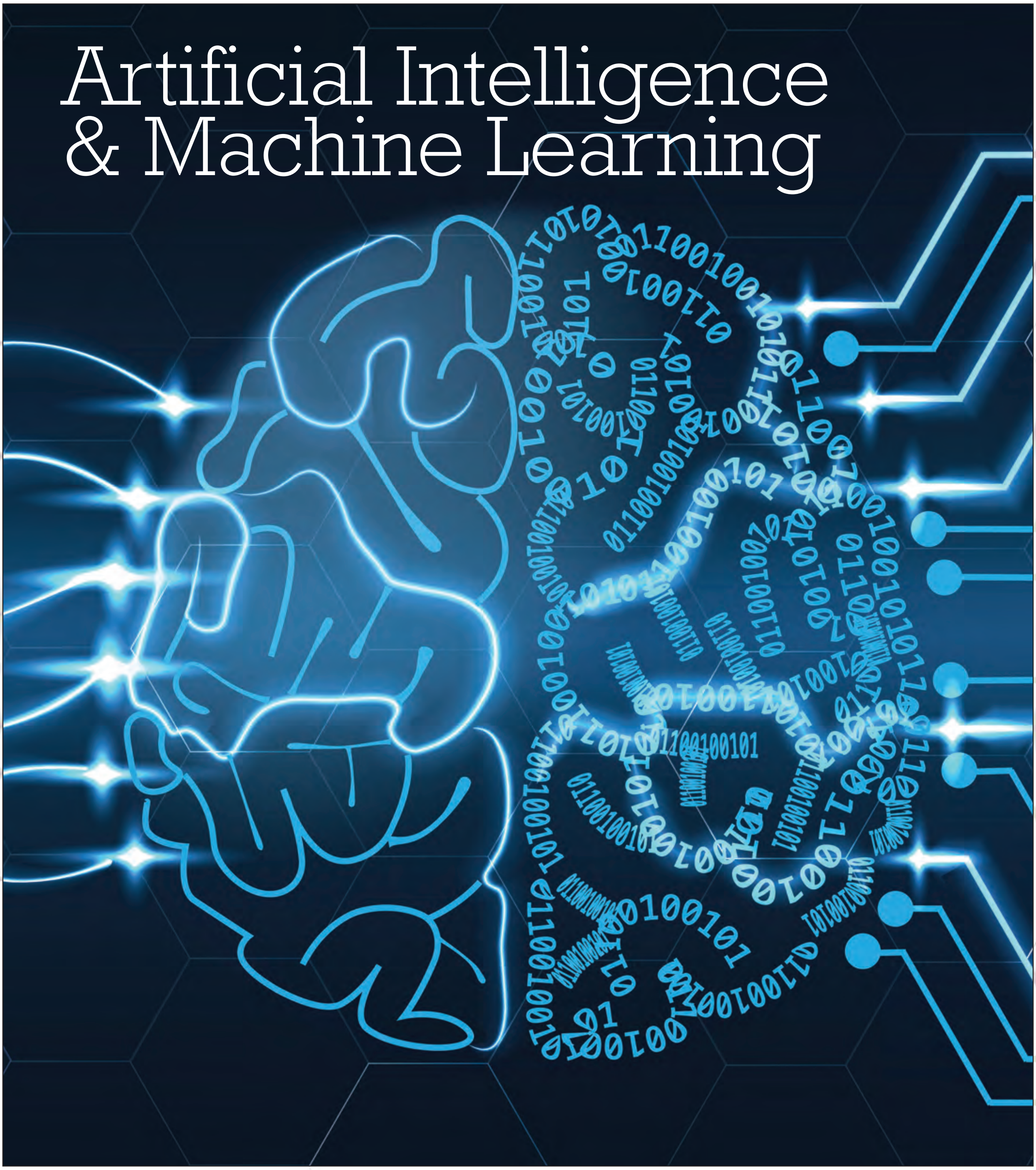
BENCH MARK

July 2018 issue . . .

THE INTERNATIONAL MAGAZINE FOR ENGINEERING DESIGNERS & ANALYSTS FROM **NAFEMS**



How Machine Learning Empowers Models for Digital Twins

Dmitry Frolov, DATADVANCE

Although the component structure of digital twins is not yet defined clearly and can differ drastically from implementation to implementation, we can see that one thing is present everywhere and powers the whole process and that is predictive models. Predictive models are used in the core of almost every digital twin that is already implemented or being developed and by using machine learning to create such models, the full range of data, from all sources, can be used. By predictive model we mean a digital representation of asset's behavior that not only gives the understanding of the behavior itself but often collects individual changes of the asset and adjusts itself accordingly.

Machine Learning for Building Predictive Models

Traditionally, predictive models can be built using existing parametrized simulations or something more sophisticated from an engineering point of view, like full-scale model-based system engineering (MBSE). Evaluation of these simulation-based models, of course, takes significant time. To clarify, in this article we refer to models that use physics-based, numerical computations, such as those using FEA or CFD techniques, as simulations.

While industry moves towards web-based applications, cross-department collaboration and engineering software democratization, predictive models should become more accessible to build, safer to distribute and faster to operate. This means that the available design, simulation, production and operational data and models should be converted from thousands of items and resource intensive simulations into easy-to-use and fast functions.

Machine learning algorithms come in handy in converting such data into such functions. Machine learning is an umbrella term that often covers different techniques and algorithms used in data mining, deep learning, predictive analytics, etc. Engineers often implement algorithms like Gradient Boosted Regression Trees (GBRT), different variations of Gaussian Processes (GP), High-Dimensional Approximation (HDA), Mixture of Approximators (MoA), Tensor Approximation (TA), Piecewise Linear Approximation (PLA) and many others. All these algorithms help to make data-driven predictions or decisions, through building a model from sample input and output data series. In a machine learning context, this process is known as training.

As a result of implementing machine learning algorithms, the user gets a data-based predictive model. Essentially, a data-based model is often nothing more than a complex polynomial that describes the multidimensional response surface of the model or, in other words, a substitution ("black box") of existing experimental or simulation data.

Different software vendors use different terms and meanings for data-based models, for example, approximation models, Response Surface Models (RSM), Reduced-Order Models (ROM), metamodels, surrogate models etc., but essentially, they are all pretty similar by function – they predict or allow model outputs to be assessed in advance.

The first and obvious question here is how many points in those data series do we need to create a reliable data-based model? There is no universal answer to that question. In most of the cases more points mean better quality, but at least it will need N points if the problem is linear and N^2 when it's nonlinear, where N is the number of input parameters. Using fewer points is just senseless. The crucial question here is the quality check of the model afterward; a process known as model validation. For that purpose, we should have a separate set of data points (test data) for comparison, which were not used during the training. After training is finished, the model predictions are compared with the test data on the same inputs, as shown in Figure 1. Comparing the difference between the output values gives accuracy estimation of the model.

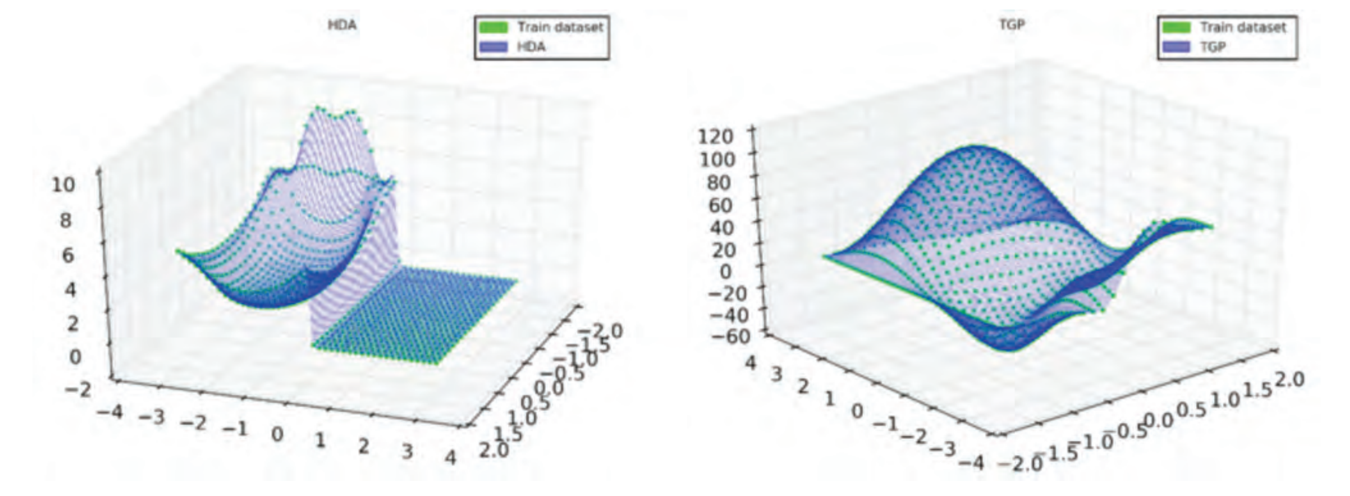


Figure 1: Example of data-based models built with machine learning algorithms

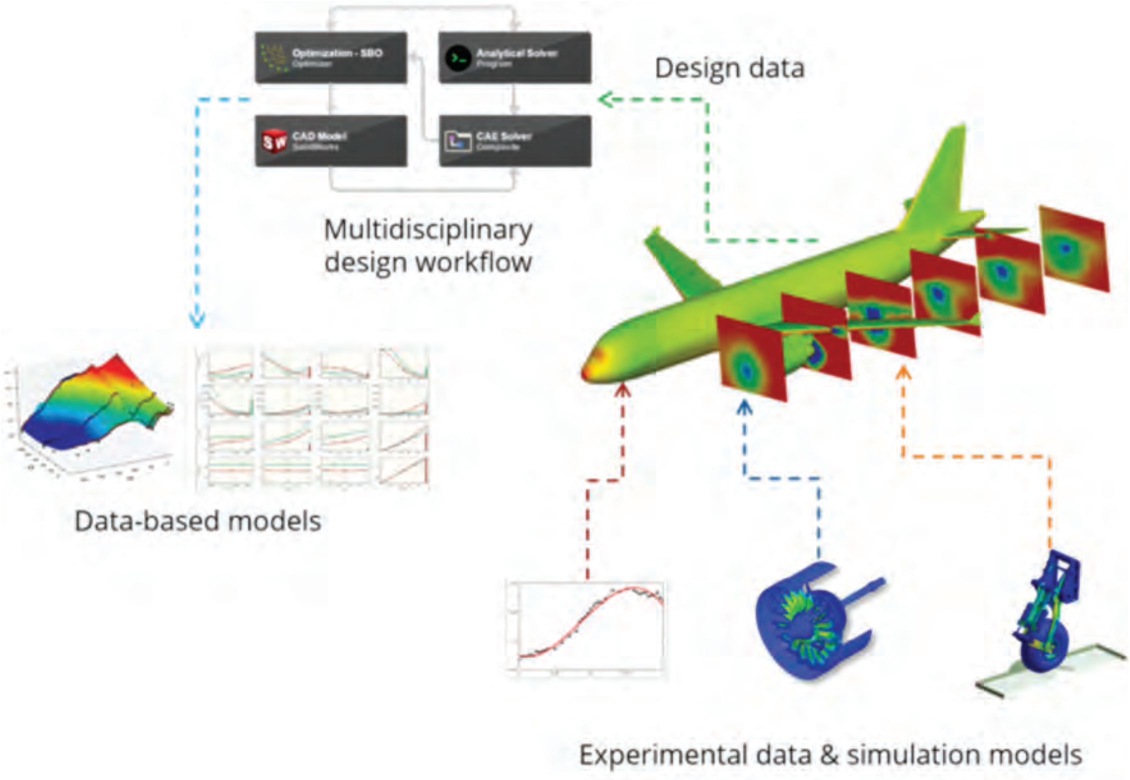


Figure 2: Design data turned into data-based models

Data-Based Models for Design

Performance optimization of an asset at conceptual or later design stages leads to significant iterative changes, requiring the performance prediction simulations to be repeated with different parameters. Using characterization and reduced order models can save a lot of computational time and resources. Which means that data-based models are integrated into the design process, as shown in Figure 2.

In situations when companies develop assets based on standard but still variable components, capitalizing on such data-based models can bring significant competitive advantage in shortening of time to market.

Data-Based Models for Manufacturing

During production, the characteristics of materials or quality requirements can vary in specific ranges and being able to predict the effects of these variations on site, allows the production processes to be optimized accordingly. In more sophisticated scenarios, production processes can be optimized on the fly automatically with respect to online sensor readings.

Predictive models for production processes can be built from historical, experimental, test, analytical or simulation data. To get the best of the available data, sometimes it is good to take advantage of data fusion. For example, data from physical experiments and simulations can be combined in the resulting data-based predictive model.

In 2016, a project with the Skolkovo Institute of Science and Technology (Skoltech) accomplished the task of optimizing manufacturing parameters of a composite beam pultrusion process[1].

As a first step of the solution, a uniform Design of Experiments (DOE) was conducted to study the simulation model behavior and sensitivities. A sample of 45 points was obtained with the Latin hypercube sampling method and the points evaluated using coupled thermal-structural analysis in Abaqus. The simulation model was in advance calibrated on the real-life experiment. Sensitivity analysis was conducted on this data to estimate how variations in the model output can be attributed to variations in the model inputs.

Based on this data, a predictive model based on Gaussian processes was built with a maximum Relative Root Mean Square (RRMS) error (based on training sample) of 0.04 thanks to extensive internal validation algorithms. The resultant model can predict in advance the cure degree and stress in the beam using the combination of input parameters, like pulling speed, die temperatures and so on.

Since one of the optimization goals was to reduce deformation of the pultruded part, the springback angle distribution over the parameter space was also studied and found that an increase in pulling speed does not lead to significant deformation; this finding allowed the engineers to expect a flat Pareto frontier when carrying out the Pareto (multi-objective) optimization of springback angle vs. pulling speed.

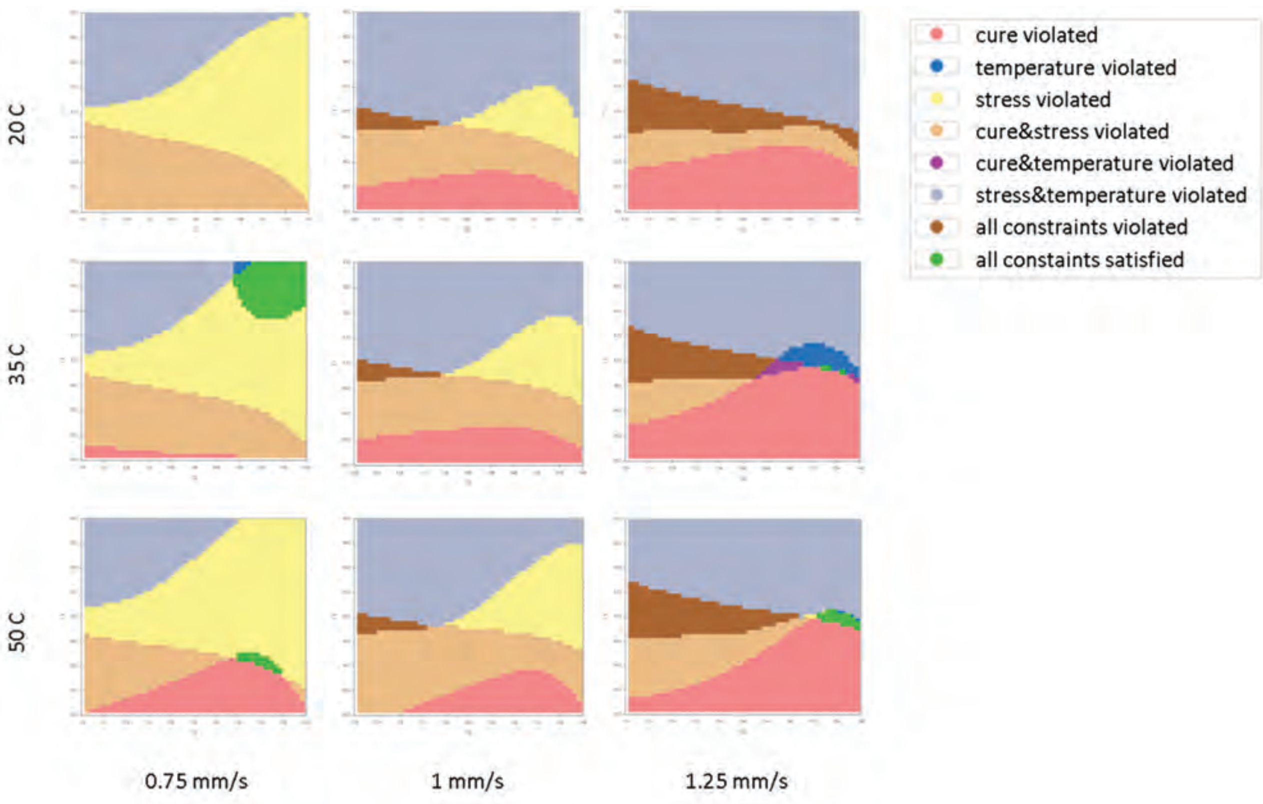


Figure 3: The predictive model allows visualization of the areas for different constraint violations of the pultrusion process. In this figure, the global X axis represents pulling speed, global Y – initial temperature, local X and Y – two die temperatures.

The obtained predictive model allows visualization of the areas for different constraint violations. Its study showed that it is possible to obtain configurations that satisfy all constraints. However, the allowed solution area (shown in green in Figure 3) is quite small.

The more data that is used for building such predictive models - the more sophisticated predictions they can make. This allows adjustment or even optimization of the production process on the fly, for example, when the material composition or technical conditions of

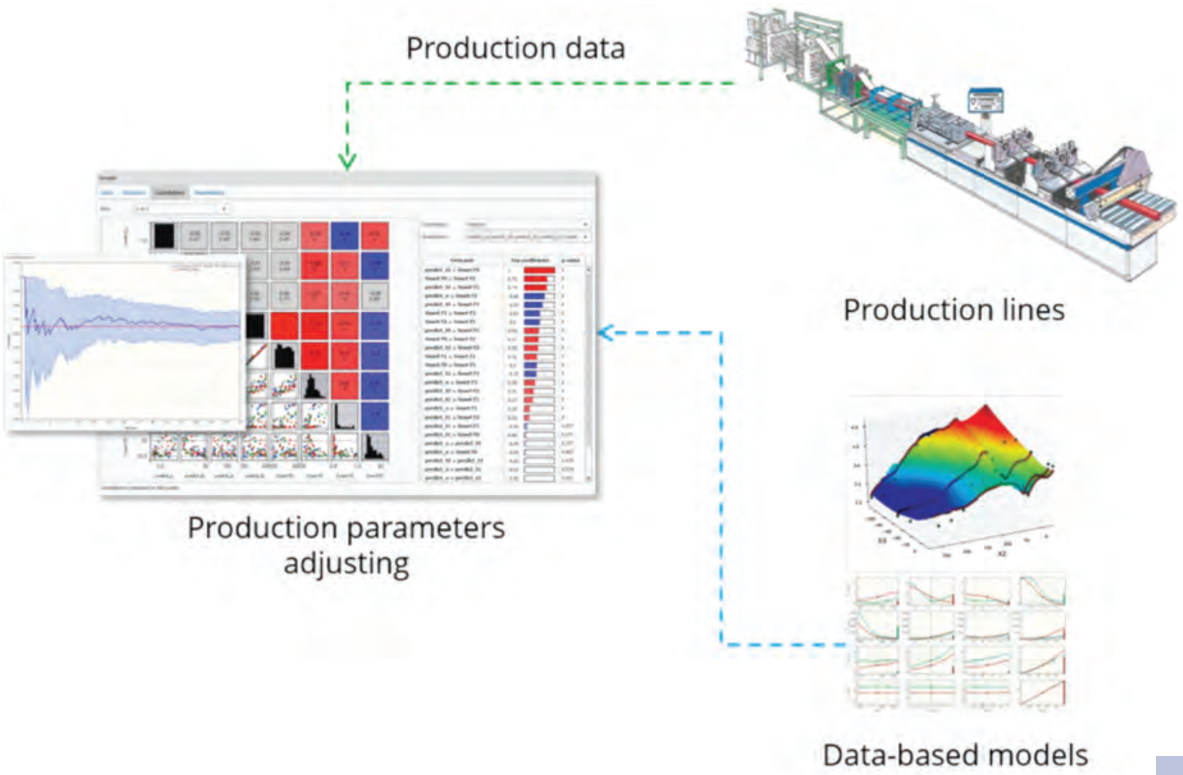


Figure 4: Production lines adjustment based on data-based models

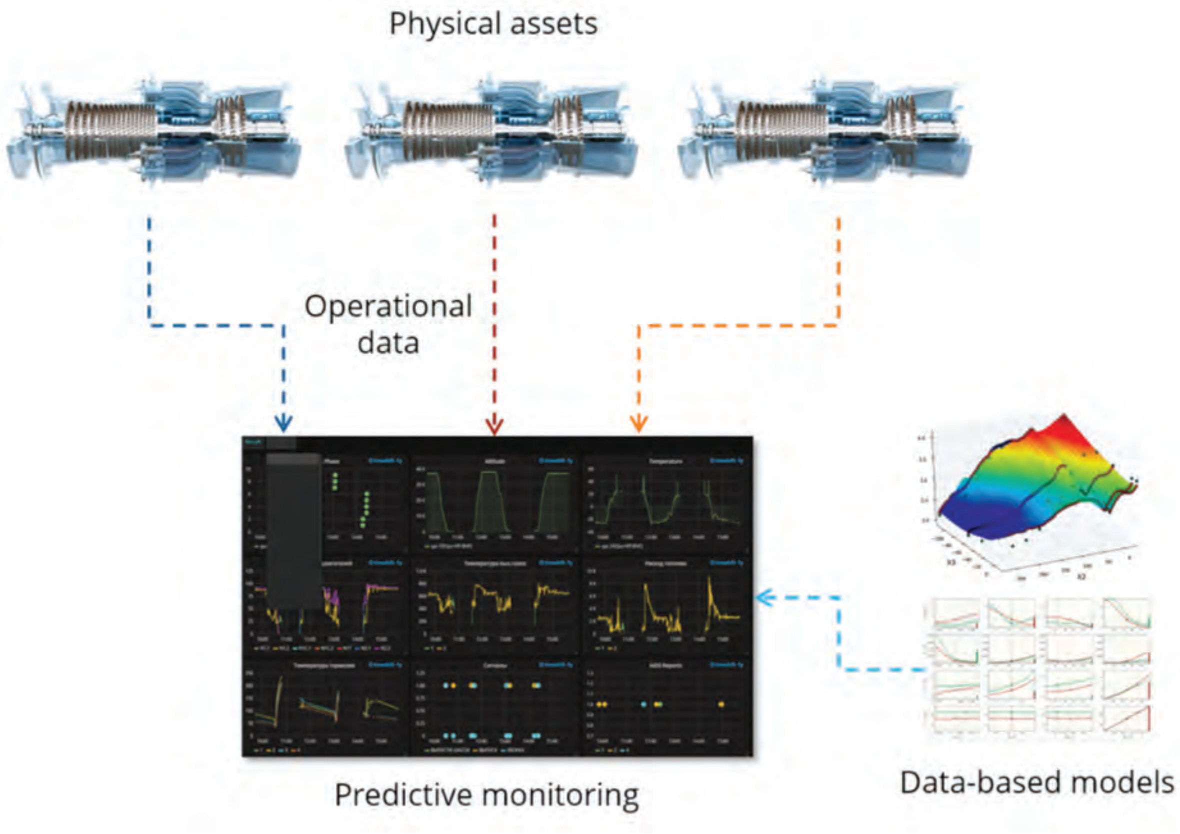


Figure 5: Asset behavior monitoring systems based on data-based models

production have slightly changed, as shown in Figure 4. By using multi-objective surrogate-based optimization and a fast-to-evaluate predictive model built using machine learning and based on simulation of the process, the pulling speed was increased by 18%, while satisfying all the mechanical constraints.

Data-Based Models for Operation

If the asset is operating under normal conditions, the corresponding digital twin collects data for future use. However, when there is a requirement for real-time performance optimization or upcoming/occurred malfunctions, requiring decisions to be made in a short period, then the speed of predictive model response becomes mission-critical.

If a full-scale 3D simulation is used for the underlying predictive model, the response needed can be acquired long after the useful window of opportunity. Another approach is to use machine learning with the historical data from asset operation in normal conditions together with data series containing pre- and post-malfunction behavior for comparison with sensor readings coming from the physical asset. This approach is widely used in Predictive Maintenance as a part of MRO (Maintenance, Repair and Overhaul) activities.

In 2018 a pilot project, shown in Figure 5, was carried out with S7 Airlines to develop a predictive maintenance system for the Airbus A319 aircraft fleet. This system makes long-term predictions of the potential failures of each S7 Airlines aircraft based on the analysis of the historical datasets on aircraft maintenance and individual component operation.

The goal of system implementation is to reduce the number of flight delays caused by technical issues in jet engines and hydraulic systems. The system estimates the probability of various types of failures in the defined upcoming period. If this probability turns out to exceed the preset level, additional aircraft diagnostics is recommended.

To build such a system, terabytes of parametric operational data are enriched and synchronized with data about technical maintenance works and weather conditions. Resulting time series are then transformed into thousands of features. Such features, for example, can be minimum, mean or maximum values of parameters on a fixed time interval, etc. Of course, training predictive models using all these thousands of features at once leads to overfitting, which is why optimization algorithms were used to find an optimal combination of features that leads to good quality for each particular model.

The main difficulty of this approach in aerospace is that the ratio of operational data with and without malfunctions for airplanes is somewhere near 1 to 10,000. It means that validation of the trained models (as explained earlier) is more than critical. Machine learning alone doesn't solve this problem. Applying engineering approaches and understanding of the subject area is very important. For example, airplane symmetry allows the training data set to be increased with malfunctions for specific parts of the jet engine; because data from an engine on one side of the aircraft can also be used for predictions for engines on the opposite side.

It's too early to forecast exact financial results coming from this project, but the initial aim of reduction of unplanned downtime by 10%, already leads to savings of millions of dollars each year. The next step is to implement the production system with increased efficiency, enabling real-time online data export and with a user-friendly interface.

Conclusion

Each industrial organization has yet to determine the usefulness of machine learning and digital twins for their products and processes. Some may benefit rapidly, some in the long run and some may just lose time trying to implement the wrong concept or techniques, but with time, implementing machine learning algorithms in design, production and operation stages will be as traditional as todays engineering software tools. ■

References

[1] Bruce Jenkins, Digital Engineering (2016). Optimizing Pultrusion of Glass-Fiber Reinforced Plastic Components nafe.ms/2N1CYWu

Dmitry Frolov graduated from the Ural Federal University in 2010 with a degree in Computer Aided Design and Engineering. After graduation, he worked in the technical support of ANSYS software for mechanical analysis and optimization. He participated in the development of a new family of engines in the Ural Diesel-Motor Works LLC as a leading design engineer. He has worked as a marketing director in the largest Russian distributor of ANSYS software and is now Marketing Director at Datadvance (www.datadvance.net). He has been chief editor of two technical product journals and is the author of several articles, including optimization studies.