D5.7 – Data-Driven Continuous Process Qualification, v2

WP5 – BUILD: Rapid Manufacturing Line Qualification and Reconfiguration

i4Q has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958205.

Document Information

GRANT AGREEMENT NUMBER	958205	ACRONYM	i4Q	
FULL TITLE	Industrial Data Services for Quality Control in Smart Manufacturing			
START DATE	01-01-2021	DURATION	36 months	
PROJECT URL	https://www.i4q-project.eu/			
DELIVERABLE	D5.7 – Data-Driven Continuous Process Qualification v2			
WORK PACKAGE	WP5 – BUILD: Rapid Manufacturing Line Qualification and Reconfiguration			
DATE OF DELIVERY	CONTRACTUAL	31-Dec-2022	ACTUAL	30-Dec-2022
NATURE	Report	DISSEMINATION LEVEL	Public	
LEAD BENEFICIARY	TUB			
RESPONSIBLE AUTHOR	Jan Mayer (TUB)			
CONTRIBUTIONS FROM	7-IKER, 9-UPV			
TARGET AUDIENCE	1) i4Q Project partners; 2) industrial community; 3) other H2020 funded projects; 4) scientific community			
DELIVERABLE CONTEXT/ DEPENDENCIES	This document explains the development of the i4Q Data-driven Continuous Process Qualification (i4Q^{PQ}) solution up to month 24, which corresponds to Release #2 of the software. Its relationship to other documents is as follows: For constantly evaluating requirements which were elicited in T1.4, information about the user needs is taken from D1.9 Requirements Analysis and Functional Specification and assessed by the current development status. To ensure the connection to appropriate company units, i4Q^{PQ} is being mapped to the Reference Architecture coming from D2.7 i4Q Reference Architecture and Viewpoints Analysis.			
EXTERNAL ANNEXES/ SUPPORTING DOCUMENTS	None			
READING NOTES	None			

ABSTRACT	This document is a Technical Specification document about the development of the i4Q Data-Driven Continuous Process Qualification until M24. This document provides a thorough description and analysis of the functionalities, features, and the current implementation status. Furthermore, an in-depth technical overview of the principal functional sub-components (i.e., features) of the solution is described.

Document History

VERSION	ISSUE DATE	STAGE	DESCRIPTION	CONTRIBUTOR
0.1	07-Nov-2022	ToC	Create ToC and Section Description	TUB
0.2	20-Nov-2022	Draft	Insert Section 1 & 2	TUB
0.3	23-Nov-2022	Draft	Insert Section 3 & 4	TUB
0.4	25-Nov-2022	Draft	Insert Conclusion and Executive Summary – available for internal review	TUB
0.5	02-Dec-2022	Internal Review	Internal Review	UNINOVA, BIESSE
0.6	09-Dec-2022	Draft	Addressing comments from the reviewers	TUB
1.0	30-Dec-2022	Final Document	Final quality check and issue of final document	CERTH

Disclaimer

Copyright message

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS/ACRONYMS

API	Application Programming Interface
AutoML	Automated machine Learning
Cpk	Capability Index for Process Performance
DB	Database
Dppm	Defective Parts Per Million
GUI	Guided User Interface
HTTP	HyperText Transfer Protocol
i4Q	Industrial data services for Quality Control in Smart Manufacturing
LCL	Lower Critical Limit
LTL	Lower Tolerance Limit
PQ	Data-Driven Process Qualification
SPC	Statistical Process Control
UCL	Upper Critical Limit
UI	User Interface
UTL	Upper Tolerance Limit

Executive summary

This document provides compressed information about the development tasks up to M24 about the i4Q Data-Driven Process Qualification (i4Q^{PQ}). The main sections present the general description, the technical specifications, implementation status and the conclusion. This is an updated deliverable with respect to its previous version D5.1 Data-Driven Continuous Process Qualification.

i4Q^{PQ} is a user-friendly forecasting and real-time monitoring tool to describe the current behavior of single quality measurements of manufactured products. It provides automated machine learning (AutoML) capabilities and visualizations for direct interpretation of the analysis output. The containment of a static and dynamic data connection facilitates the implementation to existing storage technologies. The graphical user interface (GUI) guides through all necessary steps which end up with the analysis results containing current and future process qualification, current distribution details and the defective parts per million (dppm) of the current process.

The Source Code of the i4Q^{PQ} Solution that is in a private repository of Gitlab: https://gitlab.com/i4q/PQ.

This document **i4Q** D5.7 v2 is an update of v1 of D5.7., for this reason it contains information of the 1st version together with the updates developed in this 2nd version.

Document structure

Section 1: Contains a general description of the i4Q **Data-Driven Continuous Process Qualification**, providing an overview and the list of features. It is addressed to final users of the i4Q Solution.

Section 2: Contains the technical specifications of the i4Q **Data-Driven Continuous Process Qualification**, providing an overview and its architecture diagram. It is addressed to software developers.

Section 3: Details the implementation status of the i4Q **Data-Driven Continuous Process Qualification** explaining the current status, next steps and summarizing the implementation history.

Section 4: Provides the conclusion.

APPENDIX I: Provides the PDF version of the i4Q **Data-Driven Continuous Process Qualification** web documentation, which can be accessed online at: http://i4q.upv.es/17_i4Q_PQ/index.html

1. General Description

This section provides general background knowledge about the software solution i4Q^{PQ}. Furthermore, it describes the main features and their functionality and maintains the close alignment to the user requirements.

1.1 Overview

Information about process qualification is an established procedure to describe the current and former status of products which have been manufactured within a given quality range. This procedure is used when evaluating a continuous quality measurement, e.g. length, diameter, roughness. The quality range describes the acceptance thresholds of inner and outer quality. Using descriptive statistics enables the assumption to a specific tolerance value, if a scaled value of the current parameters is within or outside the quality range. Therefore, around 130 manufactured products are recommended to evaluate. A conversion can be done to describe this calculation into a single value, called the C_{pk} value. If this value is above 1.33, no process interventions are needed. The industrial term to this practice is Statistical Process Control (SPC).

i4Q^{PQ} applies this procedure with novel technologies coming from AutoML and real-time monitoring. The combination of innovative software and established manufacturing calculations builds up on implicit process knowledge of the quality engineer. Thus, the application points out improvements of interpretation and adapts to multiple environments. Therefore, the data connection can be chosen in static and dynamic format, representing different technologies. To enlarge the potential of AutoML forecasting, parameters of the algorithm like the variable of interest and the forecasting window can be adjusted separately. As visual interpretation support, both data connections provide a process chart and histogram. In addition, the histogram describes the current dppm and implies scale and location of the process distribution. The output of the forecast is visualized with traffic lights colors containing possible alerts of the converted C_{pk} value.

1.2 Features

The main features which are contained in the software can be seen as functionality of i4Q^{PQ} which are executed during production:

- **GUI:** Provides guidance for the novel approach to forecast the process qualification and acts as main integration part for other software. Furthermore, it is also extendable if other software is demanded.
- **Data Connection:** Provides an interface to popular storage systems like MongoDB, HTTP or csv-files.
- **Visualizations:** Provides two established visualizations (Process Chart and Histogram) to find observations outside the quality range and interpret statistical measurements across the complete range of the scaled distribution.
- **Distribution interpretation:** Provides information about the current distribution and the dppm.
- **Forecast:** Provides AutoML capabilities and individual adjustments to forecast the Process Qualification and calculate the C_{pk} value.

2. Technical Specifications

To describe the utilization of i4Q^{PQ} regarding its technical specification, this section is providing information about how i4Q^{PQ} is structured and what distinct technologies were used. Specifically, the interaction to other company areas, the guided interaction with users and the theoretical background behind the statistical analysis and the AutoML-Pipeline is examined. Regarding the application of the software in a manufacturing environment, provisional technologies for the easy installation is provided.

2.1 Overview

Certain technology and human expertise must be used to link processes for continuous monitoring. In terms of technical implementation, these enabling technologies are necessary tools for i4Q^{PQ} to add value. Because real-time technologies aren't always used in prospective industrial use cases, processing data from numerous sources has extraordinary benefits for the versatile deployment of i4Q^{PQ}. Additionally, integrated data management applications are included in the backend functionality to save and load the developed AutoML pipeline, for example. However, using streaming technologies is a must for deployment on the edge or in another environment, and accurate data must be inserted in order to evaluate the analytic results. The human collaboration is essential for the practical outcome because the final decision is delegated to process owners.

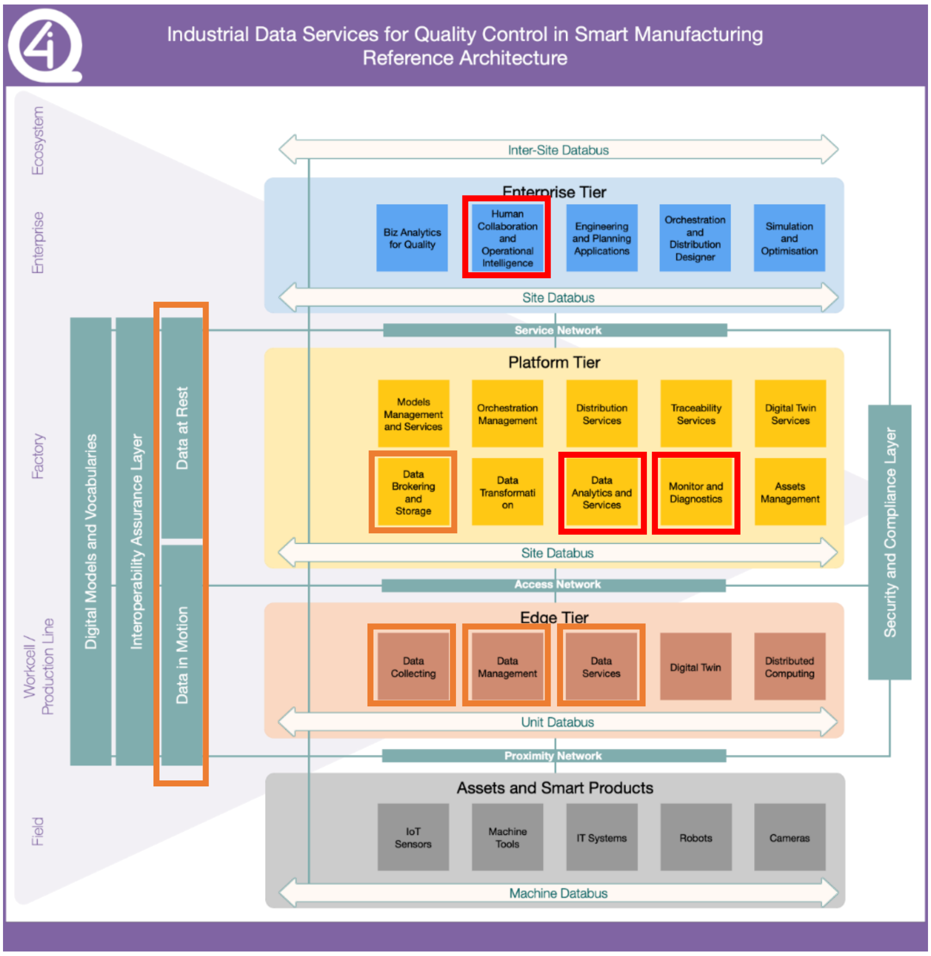


Figure 1. Mapping of elements to Reference Architecture of i4Q^{PQ}

By fusing statistical methods with cutting-edge technologies, i4Q[PQ] can be described as an enabling tool for process owners and manufacturing. In terms of process management, this adds value, but it also places dependencies on the infrastructure of the relevant information systems. In conclusion, a number of the reference architecture's components are involved both as providers and as prerequisites. Since some of the domain descriptions were already made in T5.1, only new developments are included in this section:

- **Platform Tier:** A special connection needs to be made since Apache Kafka, the data streaming technology it contains, is a requirement for real-time monitoring. The backend of i4QPQ features prospect reading and processing. This link enables the AutoML pipeline to execute customized forecasts as a service to track the procedure based on the desired variable. In addition, the analytics regarding the process distribution is done via this Tier.
- **Network:** The reliance on the edge also has an immediate effect on transferred data that is in motion or at rest. The end users can select either connection, but they both rely on the efficient data storage and brokering mechanism.

2.2 Interpretation and Results for Industrial Application of Pilot 6 (FARP)

As an example, a continuous variable of Pilot 6 is evaluated.

The first step would be to implement the specification/tolerance limits. In this case, it would be a quality range between 40 as lower tolerance limit (LTL) and 110 as upper tolerance limit (UTL). The basic unit is not required since the interpretation relies on the implicit process knowledge. After the selection of the variable of interest and the forecast window, here 30 time units, the checkbox “Create Process Qualification Information with the above selected criteria” is the trigger to create the visuals and process qualification forecast, as well as the C_{pk} values.

The first result displays a process chart to see the last observations and if they were outside the quality range. The grey line represents former (real) observations, the purple line is the forecast which was created by the AutoML algorithm.

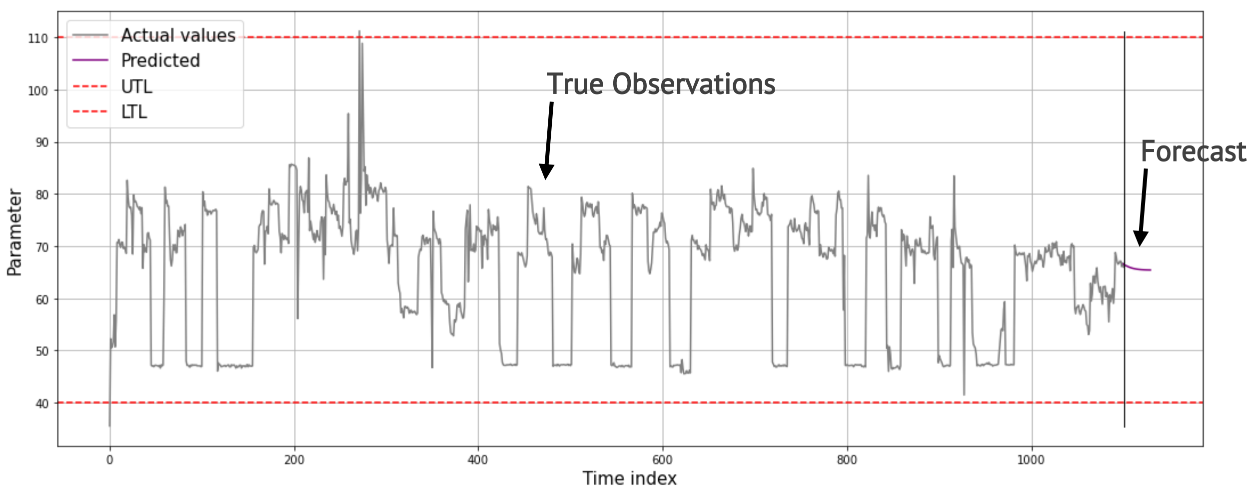


Figure 2. Time Series Forecasting for i4Q[PQ]

The output of the statistical parameters would be the following:

Current C_{pk} value:

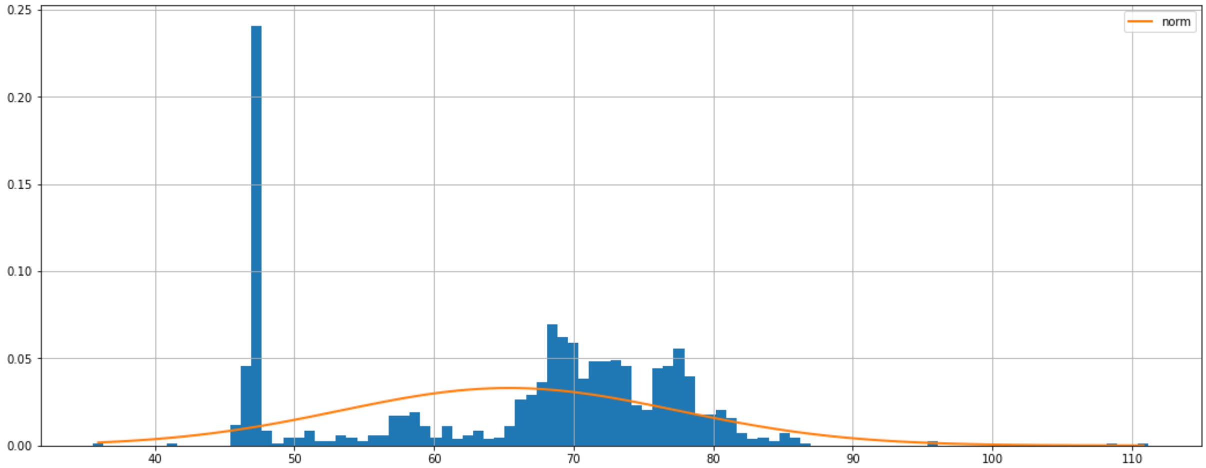
```
Measurement_mean    63.519923
Measurement_std      6.364902
Measurement_cpk      1.231751
```

Predicted C_{pk} value:

```
Measurement_mean    64.667847
Measurement_std      4.100199
Measurement_cpk      2.005419
```

The interpretation would be to leave the process as is without any intervention since the Cpk value is around 1.33 and increasing. The standard deviation describes the deviation around the mean for around 66% of all produced products. Since this number is decreasing, the variance is more concentrating around the mean for the next 30 produced products. Thus, no other intervention is required.

The second result would be to check how the process is distributed. This supports to interpret if some observations are made in a higher frequency than others. In this case, the statistical test assumes a normal distribution, however there is a concentration of the observations in the lower interval of the quality range. Thus, the process might be in risk to produce in a specific interval. This result is dependent on the implicit process knowledge of quality engineers in the company. Keeping the process on this level would represent an estimated dppm value of 18'041.




```
The process you are evaluating is currently normal distributed.
This assumes a natural variance in the process with data equally distributed around the mean.
Your current amount of defective parts per million is 18041.
```

Figure 3. Process Distribution and Interpretation of Results

3. Implementation Status

Pretesting i4Q^{PQ} is essential for the software application's implementation because i4Q^{PQ} deployment depends heavily on the test phase of the i4Q project.

3.1 Current implementation connected to user requirements

To support the persistent arrangement of i4Q^{PQ} with user requirements coming from the industrial partners of the venture, **Table 1** presents the current organize of the improvement per Pilot. The advancement can be compared to the latest development status of the last reporting phase.

Pilot	Requirement	Development Stage M18	Further Development	Development Stage M24
Biesse	Perform an uncertainty analysis with the aim to define the reliability of the developed algorithms	66%	Retraining phase will be included in the backend to check if the algorithm can be improved.	100%
Biesse	Scalability of the data and/or thresholds when change the machine model or the boundary conditions of the test	100%	NA	100%
Factor	Predict quality measurement	100%	NA	100%
Factor	Check if the problem has been solved	0%	Retraining phase will be included in the backend to check if the algorithm can be improved. Interpretation guidelines will be included in the interface	100%
Factor	Ensure that the process is being optimized	100%	NA	100%
Riastone	Quality check on the analysis of raw matter's composition and granulometry	100%	NA	100%
Farplas	Acceptance tests for initial production of the injection machine	50%	Cpk-calculation is still pending since data connection logic	100%

Pilot	Requirement	Development Stage M18	Further Development	Development Stage M24
			needs to be implemented beforehand.	
Farplas	Automated defect detection	0%	This requirement is in discussion with i4Q[QD] as the detection part is not included in i4Q[PQ].	100%
Overall	**8 Requirements**	**71%**		**100%**

Table 1. Development progress between M18 and M24 concerning User Requirements

3.2 New improvements

Similar to the above-mentioned user requirements, the improvements can be tracked with a development table, which has been introduced in the last deliverable.

Points which have been not implemented, but are additional to the overall effectiveness and functionality are the following:

- **Implement retraining logic:** The automated retraining logic has been examined by IKER to check if the change of process distributions can have enough impact to bias the model output and thus the result of the forecast. Due to the non-investigated correlations between different underlying distributions and the model output, the retraining won't be part of the solution in M24. However, there is existing code which can be implemented if needed.
- **Implement connection to Min.io cloud:** Since the Min.io cloud connection must be established once the software is in production, the tool will be in charge once the i4Q[PQ] goes live.
- **Create Docker Compose File for deployment routine:** As the evaluation phase is providing valuable feedback about prerequisites of the installation routine, the docker file will be updated, once adjustments about robustness have been made.

Topic	Development Task	Fulfilment M18	Further Development	Fulfilment M24
UI	Create multi-layered app framework	100%	NA	100%
Data Connectors	Implement connection logic to sidebar for static data connection.	100%	NA	100%

Topic	Development Task	Fulfilment M18	Further Development	Fulfilment M24
Data Connectors	Implement connection logic to sidebar for dynamic data connection.	0%	Inserting text boxes for creating individual data stream connection	100%
SPC	Implement Cpk-value traffic light behaviour with current and future Process Qualification	0%	Calculate Cpk-value, catch forecast values, implement in Process Qualification	100%
SPC	Implement histogram	0%	Insert plots for both dynamic and static connection	100%
SPC	Implement process chart	0%	Insert plots for both dynamic and static connection	100%
SPC	Implement UTL and LTL to individualization logic	0%	Create more text boxes for the input of UCL and LCL	100%
SPC	Implement distribution fit	0%	Implement function to define current process distribution	100%
AutoML	Implement pipeline logic	100%	NA	100%
AutoML	Implement individualization of pipeline	100%	NA	100%
AutoML	Implement retraining logic	0%	Implement behaviour to continuously evaluate algorithm	0%
AutoML	Implement Line Chart for future process quality	0%	Add another line chart to visualize future behaviour of the process	100%
AutoML	Implement connection to Min.io cloud	0%	Add connection via API calls to Min.io	0%

Topic	Development Task	Fulfilment M18	Further Development	Fulfilment M24
Deployment	Create Docker Compose File for deployment routine	50%	Reuse Docker Compose from i4Q^{QD}	80%

Table 2. Development progress between M18 and M24 concerning self defined programming tasks

3.3 History

Version	Release date	New features
V0.1.0	01.04.2022	Framework of UI
V0.2.1	05.04.2022	Static Data Connection
V0.3.1	29.04.2022	AutoML Pipeline
V0.3.2	05.05.2022	AutoML Pipeline Individualization
V0.4.1	05.08.2022	Dynamic Data Connection (Kafka Integration)
V0.4.2	01.10.2022	Process Distribution Interpreter
V0.4.3	15.11.2022	Docker preparation
V1.0.0	30.12.2022	Final Version

Table 3. Version History

4. Conclusions

Deliverable D5.7 Data-Driven Continuous Process Qualification v2 contains a compressed and detailed description of the software tools and improvements in regard to the first reporting period in M18. Furthermore, it provides an in-depth specialized outline of the i4Q^{PQ} arrangement. This report depicts the advancement work which has been done concerning both conceptual as well as program advancement work. The main aspect of PQ lays in the change of the application of SPC in fabricating forms. Furthermore, it classifies the novel process qualification to particular company domains which stand as enablers regarding the application and continuous monitoring aspect of the software.

The main features are aligned to the required needs from industrial partner and the postulated information of the i4Q Description of Action, where PQ is responsible regarding the innovation action of the SPC. This process is utilized by i4Q^{PQ} in conjunction with cutting-edge AutoML technologies and real-time monitoring. Innovative software combined with tried-and-true production calculations expands on the quality engineer's implicit process knowledge. As a result, the application identifies improvements in interpretation and can be adjusted for various environments. As a result, different technologies can be represented by the static and dynamic formats of the data connection. The forecasting window and the variable of interest can both be changed independently to increase the potential of AutoML forecasting. Both data connections provide a process chart and a histogram to aid in visual comprehension. The histogram also indicates the scale and position of the process distribution and explains the current dppm.

Appendix I

***i4Q* Data-driven continuous process Qualification** *web documentation can be accessed online at: https://i4q.upv.es/17_i4Q_PQ/index.html*