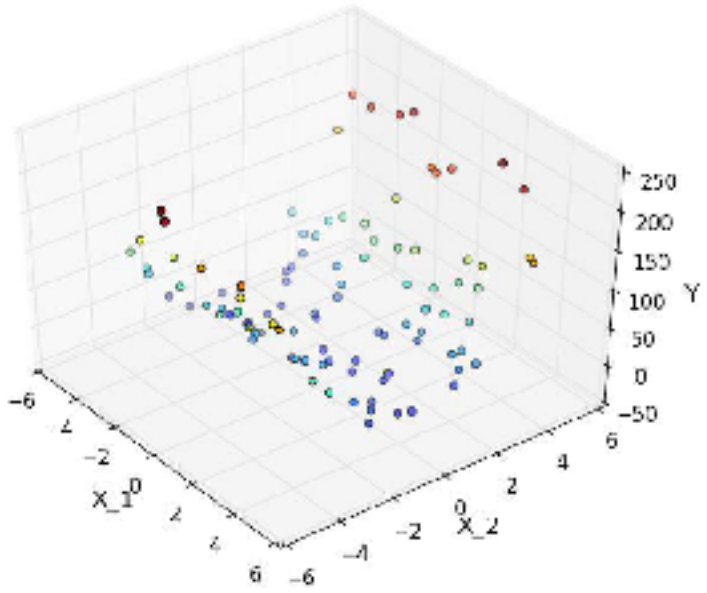
DATADVANCE

Sensitivity and Dependency Analysis (SDA)

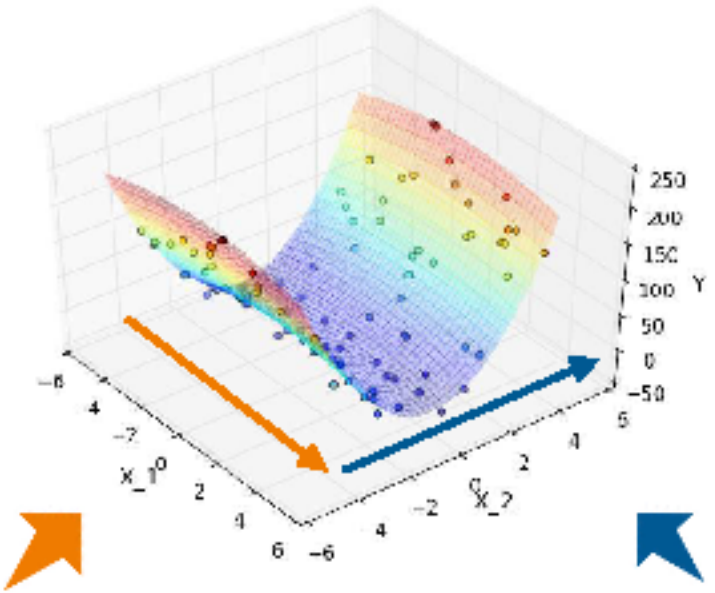
How data variables are **correlated**?

Which input variables are **more/less important** for the response function?

Which input variables can be **discarded/ignored**?



Data set with input-output values



Not important input variable

Important input variable

DATADVANCE

SDA in pSeven

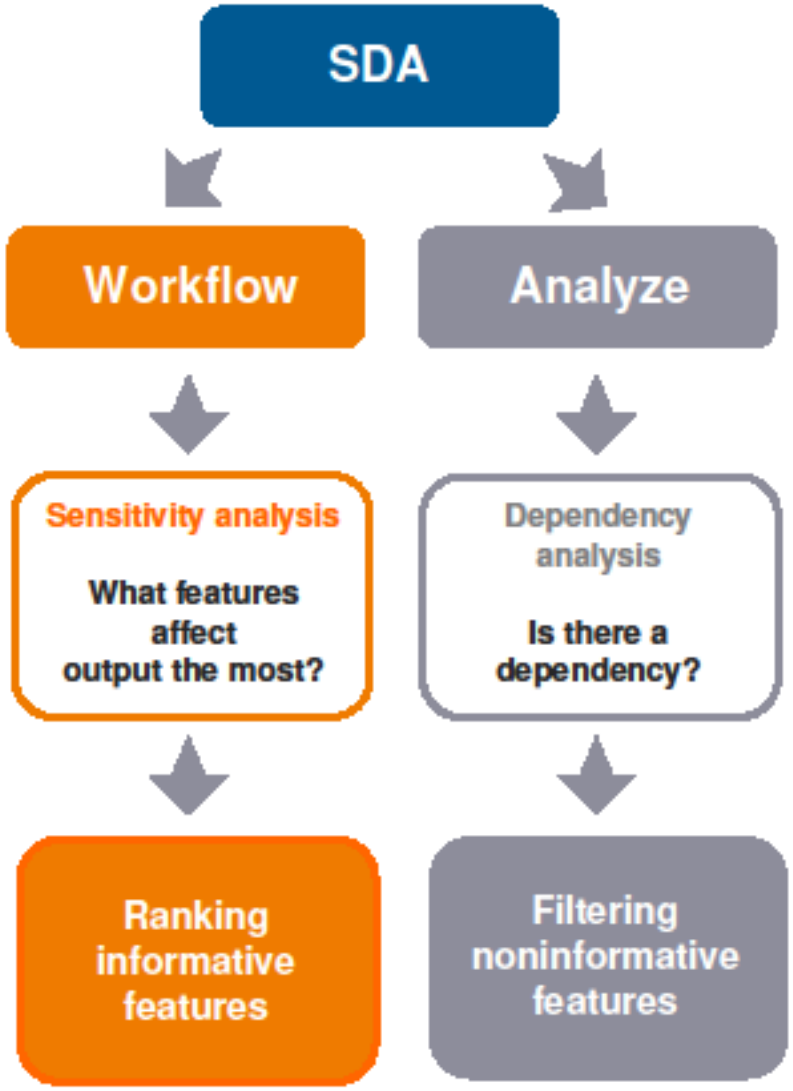
pSeven provides rich set of state-of-the-art techniques for sensitivity and dependency analysis.

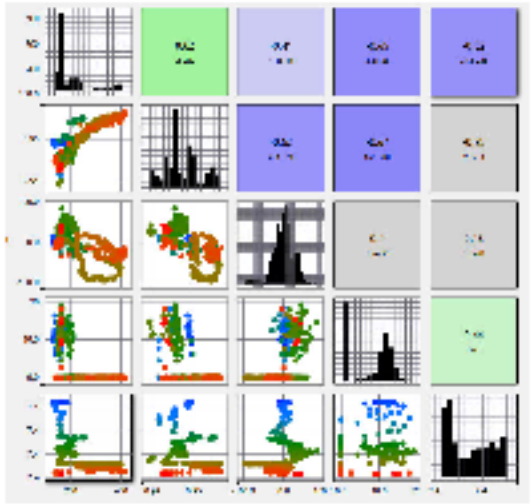
Sensitivity analysis:

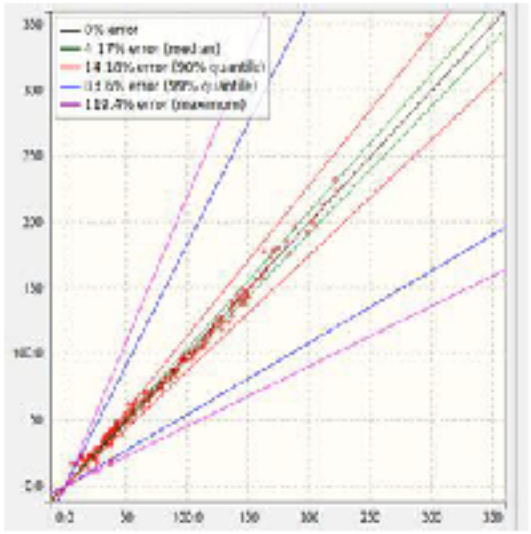
- Blackbox-Based (Elementary Effects, Fourier Amplitude Sensitivity Testing)
- Sample-based (Mutual Information, Ridge Regression, Surrogate Model Based FAST)

Dependency analysis:

- Linear correlation (Pearson, Partial Pearson, Robust Pearson)
- Rank correlation (Spearman, Kendall)
- Nonlinear correlation (Distance, Partial Distance, Mutual Information)







DATADVANCE

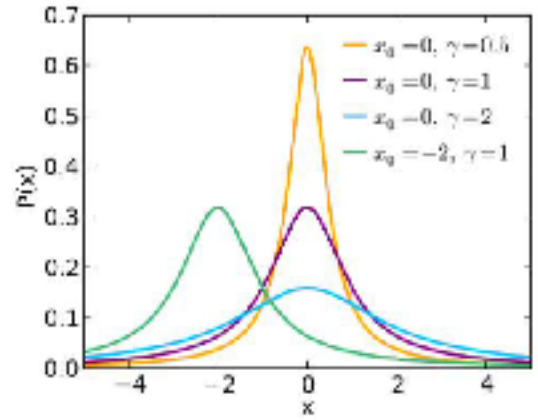
Uncertainty management

In reality the values of model parameters (geometry, material properties, load magnitudes etc.) always contain some uncertainty.

This uncertainty can be both caused by **technological limitations on the accuracy** and by the **natural variability** of a parameter.

In some cases, common analysis tools and methods may not be sufficient for an engineer who wants to:

- **Validate** product **robustness** under various conditions
- **Study product behavior** and possible ways to improve it



DATADVANCE

Uncertainty management in pSeven

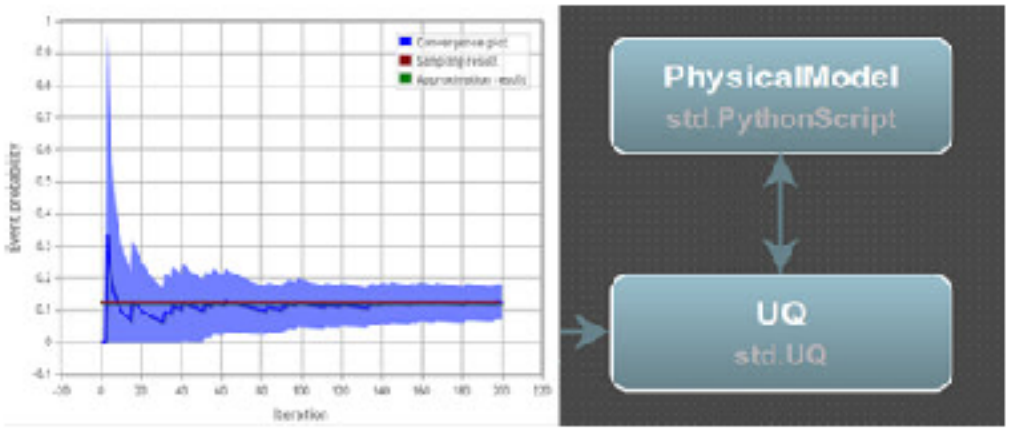
pSeven provides an easy to use tool to evaluate the influence of uncertain parameters of a product on the technical and operational characteristics.

Uncertainty quantification:

- Auto-selection of distribution type for parameters sample
- Create parametric and non-parametric probabilistic models
- Dependencies of input parameters

Reliability analysis:

- Failure probability, reliability index
- Variety of algoritms (FORM, Monte Carlo, LHS, Directional sampling)



DATADVANCE

Use your favorite tools for data analysis

You may add complementary modules and models to pSeven through Python integration.

Smoothly use whole universe of python data analysis modules

PyBrain · scikit-learn · pandas · Bokeh

- Classification
- Clusterization
- Time series analysis
- Data representations
- Manifold learning

python

pSeven

Create, store and view custom interactive reports and visualizations inside pSeven

DATADVANCE

Agenda

- Design Space Exploration with pSeven
  - Data & Model Analysis
  - **Predictive Modeling**
  - Design Optimization
- pSeven Platform
- Summary

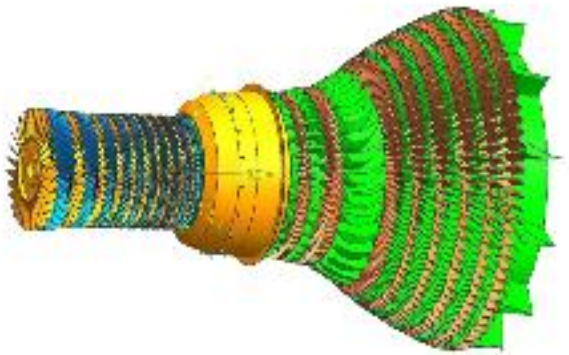
DATADVANCE

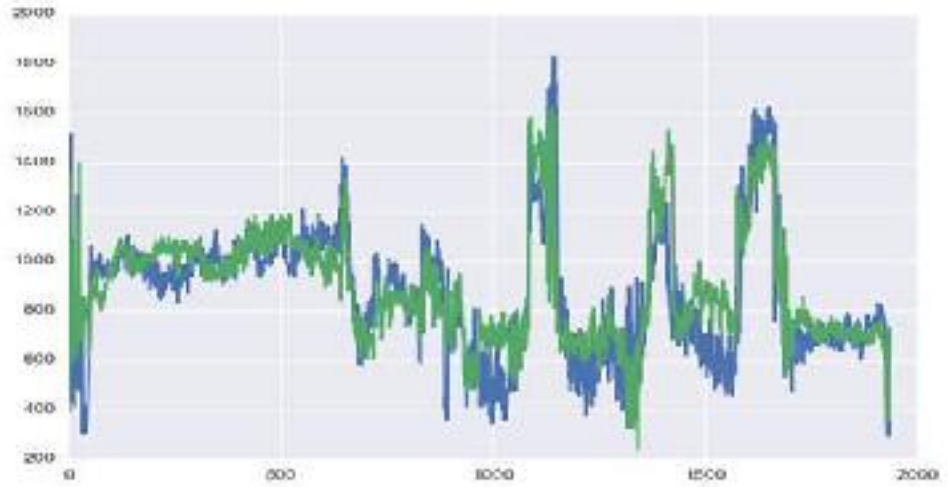
Predictive modeling

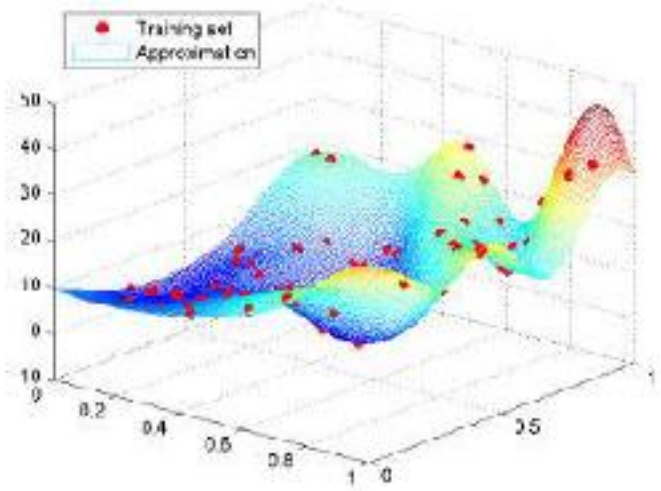
How to **predict** product behavior in various conditions?

How to process **data** from **experiments** and **simulations** together?

How to use huge data samples and simulations **faster**?





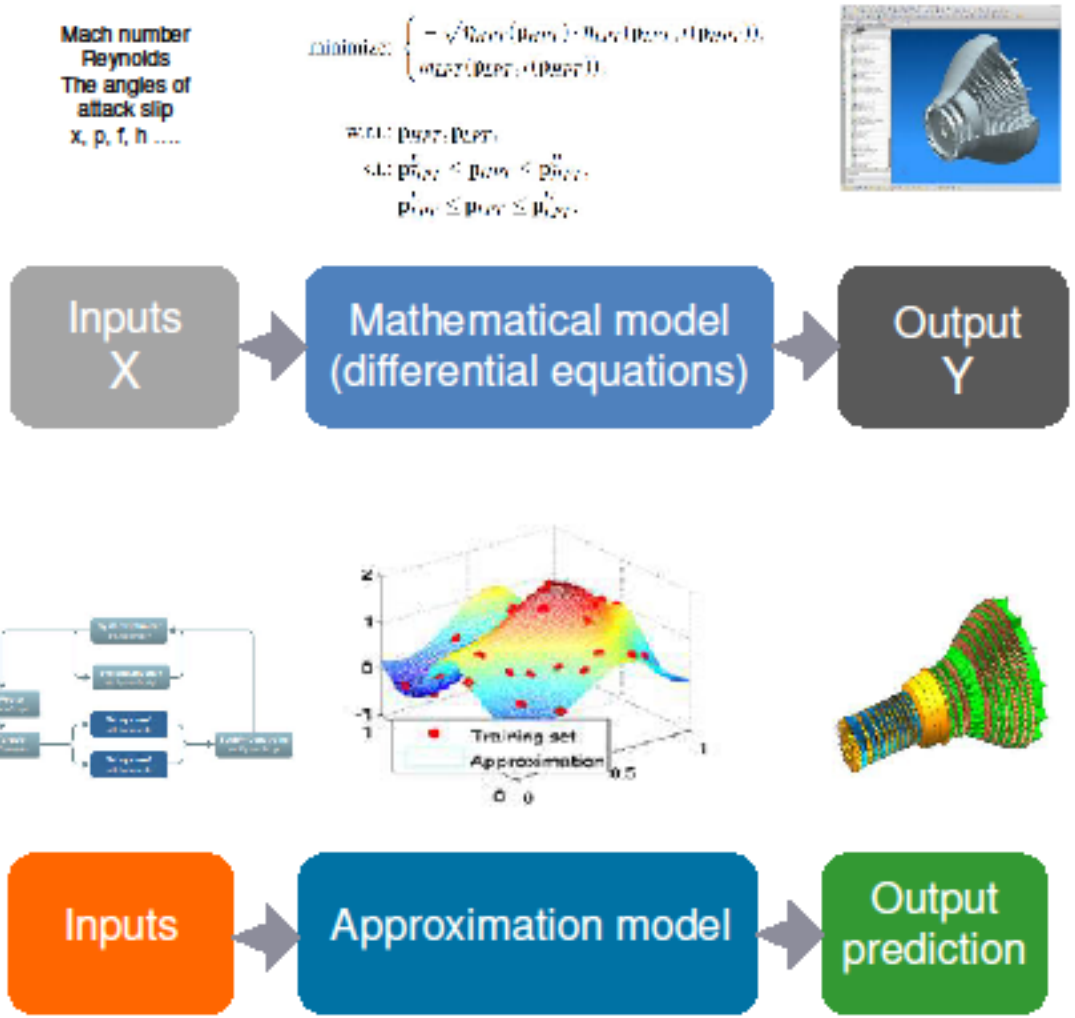


Predictive modeling is an approximation of available data based on creating **surrogate models**.

Surrogate models

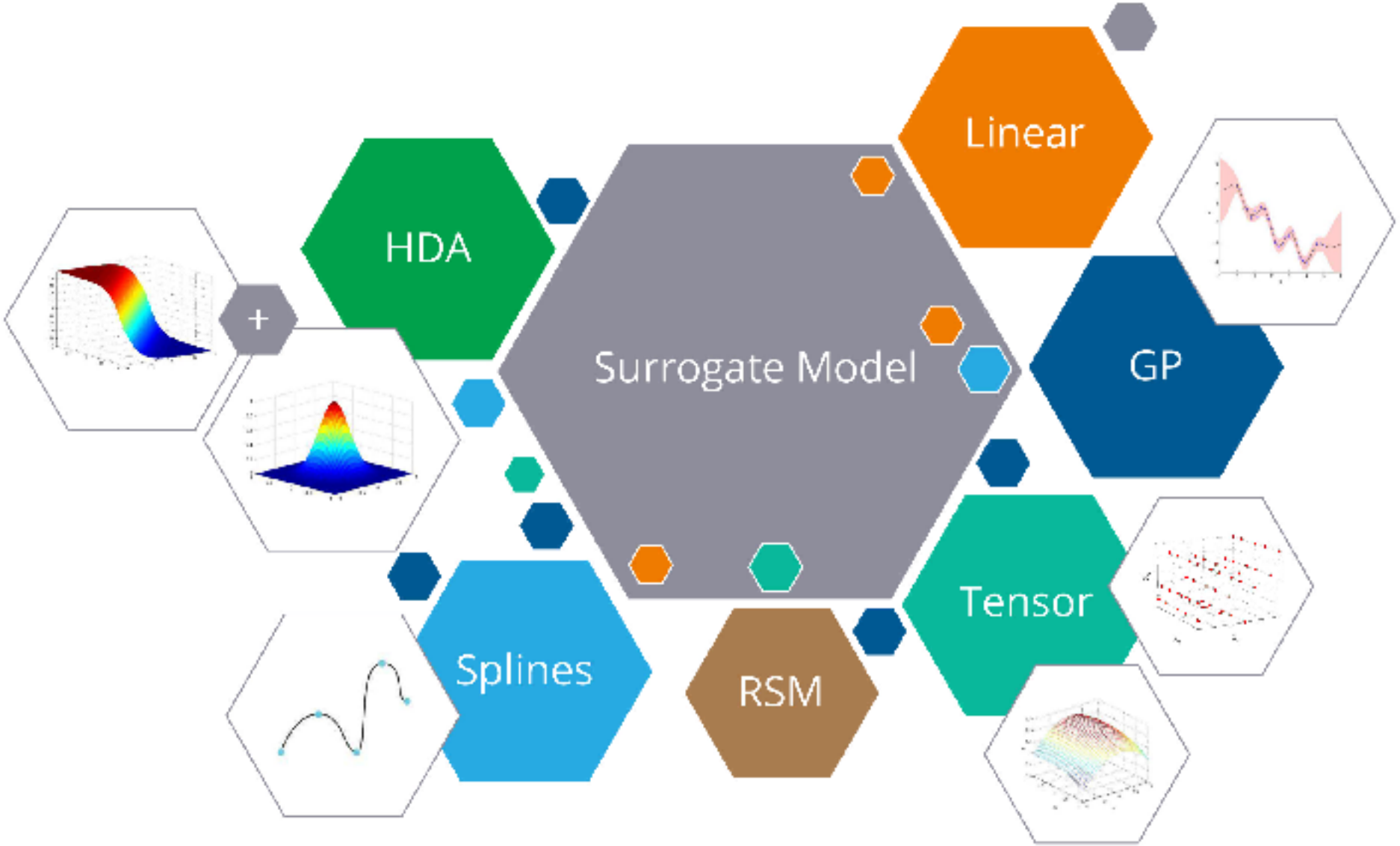
Surrogate models are the **substitution** ("blackbox") of existing **data** and **simulation** models.

- **Predict** response function values for new designs
- **Accelerate** computation of complex simulation models by many orders of magnitude
- Use fast surrogate models in **parametric and optimization** studies
- **Capture** essential knowledge from vast amounts of data
- Easily and safely **exchange** surrogate models between partners preserving IP rights



DATADVANCE

Surrogate modeling techniques in pSeven



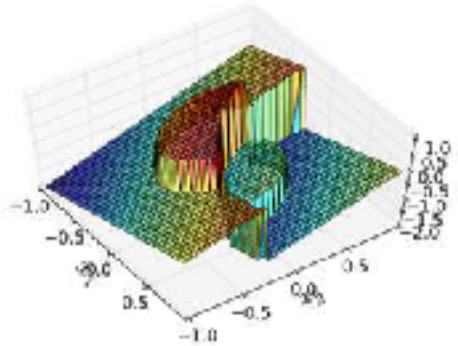
DATADVANCE

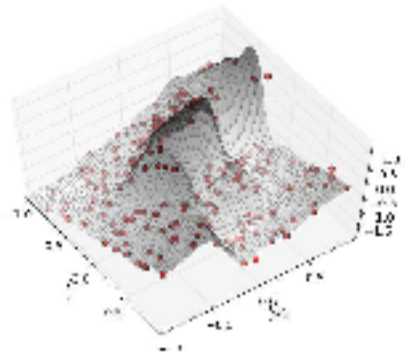
Surrogate modeling techniques in pSeven

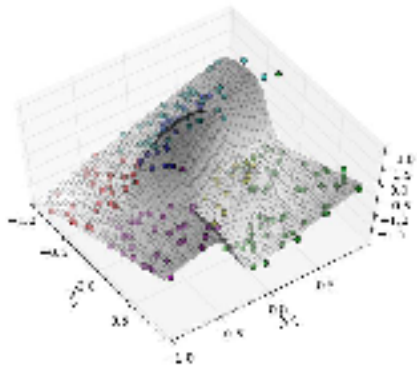
pSeven provides industry proven techniques for surrogate modeling:

- Piecewise Linear Approximation (PLA)
- 1D Splines with tension (SPLT)
- Response Surface Model (RSM)
- Gaussian Processes (GP)
- Gradient Boosted Regression Trees (GBRT)
- High Dimensional Approximation (HDA)
- Tensor Approximation and Incomplete Tensor Approximation (TA, iTA)
- …and other in-house techniques

13
Techniques



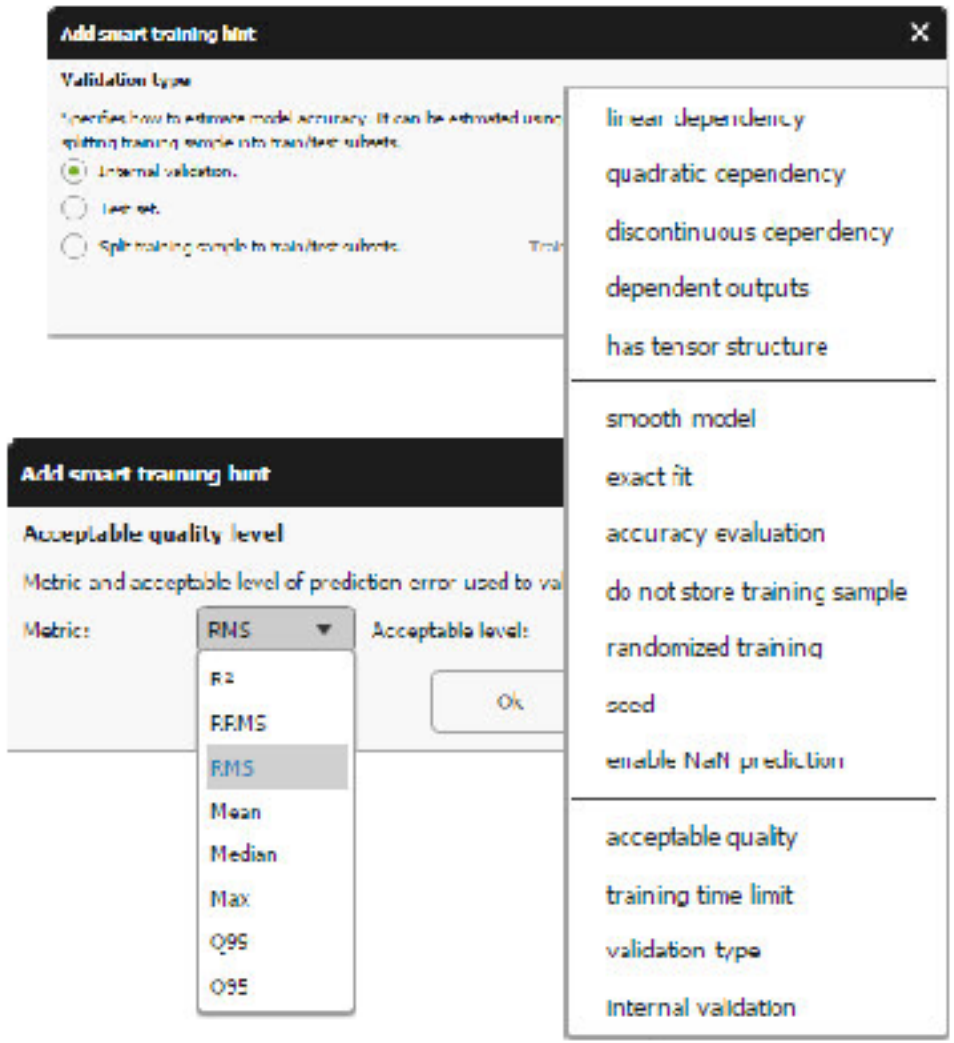




Surrogate modeling configuration

pSeven doesn't require knowledge of a specific surrogate modeling technique and its settings from user:

- **Set of options and hints** helps user to describe problem and desired solution from his point of view, not from the algorithmic point of view:
  - Provide hints about the data: linear, quadratic, discontinuous etc.
  - Specify desired model properties: smooth, exact fit, accuracy evaluation, NaN prediction etc.
  - Specify time constraints and required quality: acceptable quality, training time limit, validation type, internal validation
- **SmartSelection™** automatically selects the most efficient technique for a given problem and data, so users can concentrate on the engineering problem itself.



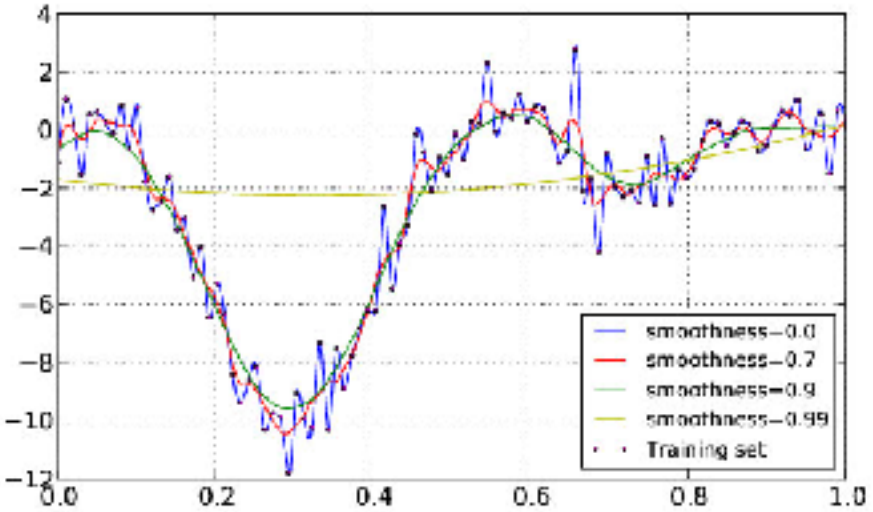
DATADVANCE

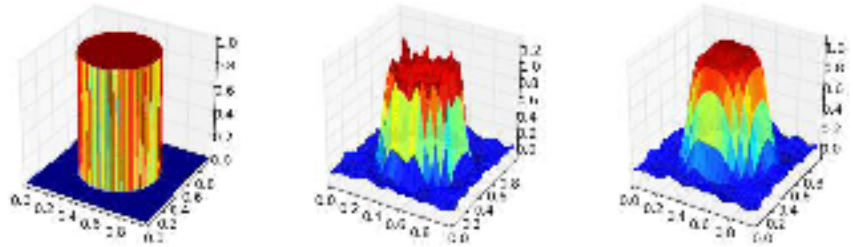
Surrogate modeling key features

Data can be very different in dimensionality, size and noisiness. Modeling may also require additional pre-and post-processing to collect data and assess the results.

pSeven is a **«Swiss army knife»** for creating surrogate models:

- **Data fusion** – construction of models from multi-fidelity data sources
- **Accuracy** and **error** assessment of constructed models
- **Exact fit** and **smoothing**
- Full control of the model construction **time**
- Handling of **missing data** and **discontinuities**
- **Updating** existing models with new data
- **Combining** of models
- **Export** to C, Octave and FMI





DATADVANCE

Agenda

- Design Space Exploration with pSeven
  - Data & Model Analysis
  - Predictive Modeling
  - **Design Optimization**
- pSeven Platform
- Summary

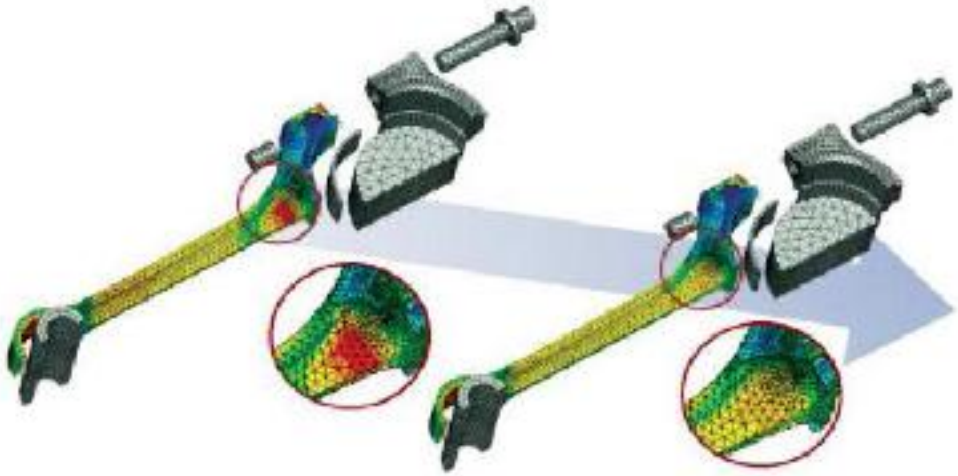
DATADVANCE

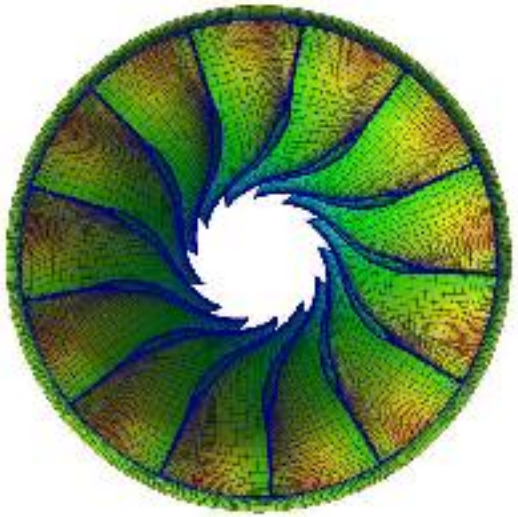
Design optimization

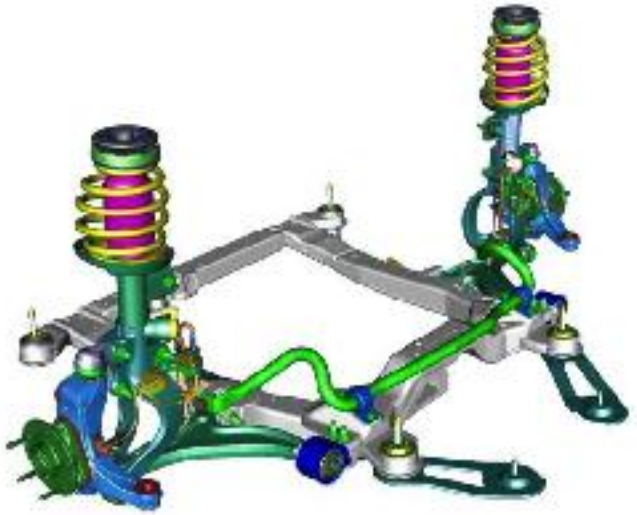
Which product design parameters are the **best**?

How to **improve** product characteristics?

How to decrease effect of parameters **variability** on overall product **behavior**?



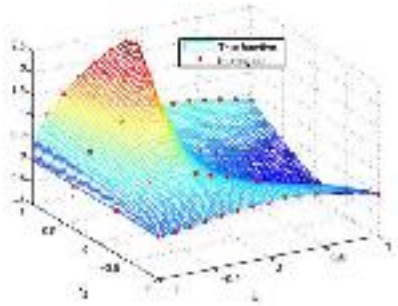


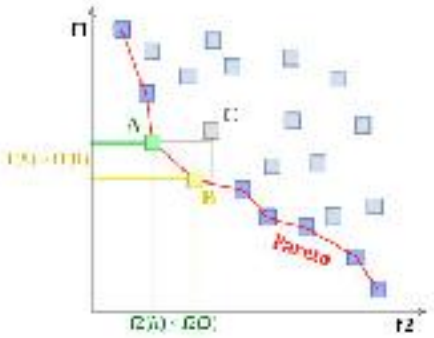


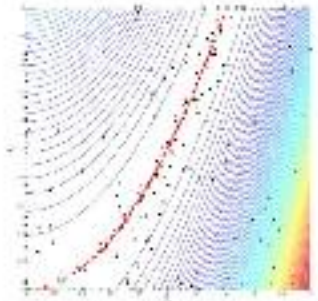
Design optimization helps engineers to answer these questions.

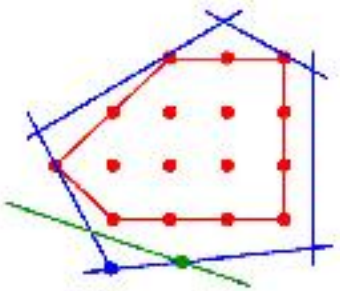
Design optimization made easy

- Problem statement: single- or multi-objective, multidisciplinary, robust- or reliability-based
- Large dimensionality
- Continuous and/or discrete input parameters
- Nonlinear, multimodal or noisy objective functions and constraints
- Presence of implicit constraints (domains of undefined behavior)
- Presence of uncertainties
- Long calculation time

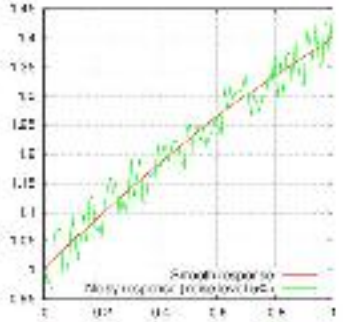








pSeven provides easy and effective solution for most of industry optimization problems!



DATADVANCE

Solving optimization problems in pSeven

Optimization problem statement → **Hints setup** → **SmartSelection™ chooses algorithm** → **Solution**

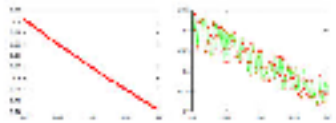
- Single-objective
- Multi-objective
- Multidisciplinary
- Robust-based
- Reliability-based

$$\min_{\vec{x}} \vec{F}(\vec{x})$$

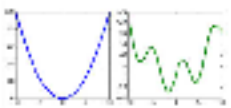
$$\vec{x}_0 \le \vec{x} \le \vec{x}_1$$

$$\vec{C}_0 \le \vec{C}(\vec{x}) \le \vec{C}_1$$

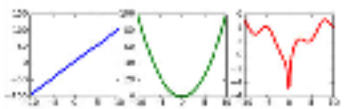
- Noisy



- Multi-extremal



- Linearity type



- Computational time

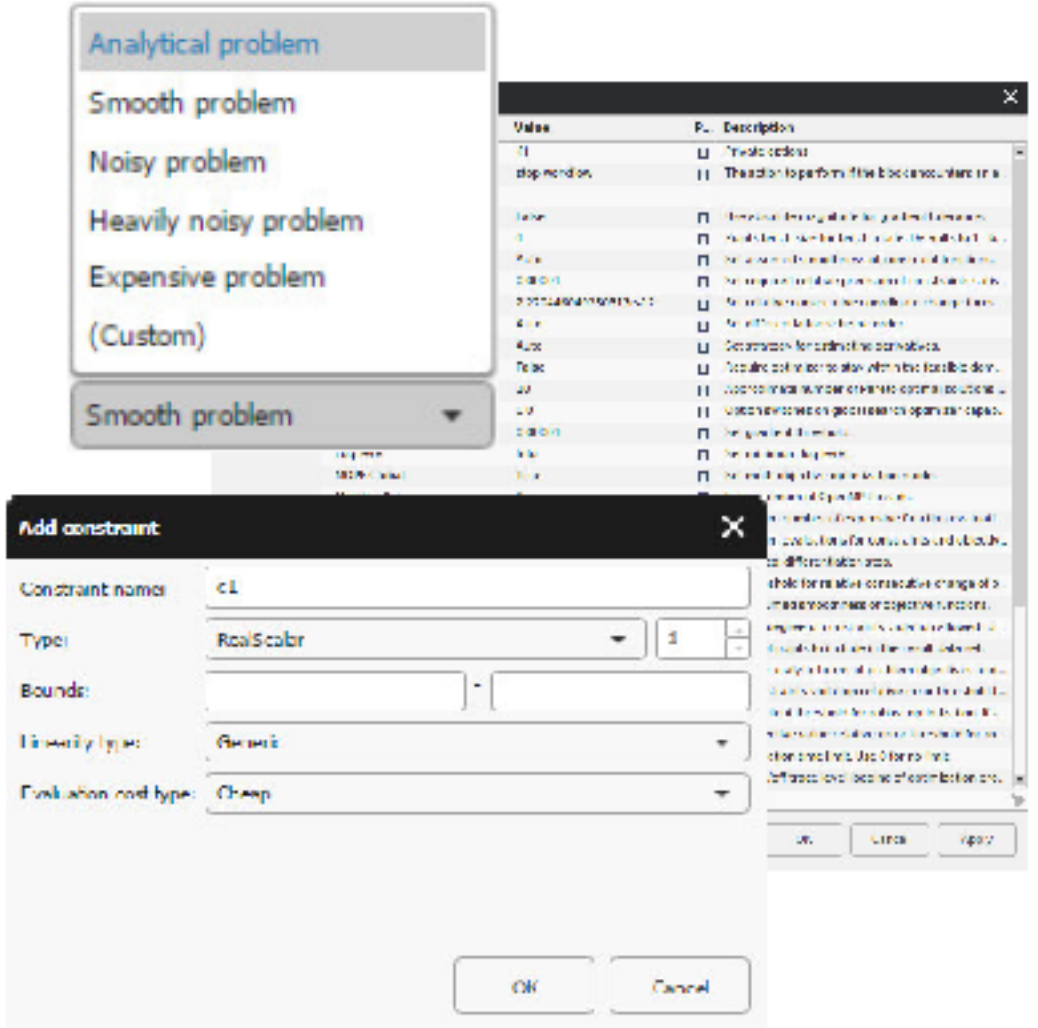
- **pSeven includes full set of optimization algorithms:**
  - QN – Single or Multi-Objective Quasi-Newton
  - QP – Quadratic Programming
  - SQP – Sequential QP with Filter
  - SQCQP – Quadratically Constrained SQP
  - RDO - Robust Optimization
  - SBO – Surrogate-Based Optimization
  - …and other local, governing and global algorithms

16 Algorithms

Optimization configuration

pSeven doesn't require knowledge of a specific optimization algorithm and its settings from user:

- **Set of options and hints** helps user to describe problem and desired solution from his point of view, not from the algorithmic point of view:
  - Hints for variables and responses: expensive/cheap, linear/quadratic/generic.
  - Options presets: analytical problem, smooth problem, noisy problem, heavily noisy problem, expensive problem.
  - High-level options: optimization stop criteria, globalization intensity, number of Pareto points.
- **SmartSelection™** technology chooses the optimal algorithm automatically based on hints, options and optimization behavior.



DATADVANCE

Single- and multi-objective optimization

Single-objective algorithms:

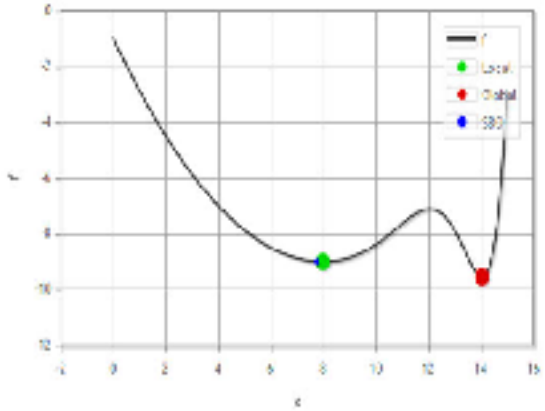
- Most of them originally implemented and specifically tuned for engineering problems

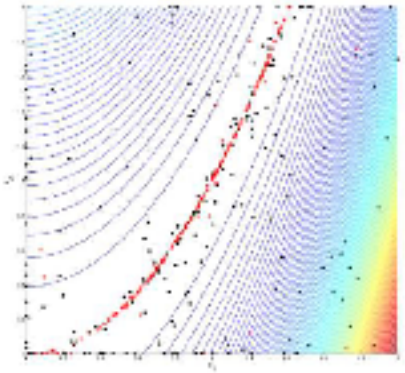
Multi-objective algorithms:

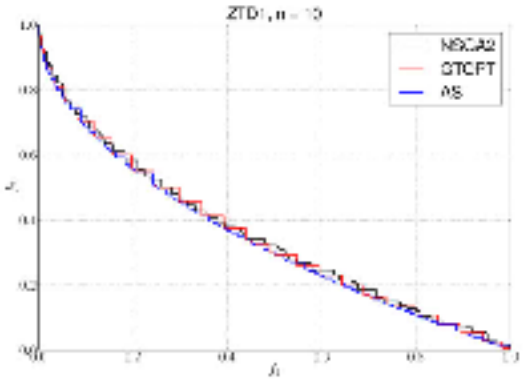
- Avoids evaluations far from Pareto frontier
- Beat genetic algorithms and scalarization techniques on most of the problems

Algorithms features:

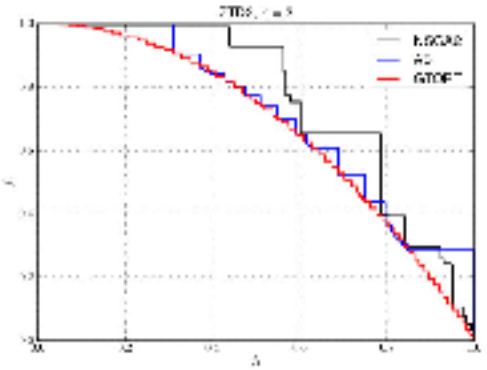
- Run evaluations in parallel
- Unique technology for handling problems with noise
- NaN support: functions have incomputable areas







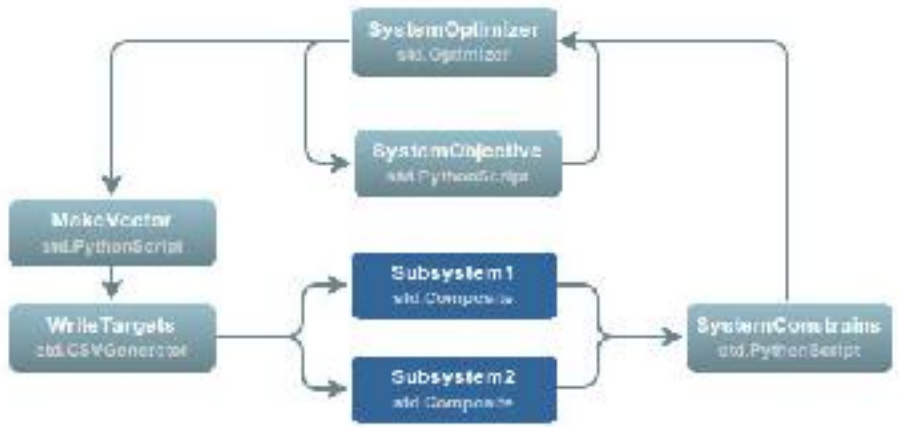
NSGA-II – 2368 iterations
Adaptive Scalarization – 3765 iterations
GTOpt – 488 iterations

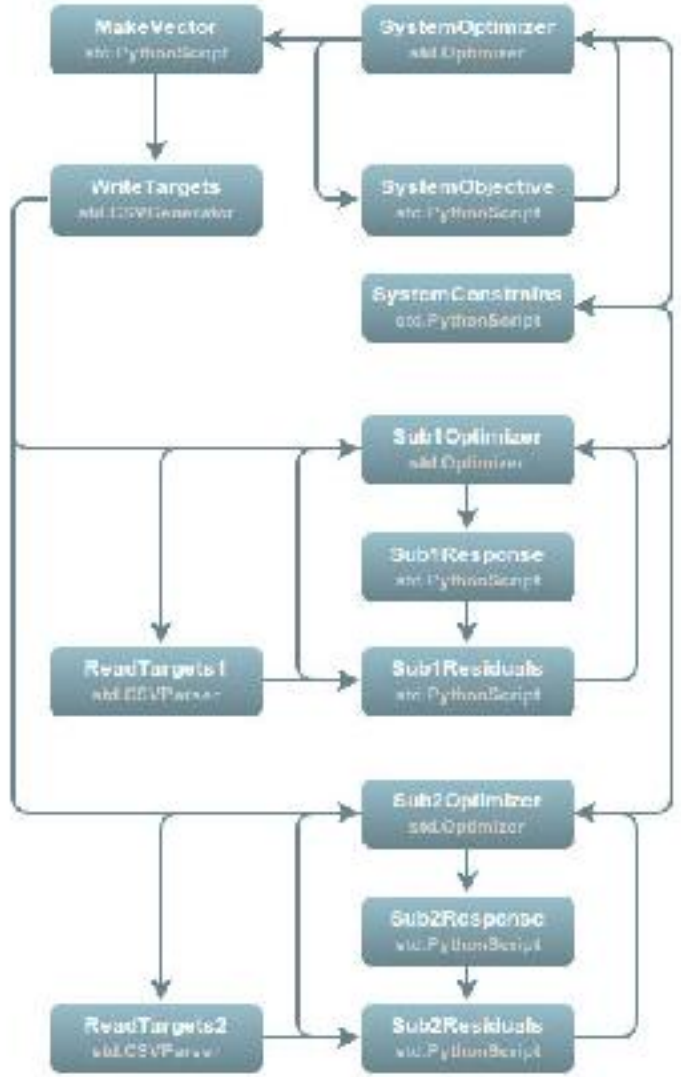


NSGA-II, Adaptive Scalarization,
GTOpt – 280 iterations each

Multidisciplinary Design Optimization (MDO)

- pSeven allows you to **integrate a wide range of solvers** and create arbitrarily complex and nested workflows.
- It makes possible to apply different **MDO strategies** to your problem, including multi-level ones:
  - Collaborative Optimization (CO)
  - Analytical target cascading (ATC)
  - Bi-level Integrated System Synthesis (BLISS)
  - Concurrent Subspace Optimization (CSSO)



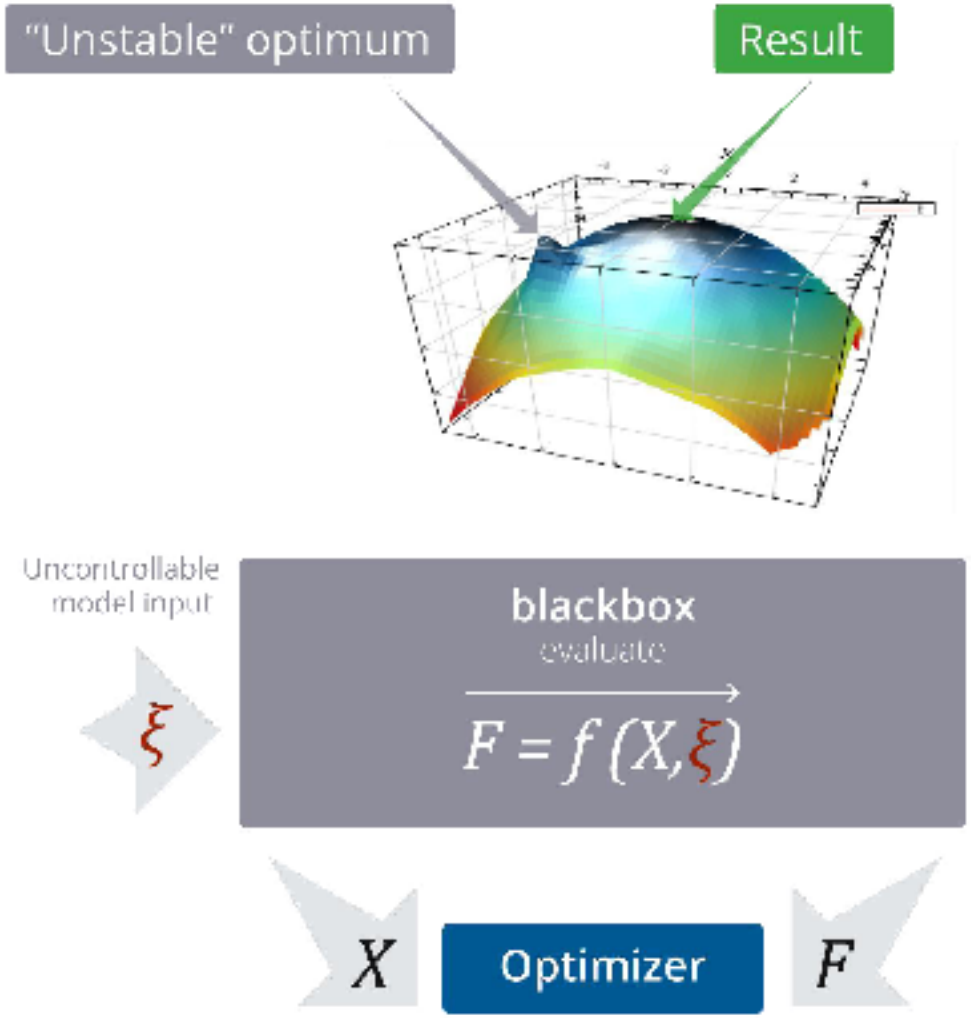


DATADVANCE

Robust- and Reliability-Based Design Optimization

- **Robust Design Optimization (RDO)** – uses a measure of the robustness of the system or component as optimization constraint or objective in order to meet the best robust performance possible.
- **Reliability Based Design Optimization (RBDO)** - uses the mean values of the random system parameters as design variables, and optimizes the cost or objective function subject to prescribed probabilistic constraints.

pSeven supports virtually all possible robust formulations, including probabilistic and quantile type constraints.

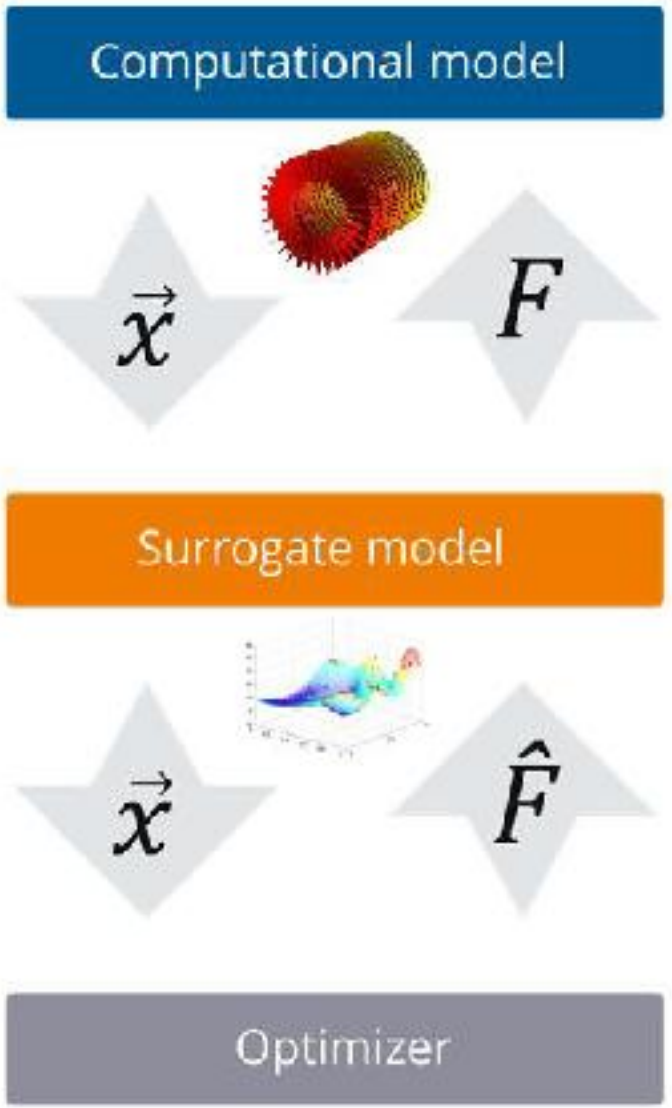


Surrogate-Based Optimization (SBO)

Surrogate-Based Optimization (SBO) is a class of optimization algorithms involving constructing and optimizing surrogate models as auxiliary steps.

Features:

- Intelligently spends evaluations budget
- Allows to effectively solve large scale problems (up to 100 design variables) based on unique implementation of multi-resolution GP.
- In-house developed DoE strategy, which respects as much feasibility domain of the problem as possible
- Single slider regulates the complexity of applied global methods



DATADVANCE

Agenda

- Design Space Exploration with pSeven
  - Data & Model Analysis
  - Predictive Modeling
  - Design Optimization
- **pSeven Platform**
- Summary

DATADVANCE

Visual process integration

Capture your design process with pSeven

- Integrate simulation
- Perform multidisciplinary design optimization
- Use predictive modeling
- Automate trade-off studies

Using

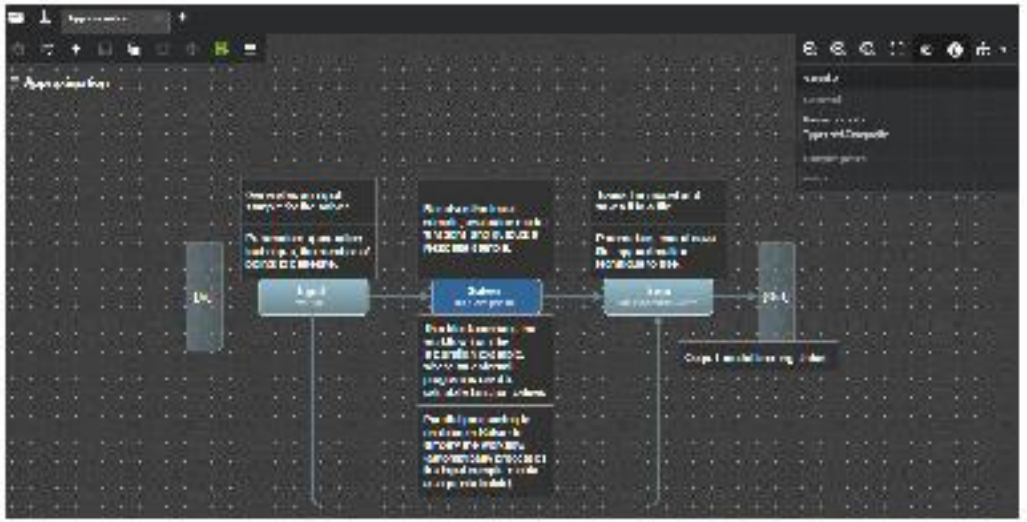
- Creation of simple visual workflows
- Automatic file management
- User-friendly graphical interface
- Full support of Python scripting

Process modeling

Design process in pSeven is represented as a sequence of computations with defined execution order or **Workflow.**

Workflow system provides:

- Intuitive definition of complex computations
- Data reuse
- Data caching
- Parallel execution
- Full history of your computations
- Possibility to construct nested loops



DATADVANCE

Building workflows with blocks

Basics:

- Workflow consists of blocks and links
- Each block represents some kind of activity

General:

- Handle your data flow with specialized blocks
- Composite blocks - Create cached regions in a workflow, export and import blocks

Logic:

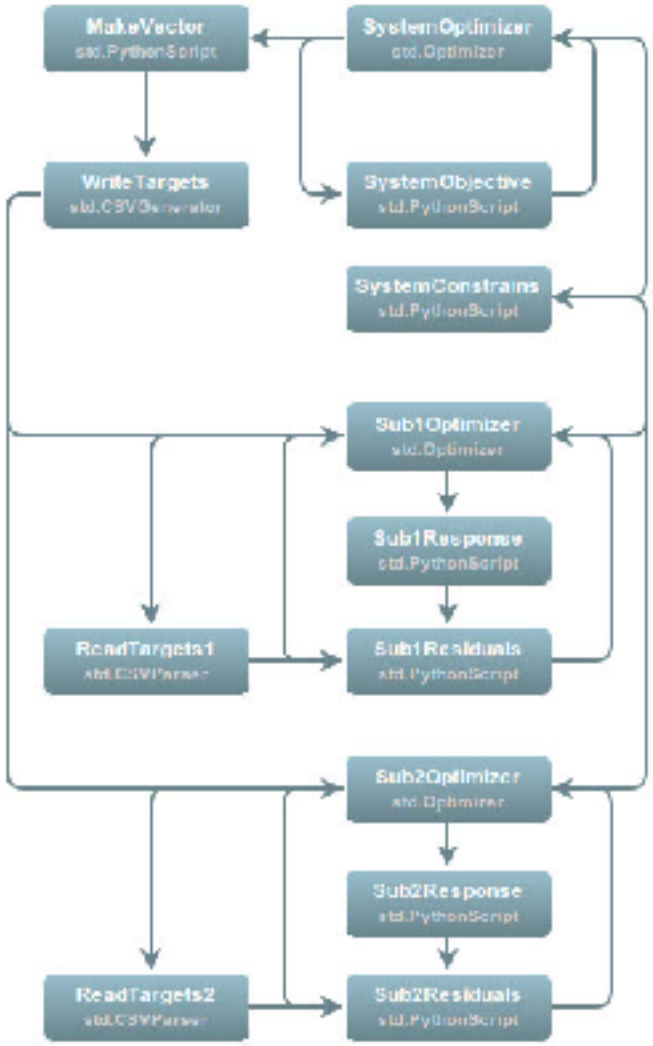
- Branching - Create links that connect output port to different input ports. Independent branches are executed in parallel, automatically increasing performance
- Looping - Add loops to workflow, including nested optimization loops which are essential for MDO

Integration:

- Direct integration with CAD & CAE systems
- Integration of analytical models & 3rd party software using Text files or Python scripts

Options:

- Set parameters to be changed in the workflow, map options to the ports

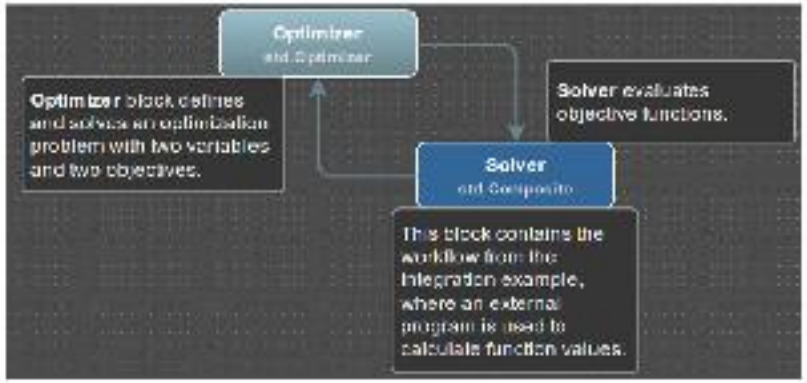


DATADVANCE

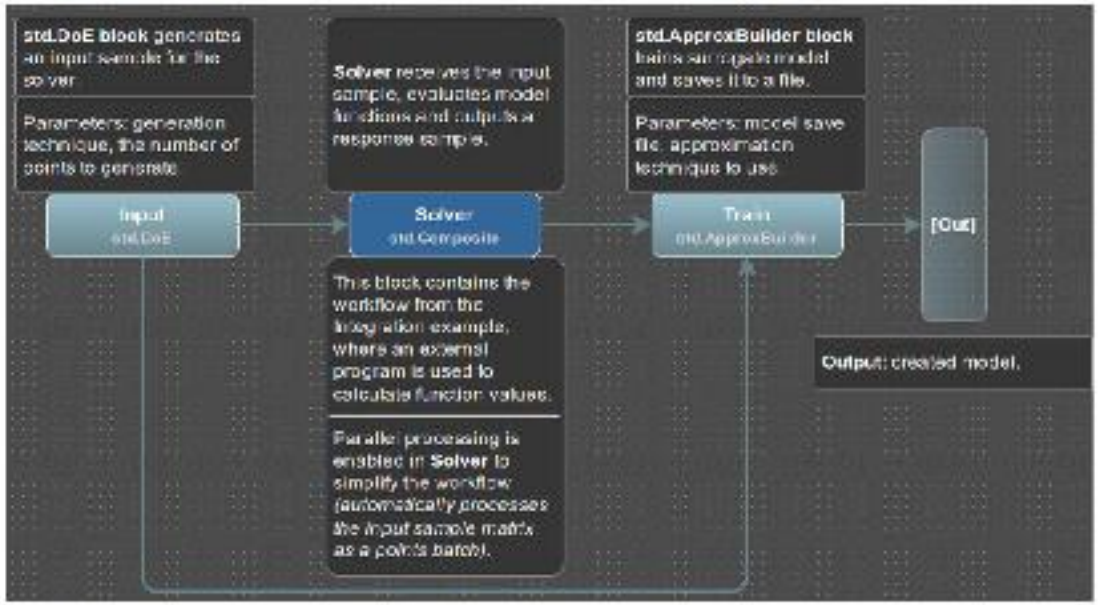
Algorithms blocks

Algorithm blocks in pSeven:

- Design of Experiments
- Optimization
- Approximation
- Data Fusion
- Important Variable Extraction
- Dimensionality reduction
- Uncertainty quantification



Optimization workflow



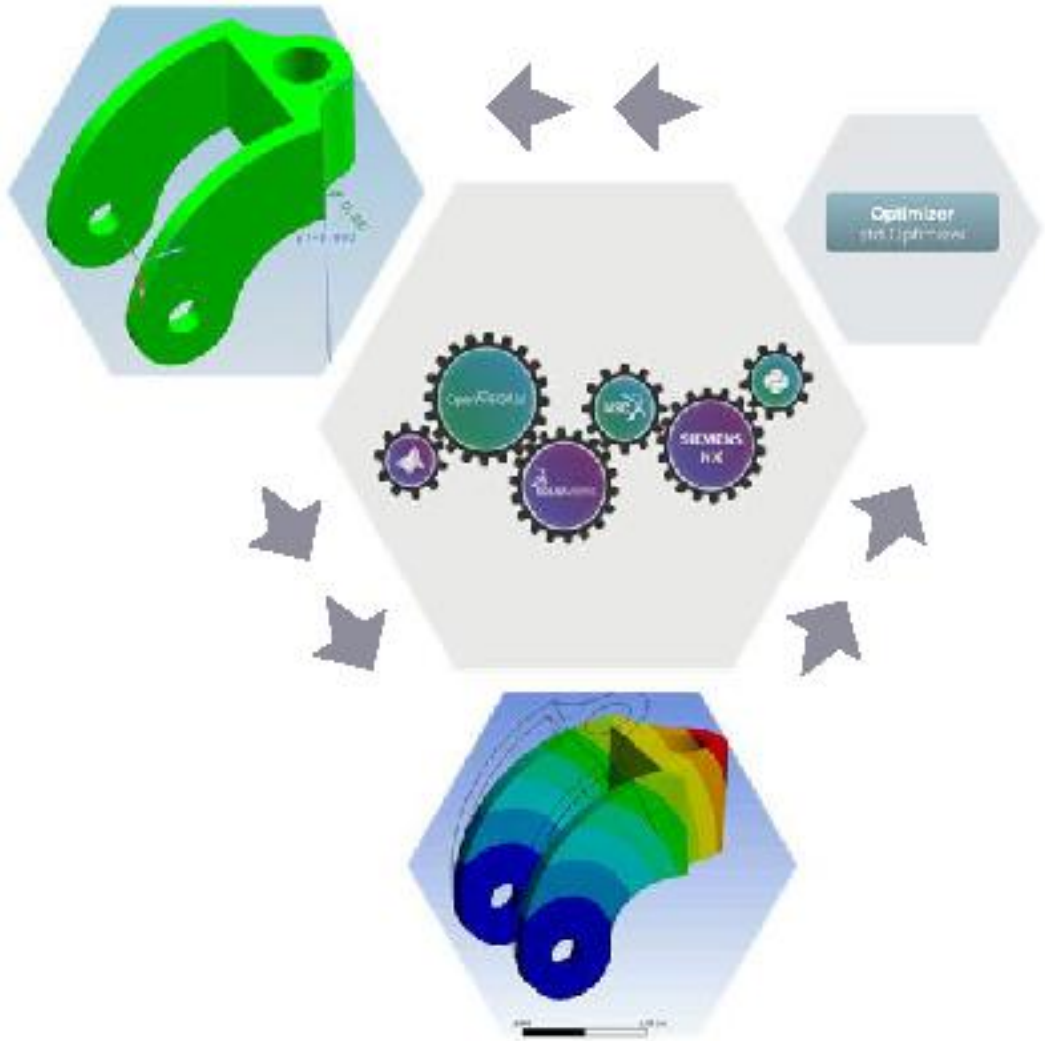
Creating DoE and constructing surrogate model with it

DATADVANCE

CAD & CAE Integration

pSeven integration blocks allow to couple CAD and CAE applications with design exploration algorithms.

SOLIDWORKS
ANSYS
CATIA
MSC Software
SIEMENS
ABAQUS
NX
Mentor Graphics FloEFD™
PTC creo™
OpenFOAM



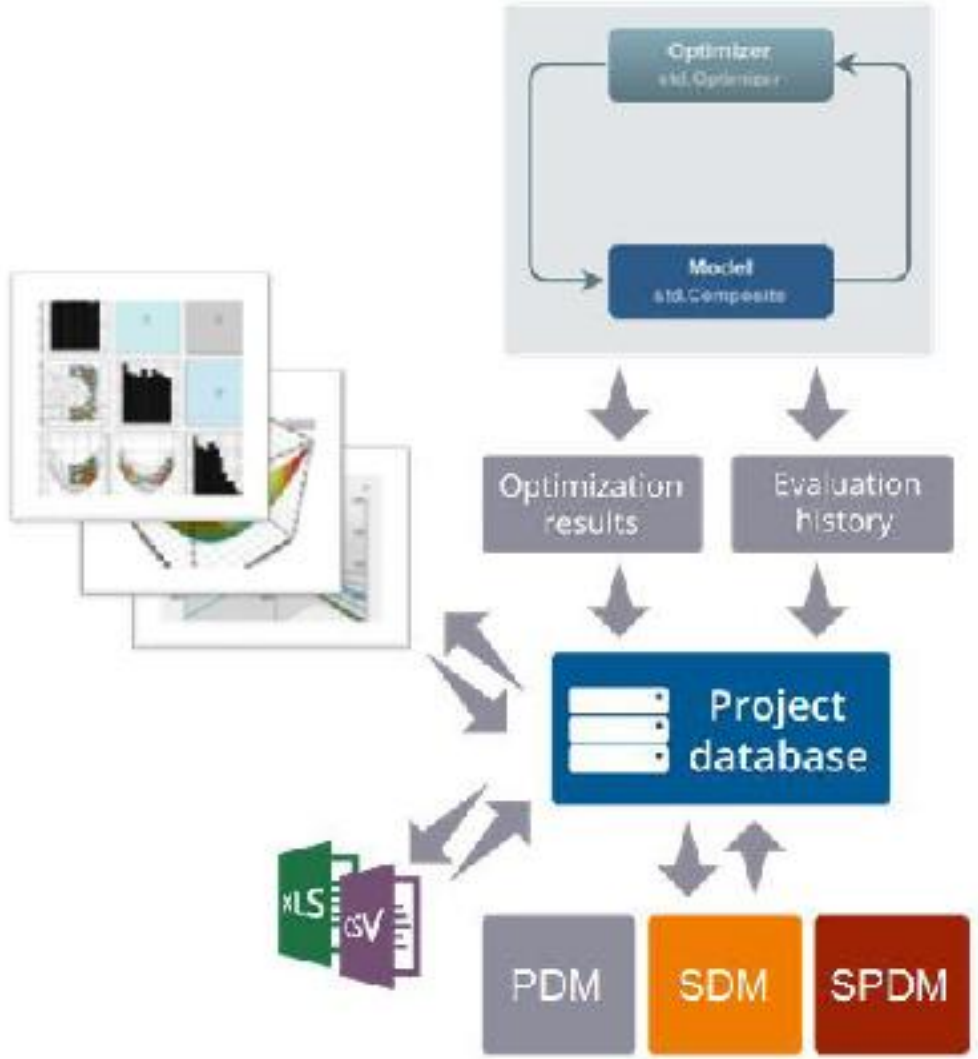
If your tool of choice is not yet integrated into pSeven, you can:

- Use generic integration blocks
- Use scripting capabilities
- Develop a custom integration block

DATADVANCE

Data is under your full control

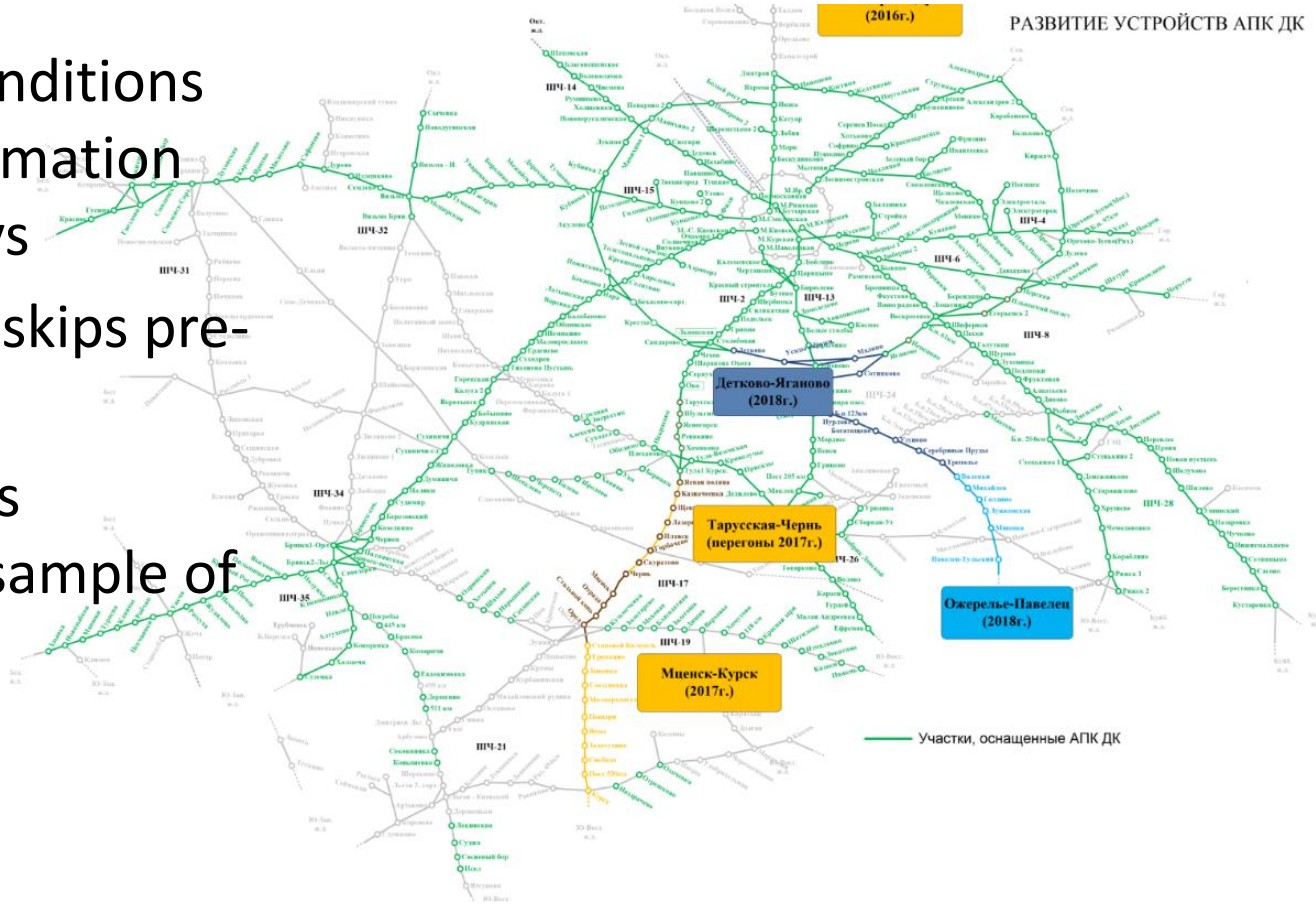
- Workflow execution history and final result are automatically stored in high-performance **project database**
- Project database provides easy-to-navigate **history** of your computations for traceability and fault tolerance
- **Import/export** supporting most popular file formats (CSV and Excel)
- Upcoming data **exchange** with PDM/SDM/SPDM systems
- Datasets in project database can be **explored** with pSeven advanced analysis capabilities



DATADVANCE

Automatic incident ranging system in Content management infrastructure Center (Moscow Railways)

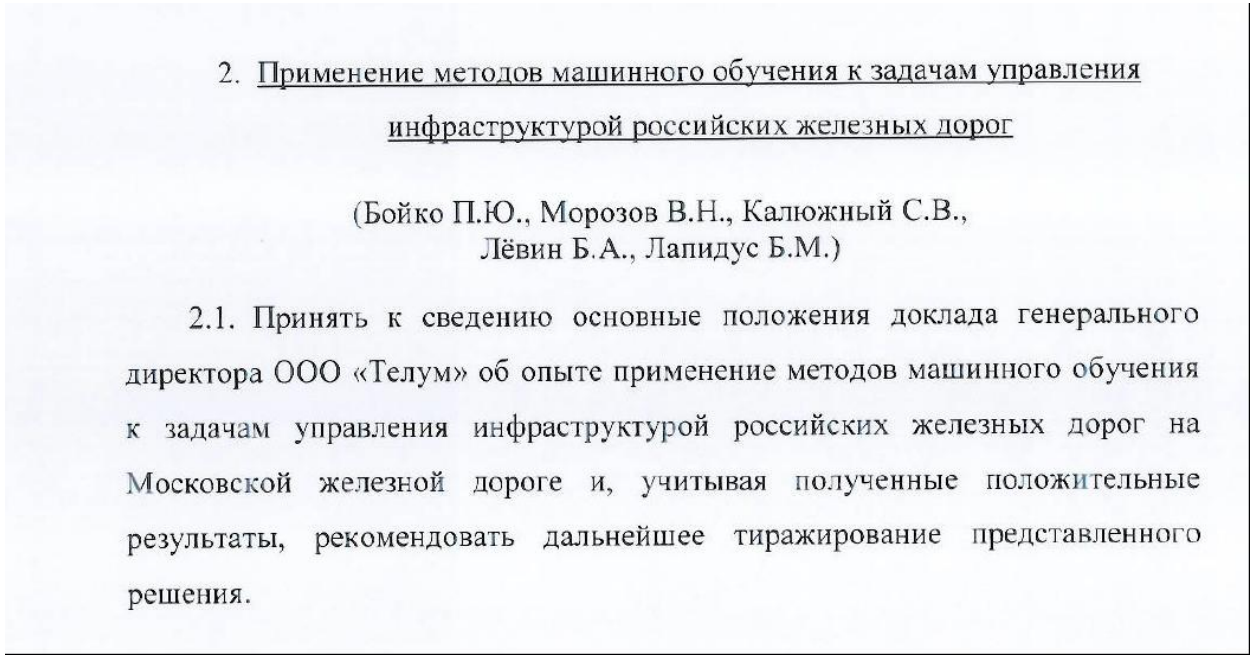
- Context: the Technical infrastructure conditions monitoring system in the sector of automation and remote control on Moscow Railways
- Problem: a large number of false alarm skips pre-orders
- Solution: an automatic signal classifier is developed and implemented; learning sample of 100+ million signals per 5 years of use



Current status

- The system has successfully passed preliminary operation on Moscow Railways
- It proved possible to reduce the number of missed pre-failure conditions to 2 times and response rate on it to 5 times
- As a result of the preliminary operation was made the report by the Joint Scientific Council of Russian Railways

Planned activities are completed in its entirety. The results of the pilot operations are reflected in the minute of the trial operation SARI on July 8, 2016. The scope, characteristics and functioning of the system technology, technical documentation comply with the requirements. During controlled operations the weaknesses haven't been identified. The system meets the requirements and generally ready to use.



2. Применение методов машинного обучения к задачам управления инфраструктурой российских железных дорог

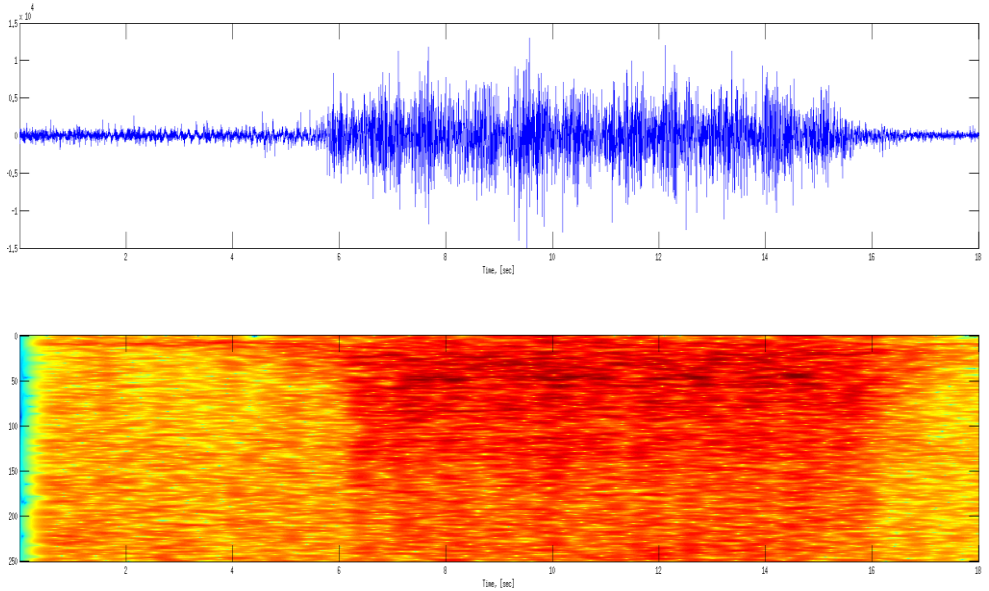
(Бойко П.Ю., Морозов В.Н., Калюжный С.В., Лёвин Б.А., Лапидус Б.М.)

2.1. Принять к сведению основные положения доклада генерального директора ООО «Телум» об опыте применение методов машинного обучения к задачам управления инфраструктурой российских железных дорог на Московской железной дороге и, учитывая полученные положительные результаты, рекомендовать дальнейшее тиражирование представленного решения.

Automatic classification of trains according to the vibro-acoustic fiber sensor

- Context: prospective application of vibro-acoustic fiber optic sensors (eg. system "Danube" production "T8") in the signaling systems
- Problem: processing the big data rate from the sensor for rolling stock positioning and determining its characteristics
- Solution: on behalf of "T8“, it was a demonstration of data processing capabilities with the "Danube" machine learning methods. The algorithms of the rolling stock maintenance during the driving were proposed
- Result: 100% classification accuracy (number of cars and electric motors) in the sample test

Example of electric classification



The classification results:

- ✓ The quantity of cars : 10
- ✓ Cars with electric motors are №№ 2, 4, 5, 7, 9

Workflow-As-a-Ready-Tool

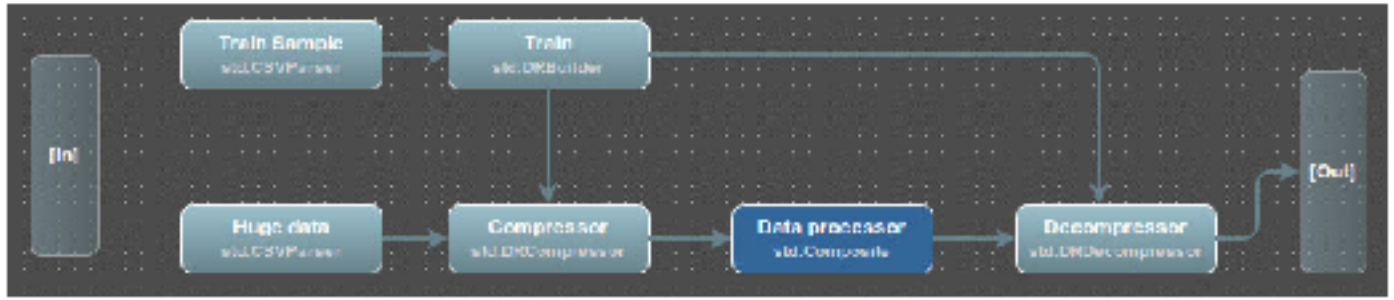
For simulation and data analysis Experts:

- Create design workflows
- Create multidisciplinary simulation models
- Specify requirements for computational resources
- Adapt and customize models/workflows

For Non-Experts and even Non-Engineers:

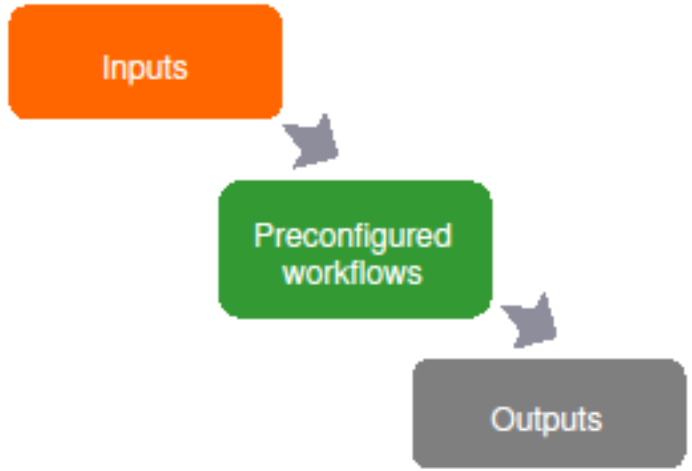
- Run model/workflow with pSeven Runner
- Analyze results and other engineering data

Configure sophisticated and multilayered WORKFLOWS



WORKFLOW can be used and reused multiple times by multiple users

Share it with your colleagues, so they so they can use them as ready tools, changing inputs and parameters that you make available on the Run screen



DATADVANCE

pSeven remote execution and HPC

Easy remote execution with a pSeven Agent:

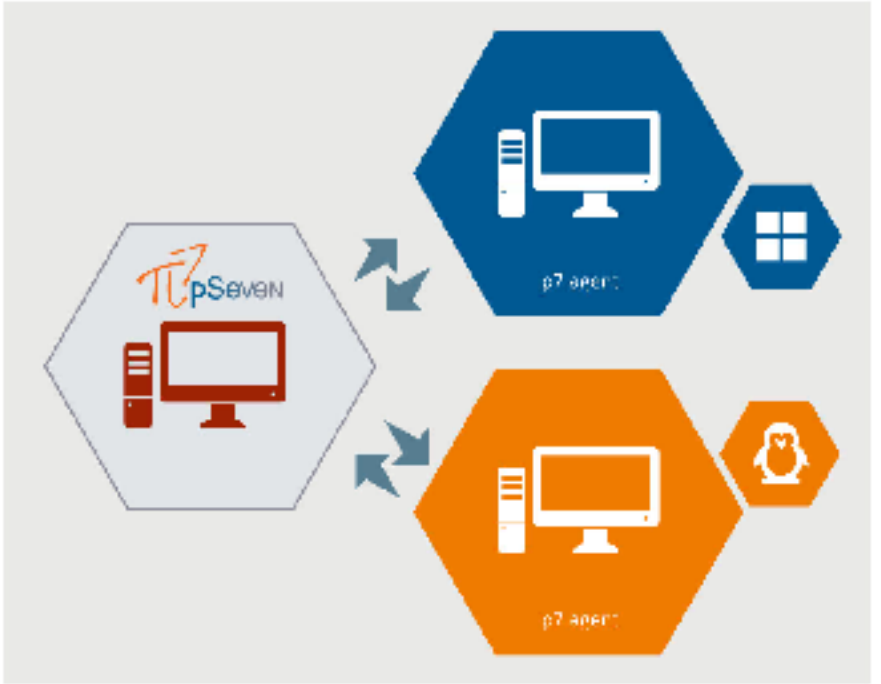
- Run remote scripts with SSH connection
- Create a flexible grid with pSeven remote agent
- Execution on Windows and Linux

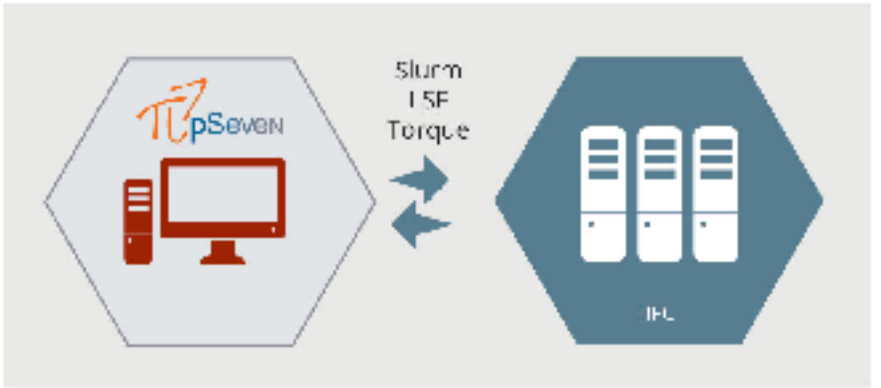
Parallel execution main features:

- Easy handling of batch input (list of parameters)
- Run as many parallel instances as you want
- CAD blocks automatically rebuild model
- Perform remote HPC calculations inside the parallel composite

HPC support:

- Built-in support of Job Array mechanism
- Direct interfaces with Slurm, LSF and Torque
- Automation of data synchronization (file management)
- Speedup the workflow just in a few clicks





DATADVANCE

Agenda

- Design Space Exploration with pSeven
  - Data & Model Analysis
  - Predictive Modeling
  - Design Optimization
- pSeven Platform
- **Summary**

DATADVANCE

pSeven key differentiators

- **Complete design space exploration toolkit**

- **SmartSelection™ for non-math experts**

- **Industry proven algorithms and techniques**

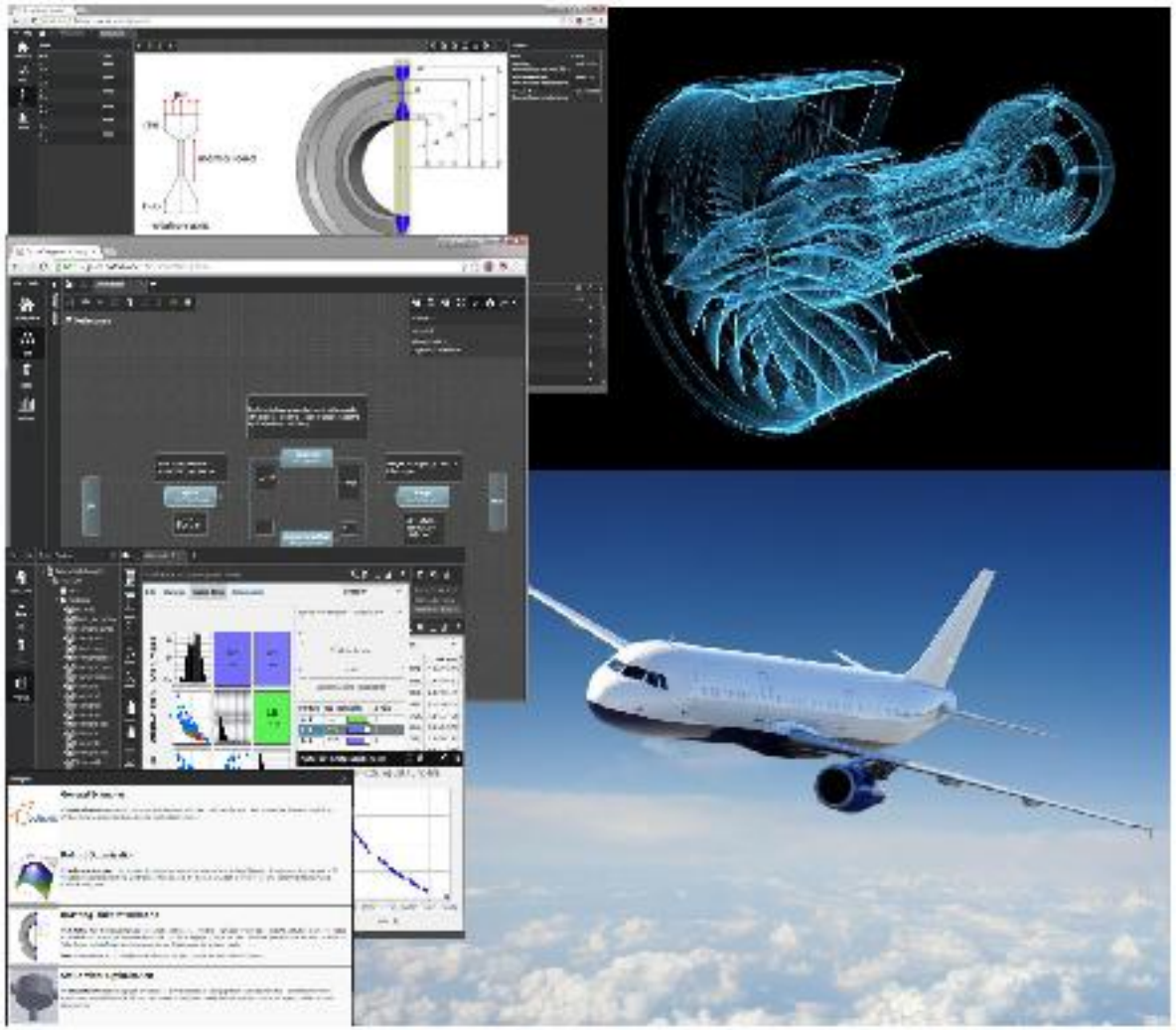
- **Platform approach and run-ready workflows**

pSeven is your Design Space Exploration tool

Efficient, reliable and scalable solution for design optimization and data analysis experts and non-experts.

Develop the best product with

pSeven



DATADVANCE

pSeven

Visit us

datadvance.net

Follow us

/DataAdvance

/company/dataadvance

Contact us

info@datadvance.net

18 rue Marius Tercé,
31025, Toulouse, FRANCE
Tel: +33 (5) 61 16 88 92

Nauchny proezd 3,
117246, Moscow, RUSSIA
Tel: +7 (495) 669-68-15

DATADVANCE