



D10.3 Model integration principles

Version 1.1

Due date	31.12.2025
Dissemination level	PU
Work package title	Single-Operation Machine Learning
Lead editor	Gunnar Mathiason (HiS)
Internal reviewer	GSW
Status	Final version v1.1 for (resubmitted 30.6.2026)

This project has received funding from the European Union under grant agreement NUMBER —101178721— ProcTwin

The information and views set out in this document do not necessarily reflect the official opinion of the European Commission. The European Commission does not guarantee the accuracy of the data included in this document. Neither the European Commission nor any person acting on the European Commission's behalf may be held responsible for the use which may be made of the information contained therein.

Authors

Name	Partner
Gunnar Mathiason	HiS
Addi Ait-Mlouk	HIS
Frank Eriksson	Swerim
Patrik Siderstam	Swerim
Olle Sandin	Swerim
Gustav Dafgård	Prevas
Daniel Scheiber	MCL
Christoph Nölle	BFI
Ana Lopez Diez	GSW

History

Date	Version	Authors	Description
04.12.2025	0.1	Gunnar Mathiason, Addi Ait-Mlouk	Document structure and content overview. Initial description of integration principles.
10.12.2025	0.2	Frank Eriksson, Patrik Siderstam, Gustav Dafgård, Daniel Scheiber, Christoph Nölle, Gunnar Mathiason, Addi Ait-Mlouk, Olle Sandin	Descriptions of physical models. Integration principles expanded.
12.12.2025	0.3	Ana Lopez Diez, Addi Ait-Mlouk, Gunnar Mathiason	GSW review comments integrated.
18.12.2025	1.0	Gunnar Mathiason	Submitted version.
10.06.2026	1.1	Gunnar Mathiason	Reduced details in section “3.2.1 SSAB” to allow dissemination level PU. Change reviewed by SSAB, and the internal reviewer GSW. Updated submission version is now 1.1



Abstract

This deliverable document describes how physics-based models of key manufacturing operations are integrated with data-driven models of the same. Physics-based models and data-driven models take different approaches for establishing process simulations. One expected key result of ProcTwin is to benefit from the integration of these different approaches that benefit from the complementary properties of them.

With a combination of data-driven models and physics-based models, the resulting model is expected to perform better, by considering and integrating existing process knowledge into data-driven models.

To later be able to integrate multiple machine operation models in a distributed data-driven model for a production route, there is a need at this stage in the project to unify the various existing physics-based models with local data-driven models of each operation. Then, locally integrated models for each operation can be joined in an overall distributed model in WP11 and WP12.

This document explains the principles for local (per operation) model integration. The purpose of establishing such integration principles is to define a generic approach that is independent of specific physics-models' designs, modelling platforms, or model setups. The first part of the document presents the principles for such integration, while the second part expands specifics for how to set up the used physics-based models to integrate with data-driven models.

Table of contents

Authors	1
History	1
Abstract	2
1. Introduction.....	5
1.1 Scope	5
1.2 Deliverable role with the project.....	5
1.3 Outline	5
2. Operation-based model integration	6
2.1 Conditions for the proposed model integration approach	6
2.1.1 Physics-based models	6
2.1.2 Validation of physics-based models.....	6
2.2 Integration principles	7
2.2.1 Publication of process data from an operation	7
2.2.2 Usage of process data in physics-based models.....	8
2.2.3 Output of physics-based models.....	9
2.2.4 Data-driven model training using process data	9
2.2.5 The virtual-sensor concept.....	9
2.2.6 Simplified physics-based models for improved execution performance.....	9
2.3 Joint training	9
2.3.1 Joint training on process data and physics-based model output	9
2.3.2 Validity of physics-based model data	10
3. Existing physics-based models	11
3.1 Production definitions and model locations	11
3.1.1 GSW.....	11
3.1.2 SSAB.....	11
3.2 Material properties (D4.1, MCL)	12
3.2.1 Model adaptation for integration with data-driven model training	12
3.2.2 Model integration specifics	12
3.2.3 Integration input specification.....	12
3.2.4 Integration output specification	13
3.3 Continuous casting (D6.1, BFI)	13
3.3.1 Model adaptation for integration with data-driven model training	13
3.3.2 Model integration specifics	13
3.3.3 Integration input specification.....	13

3.3.4	Integration output specification	14
3.4	Reheating furnace (D6.2, Prevas)	14
3.4.1	Model adaptation for integration with data-driven model training	14
3.4.2	Model integration specifics	14
3.4.3	Integration input specifications.....	15
3.4.4	Integration output specifications.....	15
3.5	Wire rolling (D6.3, Swerim).....	15
3.5.1	Model adaptation for integration with data-driven model training	15
3.5.2	Model integration specifics	15
3.5.3	Integration input specification.....	15
3.5.4	Integration output specification	16
3.6	Quenching (D6.4, Swerim)	16
3.6.1	Model adaptation for integration with data-driven model training	16
3.6.2	Model integration specifics	16
3.6.3	Integration input specification.....	17
3.6.4	Integration output specification	17
3.7	Levelling (D6.5, Swerim).....	17
3.7.1	Model adaptation for integration with data-driven model training	17
3.7.2	Model integration specifics	18
3.7.3	Integration input specification.....	18
3.7.4	Integration output specification	18
4.	Status and coming actions	19
5.	Conclusions	20

1. Introduction

1.1 Scope

This document is the third deliverable of WP10 data and model preparations. It documents the basic principles for how physics-based models and data-driven models are integrated per machine operation of a sequential production route. The activities in WP10 aim to produce datasets and models for individual machines, emphasizing feature selection, data preprocessing, and integration to enable effective model development. D10.3 establishes the principles for how local models are integrated.

1.2 Deliverable role with the project

This document describes a part of the early dataset preparation activities and outcomes, supporting the development of ProcTwin's integrated initial machine learning models. D10.3 establishes generic model integration principles, and the main outcomes of this deliverable include:

- A description of how physics-based models are set up to generate virtual sensor information that can be used in line with actual sensor information coming from the production systems.
- The usage of data-driven models that use such virtual-sensor information, in combination with actual-sensor information to train machine learning models in the combination of both.

1.3 Outline

The document describes the following principal approaches:

- How virtual sensor information is generated by configuring physics-based models in a way that can simulate off-line model training using virtual sensor information.
- How offline and online model usage can be set up, and possible limitations that need to be overcome to use such combined models.
- How machine learning models are set up and evaluated to benefit from the combination of the two data sources.
- Descriptions of all physics-based models, their relevant inputs and outputs used, and how such models are adapted and parameterized to suit a data-driven training scenario.

2. Operation-based model integration

In this stage of the project, we integrate machine models *vertically*, that is, we model each operation individually using sensor data and existing physical models available for each machine operation. We make use of existing historical sensor data collected and stored from the current data collection setup, and we also make use of new sensor data from new sensors that are first used as part of other work packages in the project.

All collected sensor information is used for data driven modeling of each operation. There are physics-based operation models that already exist for some of the operations or are being developed during the project. Historical production data is fed to physics-based models to generate output based on model knowledge. Such output can be regarded as a sensor data stream that is a result of the physics-based models' process knowledge and can be utilized for data driven model training as virtual sensors in parallel with the measured sensors.

Thus, this integrated data-driven model training is based on the combination of data-driven models and physics-based models, made up of sensor data or virtual sensor data that are available for each operation during a specifically selected production period of historic data. Models trained off-line this way are meant to be taken online for quick model inference, used in parallel with the actual production.

Physics-based models may be computationally expensive, such that model output may be lagging compared to actual sensor data streams. To mitigate this, we aim to make use of surrogate models that can mimic a full model under some assumptions or fixed parameterizations, to fulfill the constraint that the virtual data needs to be produced timely, close to real time when the combined model is used online with production. We also consider an alternative approach, to train separate machine learning models for each of the complex physics-based models that are not timely. In this way, the online model inference time can be made much shorter compared to running a non-timely physics-based model online.

2.1 Conditions for the proposed model integration approach

2.1.1 Physics-based models

A physics-based model is based on the knowledge of a physical and/or chemical process in an operation. It formulates key properties and behavior of the operation, using parameters of the operation as inputs and generates plausible output in the execution of the operation.

Physics-based models are defined at different levels of detail, and they formulate different parts of the operation. For instance, a material behavior model is significantly different from a model of an entire wire rolling flow. However, basic principles are the same: Both types involve the formulation of constraints and definitions under which the production executes also formulate the key properties of the specific ongoing physical/chemical process.

In this way, physics-based models can simulate actual operations at some level of fidelity. Every model needs to be validated against the actual physical/chemical process, under the specific assumptions for which the model was created, since every model is a reduction of the full complexity of the physical/chemical process.

2.1.2 Validation of physics-based models

Models must generate output that is true to the physical/chemical process, and relevant for the case for which it is developed. The models used for our integration are validated at various degrees, and

some models are planned to be validated during the ProcTwin project. The principles for integration remain the same regardless of the state of validation, and as soon as models have been validated, these models are integrated.

2.2 Integration principles

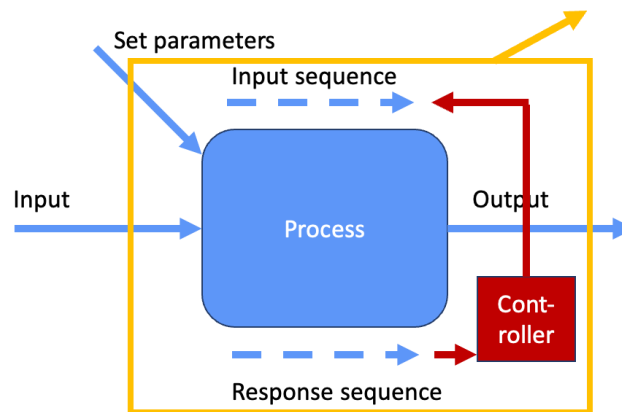
2.2.1 Publication of process data from an operation

Every process model considered in this project is complex, while only some of the total number of process variables for each process are made available for data-driven model training. The available process variables are selected based on the process experts' knowledge and made public to the distributed system by the data communication infrastructure. The principle is that the available variables of each process are *published* to the other processes to access using the communication or database infrastructure. For each process, there are four types of variables published (Figure 1): 1) *Background variables (Process input)* that define the condition for the process execution. They can be a result of the environment or a result of the proceeding processes 2) *Set parameters* are variables that configure or define the current execution of the operation, such as process recipe variables and set values. Set parameters can be fed to the process during the execution as well, as single values or as a sequence of values during the execution (*Input sequence*). 3) *Response parameters* describe the current execution, as an aggregated value, or as a sequence of readings along the execution of a process. A typical process sensor setup will produce a number of such sequences, and they are typically collected process control and production reporting. 4) *Result parameters* are metrics of the process output, typically as product quality metrics.

What is published about a process depends both on what can be published about the process and how the process is seen from the outside. Since there could be entities controlling the physical process, such as a control system or an operator, the physical process is affected by the decisions made by these entities. The published variables describe the execution of the process as seen from the outside, including the decisions taken by a controller or an operator. In contrast, a physical model typically simulates the actual physical process.

The execution response depends on how the process is controlled. Any process response depends on the inputs and control signals. Information about such decisions should be published and used for data-driven model training. A typical example of such decisions is when an operator uses her knowledge to deviate from a prescribed recipe, using experience and known inputs in combination. Such a deviation is a signal that can give a different output. It captures more information about how the process is controlled to decide response variables. Thus, data-driven process model considers the context of the physical process. A joint model is expected to be able to capture the full behavior in production.

Figure 1. Types of process data. The yellow border represents published data.



2.2.2 Usage of process data in physics-based models

The selection of appropriate time windows for production data is critical for effective physics-based model integration. Production data should be extracted from periods that represent typical operating conditions with consistent equipment performance and minimal unlabelled process disturbances. A typical time windows range from several days to months, depending on the process' complexity and the data availability.

The chosen time window must ensure sufficient data volume for statistical significance while maintaining process consistency. Seasonal variations, equipment maintenance periods, and major process changes should be considered when defining the temporal boundaries. For continuous processes, rolling time windows may be employed to capture process evolution over time.

Data synchronization across different measurement systems and process stages is essential within the selected time window to ensure temporal alignment between input parameters and corresponding outputs. Both physics-based models and data-driven models make use of the same input parameters, from the model context. Physics-based models require careful selection of a subset of context parameters that significantly influence the process outcomes while maintaining computational efficiency. Parameters are categorized into:

Variable Input Parameters (Input Sequence or Response parameters):

- Process control settings that can be adjusted during operation
- Material properties that vary between production batches
- Environmental conditions that fluctuate over time
- Equipment state variables that change during the process

Fixed Input Parameters (Background or Set parameters):

- Equipment specifications and geometric constraints
- Physical constants and material properties that remain constant
- Standard operating procedures and regulatory constraints

The selection of variable versus fixed parameters depends on the model's intended use, computational constraints, and the availability of real-time data. Parameters with minimal impact on the process outcome or those difficult to measure accurately are typically fixed to their nominal values to reduce model complexity.

2.2.3 Output of physics-based models

Physics-based models generate synthetic data streams that complement actual sensor measurements, providing comprehensive process insights for data-driven model training. The model outputs are structured to match the format and temporal resolution of actual sensor data, enabling seamless integration in the training pipeline.

2.2.4 Data-driven model training using process data

The integration of multiple data sources from different processes requires a unified approach that combines sensor measurements with physics-based model outputs, enabling multimodal data fusion for comprehensive process modeling grounded in both empirical observations and theoretical knowledge.

With every labelled production output (the combination in data entity of {product-ID, quality-metric}) there is an associated trace of sensors and virtual sensor readings that describes the process while producing the production output. Such a combination of data is seen as a training example for a data-driven model. In supervised model training, a model is trained by presenting many such examples to the learning algorithm. Thus, traceability by connecting the production output with sensor and virtual sensor data is essential

In the later stages (WP11 and WP12) of the project, the full production route is integrated. Data streams from sequential manufacturing operations are synchronized and aligned to form datasets that capture the full production chain, preserving cross-process correlations through careful preprocessing and feature engineering. Machine learning models are then trained on these *horizontal* datasets, incorporating both measured sensor data and virtual sensor outputs from multiple operations.

2.2.5 The virtual-sensor concept

Virtual sensors provide access to otherwise unmeasurable process parameters, overcoming physical inaccessibility, high measurement costs, temporal limitations, and safety constraints. They are for instance used to estimate internal temperature distributions in heating processes, stress and strain fields during deformation, chemical composition changes in metallurgical reactions, real time quality parameters normally measured offline, and equipment wear or remaining useful life. By extending observability beyond physical measurement limits, virtual sensors enable more comprehensive process monitoring and control.

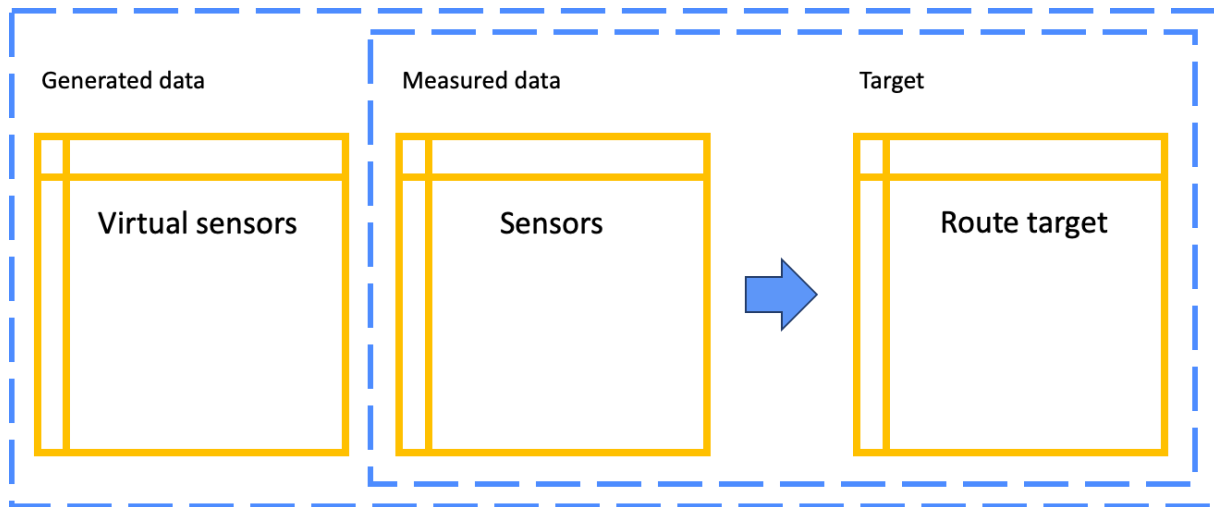
2.2.6 Simplified physics-based models for improved execution performance.

A surrogate model is a simplified mathematical representation of a complex physics-based model that preserves essential input output relationships while greatly reducing computational demands, making it suitable for real time use. Because many high-fidelity physics models are too computationally intensive to support online process control or continuous virtual sensor generation, surrogate modeling provides the necessary simplification without losing critical process behavior. Surrogate models may be developed using response surface methods, machine learning models trained on high-fidelity outputs, reduced order modeling techniques that capture dominant physical phenomena, or hybrid approaches that blend simplified physics with data-driven corrections. These models retain sufficient accuracy for their intended applications while delivering the computational efficiency required for seamless integration into real-time manufacturing environments.

2.3 Joint training

2.3.1 Joint training on process data and physics-based model output

Figure 22. Training of data-driven models using sensor and virtual sensor data.



Joint training on process data and physics-based model outputs combine empirical sensor measurements with theoretical model predictions to create robust machine learning models that leverage the strengths of both data driven and physics-based approaches (Figure 2). Training data from sensors and physics models is carefully preprocessed to ensure temporal alignment, consistent feature representation, and proper handling of differing sampling rates, uncertainties, and data gaps. The resulting models are designed to accept multimodal inputs, incorporating both measured and virtual sensor signals while accounting for the distinct characteristics and reliability levels of each source.

2.3.2 Validity of physics-based model data

It is essential that physics-based models deliver virtual sensor data that is consistent with the physical process. Validation of the model may be difficult, in particular for models that generate process information that cannot easily be verified, such as an internal temperature gradient. Some of the physics-based models in the project have proven themselves by extensive usage in production (Furnace model), while some are new (Levelling model) and need to be validated. A physics-based model that is not validated risks delivering virtual sensor data that is inconsistent with the actual sensor measurements, resulting in low-performing combined data-driven models using that virtual sensor data. Further, the physics-based model needs to represent the process well at the time that the sensor data was collected. An operation may change or be reconfigured such that the physics-based model becomes less accurate or valid.

3. Existing physics-based models

3.1 Production definitions and model locations

For each of the user cases of the project, we have decided on one specific route an unlimited production variety for each user case. Each route involves multiple physics-based models, a vertical data-driven model for each operation, and later a horizontal data-driven model for each route.

3.1.1 GSW

GSW operations route:

EAF; Ladle furnace; Continuous casting.

Reheating furnace; Shears; Multiple rolling operations.

Production output definition: Eddy current defect inspection.

GSW Production data definition:

Wire rolling dimensions 10-14 mm.

Time period for production data: 2024-06-12 to 2025-02-27 (other production periods will be collected and used for model training as well, and different periods will be compared).

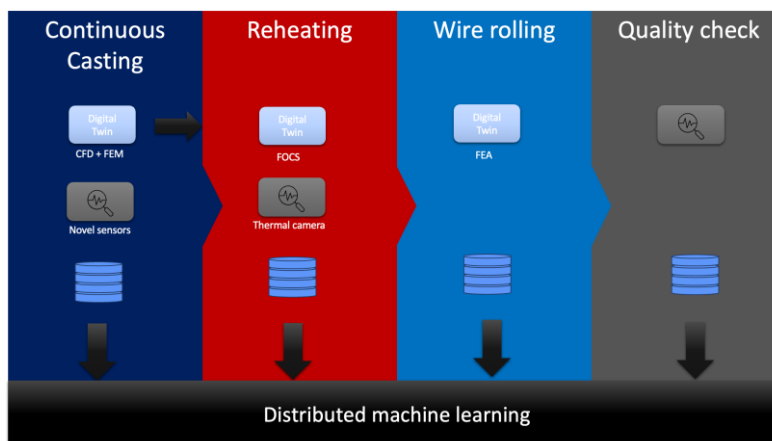
Sensor data, sources:

- Billet route (EAF, Ladle furnace, Continuous casting).
- Coil route (Reheating furnace, Shears, Rollers).

Steel grades (final grade): 16 selected high-volume grades

Using physics-based models for D6.1 Continuous casting, D6.2 Reheating furnace, D 6.3 Wire rolling (including D4.1 material model but applied for the Garret route in the project).

Figure 33. GSW production route schematic.



3.1.2 SSAB

The SSAB operations route:

Hot rolling; Heating + Quenching; Levelling

Production output definition: Shapeline flatness.

Time period for production data: January, April, August 2024

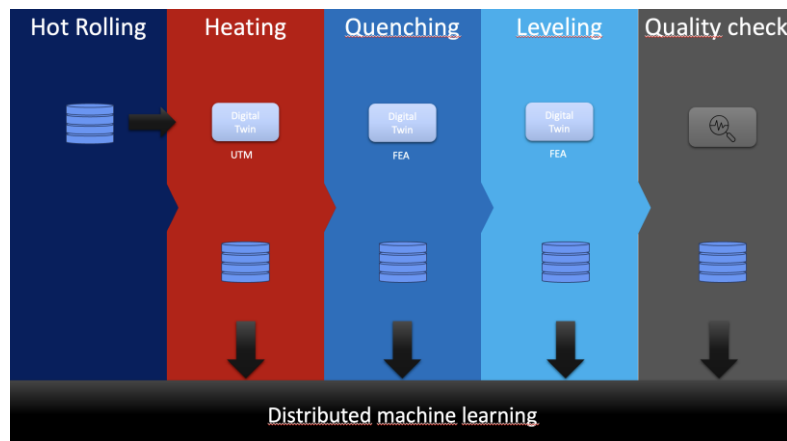
Sensor data sources:

- Rolling process
- Leveling process

Steel Grade: HX450 (wear/abrasion resistant steel). Material Strength: 1,250 MPa yield strength typically.

Using physics-based models for D 6.4 Quenching, D6.5 Levelling (including D4.1 material model).

Figure 44. SSAB production route schematic.



3.2 Material properties (D4.1, MCL)

This model is used by one or more other physics-based models, as defined below.

Model for evolution of material properties in the hot rolling process with a physical basis, if possible, or with a semi-empirical model otherwise. First approach is to go via an internal state variable model using a physical basis. If this does not give the right accuracy and speed, a semi-empirical model is considered. The internal state variable model evolves internal state variables such as grain size, dislocation density, or phase fractions as function of the processing, considering physical relations such as solute diffusivity with temperature, or deformation upon dislocation generation. As the internal state variables determine the mechanical properties, this provides also information on stress-strain curves as a result of a given set of process parameters. The model will be parametrized with dedicated experimental data on the specific materials in question from the use cases and with available literature data.

Depending on the material complexity, the model can run in real-time.

3.2.1 Model adaptation for integration with data-driven model training

The model may run in real-time, which needs to be tested upon parametrization. If model execution times are slow, a surrogate model is envisioned. The parameterization will make use of lab experiments to simulate process-relevant temperatures and strain-rates.

3.2.2 Model integration specifics

The model serves as input for FEM simulations (Swetim) of processing on steel products encoding material properties as function of processing parameters. These FEM simulations in turn focus on levelling of steel plates as well as hot rolling of wires. The material properties obtained from the material model are not observable directly during production and enhances the process simulations.

3.2.3 Integration input specification

The following inputs are used by this model:

- Material data: grade, starting microstructure (from lab characterization), thermomechanical data (from FEM simulations)
- Process data, i.e. ranges for:
 - Time and Temperatures (from FEM simulations)
 - Deformation strains (from FEM simulations)

Model parametrization will make use of lab experiments of stress-strain curves from tensile tests and Gleeble tests that simulate temperatures and deformation rates from the real processes. The specimens tested in the lab environment will further be analyzed with respect to microstructural characteristics such as grain size and phase fractions, which will also enter into the model parametrization.

3.2.4 Integration output specification

The following outputs are generated by the model and used as virtual sensor information for data-driven model training:

- Predicted material properties per plate, including
 - Stress-strain curves or selected points of the stress-strain curve
 - Microstructure properties such as grain size for modelling of mechanical properties

3.3 Continuous casting (D6.1, BFI)

This is a model of the solidification process of steel in the casting mould. It describes the flow and temperature fields as well as the formation of a solidified shell supporting the liquid steel in the inside of the casting strand. The area covered includes the complete mould and half a meter of the casting strand below the mould, after which no complex flow effects are expected any more. It is a physics-based model using a CFD approach, which takes as input the process data from the casting machine, such as casting speed, tundish temperature and electromagnetic stirring intensity. Measurement data from the mould cooling water and from the newly installed fibre-optical temperature sensors will be used for validation.

3.3.1 Model adaptation for integration with data-driven model training

The model itself cannot be executed in near real-time, but the creation of a surrogate model is envisioned. Furthermore, a set of sample scenarios will be pre-calculated and results exported, so that for a given casting situation the results from the most similar scenario could be applied. The scenarios calculations will include a parameter-sensitivity analysis, which can serve as a guide for optimizing the casting settings, but can also be fed into the data-driven model.

3.3.2 Model integration specifics

The model serves as input for a data-driven quality prediction module for the continuous casting process, which processes the same input data as the solidification model, plus some additional sensor data recorded in the later stages of the casting process, i.e., below the mould, such as a strand surface temperature measurement by means of pyrometers. The solidification model provides calculated data about the shell-thickness in the mould plus effective stresses and plastic strains, as well as temperature and flow fields, none of which is directly amenable to observations, thus the model results enhance the process data.

Besides those low-level outputs, a derived shell-cracking indicator will be reported, which should be of high relevance to the quality predictor. It is derived from the accumulated plastic strains and the principle stresses.

3.3.3 Integration input specification

The following dynamic inputs are used by this model :

Process control settings that can be adjusted during operation :

- Cutting information, incl. billet ids

Material properties that vary between production batches :

- Heat data (steel grade and chemical composition), once per casting heat

Equipment state variables that change during the process

- Process data, such as mould water flow, mould water temperature difference, tundish temperature, mould oscillations, casting speed, electromagnetic stirring intensity, etc., in 1-10s resolution
- Optional:
 - Gas flow in SEN
- Data from mould powder supervision system (when available)

Static input:

- Geometrical data of Mould for the CFD model

For model calibration and validation:

- Mould temperature measurements from fibre-optical sensors (when available)
- Billet quality indicators (obtained after hot rolling)

3.3.4 Integration output specification

The following outputs are generated by the model and used as virtual sensor information for data-driven model training:

Solidification:

- Predicted liquid steel flow field
- Predicted liquid steel and solidified shell temperature fields
- Predicted steel shell thickness evolution
 - Additionally, surface crack risks
 - Identification of defect promoting conditions

3.4 Reheating furnace (D6.2, Prevas)

A model for the material temperature in the furnace. It calculates the temperature both on the surface and inside the material. Calculations are made on a cycle throughout the furnace, so temperature over time can also be extracted. The model is based on temperature measurements in the furnace and also fuel and gas flows. When running online the model uses real measured furnace data as input. When running offline the inputs are fed through a simulator. This simulator is a simplified version of the data and does not have dynamics as real process data. Reheating furnace, temperature of steel billets are modelled by this model. The model is based on physical laws, 2-dimensional FEM model. It can run in real-time

3.4.1 Model adaptation for integration with data-driven model training

The model can be run online, but all data is stored in SQL database so it is possible to run offline also. If the model runs online input sensor data can be read from the real furnace. If the machine learning model does not accept data from SQL it needs to be exported to other format first.

3.4.2 Model integration specifics

All data for each billet can be connected to the billet ID and thus make it relatable to other processes. It is primarily recommended to use the result parameters of temperatures from each billet as input to the machine learning model, secondarily there are response parameters of temperature through time for each billet. When run online the data of approximately two months of production is stored in the database, to have more data it needs to be exported and saved elsewhere.

3.4.3 Integration input specifications

The following inputs are used by this model:

- Billet dimension
- Steel grade
- Charging temperature
- Furnace temperature (zone thermocouples)
- Fuel flows
- Air flow
- Furnace movement

3.4.4 Integration output specifications

The following outputs are generated by the model and used as virtual sensor information for data-driven model training, with the most important data for machine learning model written first:

- Temperature at discharge from furnace
 - Core temperature
 - Surface temperature
 - Temperature gradient
 - Grid of temperatures for both calculation points
- Historical temperature for each billet
 - Core, surface & average temperature (2 dimensional segments for each calculation point of the billet)
- Historical temperature of the atmosphere in the furnace

3.5 Wire rolling (D6.3, Swerim)

The wire rolling model is a physics-based thermomechanical finite element model for a 2D cross section of the billet. It performs calculations on the cross section and simulates its progression through the wire rolling process, including the different groove rolling passes. This model is used in the GSW user case, where predictions are made for temperature, geometry and material properties. The model is using the material model developed in D4.1.

Mill specifications and process data are used to define the GSW user case in the model. Input data may also come from upstream processes, such as the billet temperature upon exiting the furnace, which is used as the input temperature for the wire rolling process.

3.5.1 Model adaptation for integration with data-driven model training

The model is relatively fast, with calculations done close to real time. Although the model could potentially use process parameters in real time, the simulation time and results processing are likely longer than the actual production time. The calculation time depends, for example, on the mesh resolution used to model the cross section. Therefore, simulations will be carried out with varying input parameters to develop a surrogate model that can be integrated into the data-driven model.

3.5.2 Model integration specifics

The physics-based wire rolling model uses both fixed and variable input parameters to calculate parameters that are largely impossible to measure directly, such as the distribution of temperature, stress, and strain within the cross section. In this sense, the wire rolling model can be regarded as a virtual sensor for the data-driven model.

3.5.3 Integration input specification

The following inputs are used by this model :

- Layout of the wire rolling mill, including roll diameters and the distance between rolling stands

- Groove table, i.e. specification of the roll profiles
- Pass schedule, such as rolling pass layout, rolling speed, roll gap, and input for intermediate induction heating
- Billet temperature at the exit of the reheating furnace, preferably measured at least at head, mid, tail
- Steel grade and/or chemical composition for each billet

3.5.4 Integration output specification

The following outputs are generated by the model and used as virtual sensor information for data-driven model training:

- Temperature
- Stress
- Plastic strain
- Plastic strain rate
- Microstructure (grain size)

These output parameters are calculated for one cross section per billet per pass. Depending on the input data, more than one cross section per billet may be calculated.

3.6 Quenching (D6.4, Swerim)

The numerical water-quenching model of the SSAB user case is a physics-based thermomechanical Finite Element analysis developed to predict the final material and geometric properties of the quenched plate. It uses the material model developed in D4.1 and enhances it with capabilities for simulating phase transformations and volumetric expansion as functions of the cooling process.

The model setup is defined by measured process data and the layout of the quenching line. Its outputs, such as residual stress distribution and plate flatness, are transferred to the downstream levelling model.

Initially, the model operates offline. While optimizations of the model may reduce its computational cost, achieving execution times suitable for real-time simulation is considered unlikely.

3.6.1 Model adaptation for integration with data-driven model training

To accurately capture the key effects of the quenching process, such as residual stress, phase concentrations and plate flatness, a highly complex modeling approach is required. This is particularly critical for the material model, which accounts for phase transformations and their associated thermomechanical effects that strongly influence the residual state of the quenched plate.

As a result, the quenching model requires significant computational cost and is therefore considered infeasible for online monitoring. However, constructing a surrogate model of the process could substantially reduce computational effort since the nonlinear thermomechanical problem would not need to be solved explicitly. Such a surrogate model would require the input plate data and process data described in Section 4.6.3 and generate the output data specified in Section 4.6.4. Since the material model relies on a large number of parameters, these would need to be fixed within the surrogate model. With this approach, online monitoring could potentially become achievable.

3.6.2 Model integration specifics

The quenching model generates output that are both measurable in-line (plate temperature, dimensions, flatness) and unmeasurable (residual stress, phase concentration). Therefore, the model may be integrated as a virtual sensor providing extensive process control.

The level of output results resolution is a function of both the level of model complexity and discretization. A complex and detailed model will generate high resolution data for utilization in downstream modelling but will end up with a high computational cost. Conversely, an overly simplified model may fail to generate data of sufficient accuracy or relevance. Based on the requirements of the interaction process and the acceptable computational cost, the quenching model outputs can be provided either as high-resolution spatial data or as single-point indices.

3.6.3 Integration input specification

The following inputs are used by this model :

- Plate data
 - Dimensions
 - Internal stress state
 - Flatness value
 - Temperature of plate out from furnace
- Material data: grade, microstructure, thermomechanical data
- Process data, 100 smp/plate
 - Water flow
 - Pressure
 - Number of active cooling zones
 - Plate speed

3.6.4 Integration output specification

The following outputs are generated by the model and used as virtual sensor information for data-driven model training:

- Plate data
 - Microstructure
 - Dimensions
 - Internal stress state
 - Flatness value

3.7 Levelling (D6.5, Swerim)

The levelling model is a physics-based mechanical FEM model. A representative mesh of the plate is used to simulate the process. The mesh is pulled through a leveling setup from real-life process data used by SSAB in their leveling machine. The setup consists of nine rolls: five in the lower row and four in the upper row. The rolls in the lower row are fixed in place, all positioned at the same height. The rolls in the upper row are fixed along the length direction but can be adjusted individually in height.

The model receives input data from the end of the quenching process, including residual stresses and flatness measurements. The levelling setup can be varied between runs by adjusting parameters such as the height of the upper rolls, the roll speed, the plate thickness, and the material model. The output data from the levelling model consists of residual stresses and flatness measurements.

The objective is to collect data from multiple simulation runs with varying input parameters, thereby creating a diverse offline dataset suitable for machine learning applications for the process.

3.7.1 Model adaptation for integration with data-driven model training

The model will be simulated as an offline model to gather large datasets that will later be used in machine learning algorithms. The model was chosen to be offline since the data used as input data for the levelling setup is real process data from SSAB which is sensitive information which cannot risk being leaked outside of the project. The levelling model is also a physics-based FEA and needs a high resolution to mimic the real-life process, this leads to a computationally heavy model which results in

long simulation times. The goal is to use the data from these simulations to develop a surrogate model to be used in machine learning to make a more effective model with similar accuracy as physics-based simulations.

3.7.2 Model integration specifics

The levelling simulation completes multiple batch simulations with varying input to create a dataset from the outputs of each run. The dataset will include input parameters for each run as well as internal stresses and flatness measurements as outputs. The dataset will later be used in a machine learning algorithm to predict the output data from input variables. The idea is to make the ML simulation more efficient than FEA while keeping the accuracy.

Since the raw data of residual stress and flatness from the FEA simulations is expressed in large matrices because of the resolution of the mesh, the machine learning could be difficult on the raw data. To combat this, the data needs to be postprocessed. University of Skövde and Swerim need to agree how the data will be sufficiently expressed to successfully create a machine learning algorithm while also keeping the usefulness of a raw dataset.

3.7.3 Integration input specification

- The following inputs are used by this model: Plate data
 - Dimensions (thickness, length, width)
 - Internal stress state from quenching simulation
 - Flatness value from quenching simulation
- Material data: Mechanical properties
- Process data
 - Roll gap
 - Speed
 - Leveling setup
- Residual stress and flatness from quenching process

3.7.4 Integration output specification

The following outputs are generated by the model and used as virtual sensor information for data-driven model training:

For each successful simulation, outputs for every element and node will be available. However, using every data point for the dataset could lead to excessively large datasets. This would make it more complicated as well as computationally heavy for the ML algorithm. To combat this, another strategy to interpolate or extract the most useful data has to be chosen. The data that can be extracted from the model now is as follows :

- Predicted material properties per plate, including
 - Internal stress
- Flatness
- Forces and/or roll gap

4. Status and coming actions

- Implement a first version of a data-driven model for an operation where there is a validated physics-based model available, but do not yet use the generated data. It is suggested to use the GSW furnace operation.
- Implementation of the first integrated vertical model for the same machine operation, now using both sensor and virtual sensor data.
- Compare the contribution of virtual sensor data against a data-driven model without this.
- Repeat the same integration for a second operation, and optionally a third operation.
- Assess integration results, comparing the two integrations. If the contribution from using virtual sensor data is low, address the sources of uncertainty and rework.
- Reasons for low contribution may be:
 - o Model validity prevents useful physics information from being integrated well.
 - o Data-driven model layout is not suitable.
 - o Data preparation is not sufficient.
 - o Other.

5. Conclusions

- We have defined an approach for how to integrate multiple types of physics-based models with sensor data about the route operations.
- Physics-based models are used for specifically chosen operations of the routes.
- The effect of adding knowledge from physics-based models can be evaluated.
- The approach is meant to integrate data from a certain period of production. For offline models integrating historical data, virtual sensor information computation time is of less importance. For online calculation of virtual sensor data, timeliness of physical models for calculating virtual sensor data is a concern that needs to be solved by, for instance, using surrogate models instead.