Probabilistic Forecasting of Current Harmonic Distortions in Distribution Systems

Antonio Bracale, Pierluigi Caramia, Pasquale De Falco
University of Naples Parthenope
Naples, Italy
antonio.bracale@uniparthenope.it
pierluigi.caramia@uniparthenope.it
pasquale.defalco@uniparthenope.it

Max Domagk, Jan Meyer
Techniche Universitaet Dresden
Dresden, Germany
max.domagk@tu-dresden.de
jan.meyer@tu-dresden.de

***Abstract*—Power Quality (PQ) disturbances have an impact on electrical installations and equipment. Waveform distortions like harmonics create additional losses and reduce the efficiency of the power converter controls, and therefore should be kept below admissible limits to cope with Standards and regulatory schemes. Relying on accurate predictions of the future values of harmonics can help distribution system operators and clients in taking timely actions to comply with the admissible limits. This paper provides probabilistic methodologies to forecast current harmonics in short-term horizons. The methodologies are based on Quantile Regression (QR) models and are diversified based on the forecasting task: disaggregated forecasts of harmonics at each time interval are provided through a direct approach while daily or weekly percentile aggregated forecasts are provided through either a direct or an indirect approach. The proposed methodologies are applied on actual data collected from Low-Voltage (LV) installations to evaluate their predictive performance. The methodologies improve the performance from 22% to 43%, compared to a naïve persistence benchmark.**

Keywords—harmonics, current distortion, power quality, time series analysis, probabilistic forecasting

I. INTRODUCTION

Power Quality (PQ) disturbances are electromagnetic phenomena that divert the voltage and current waveforms from their ideal sinusoidal pattern and influence the operation of sensitive electrical installations and equipment. The management and containment of disturbances is responsibility of both distribution system operators and clients, and Standards suggest some limits on many PQ parameters to guarantee the optimal operation of the systems [1]. Limits are recommended, e.g., on voltage and current harmonics and on their Total Harmonic Distortion (THD) [2,3], at different aggregation levels and for different time horizons, suggesting taking timely actions to modify the characteristics of the systems to return below the recommended limits [4].

The current practice is to take actions after observing that the harmonic limits have been exceeded. Nevertheless, making decisions based on a predictive evaluation of the PQ levels [5] can significantly ease the job of distribution system operators and clients, by arranging countermeasures (e.g., shift some loads, change the network configuration, increase short-circuit power) before exceeding the limit. In addition, relying on accurate forecasts of harmonic values can help to reduce losses in scheduling and optimal operation of systems, not only addressing distorting loads but also considering the impact of distorting distributed energy generation systems [6].

Probabilistic forecasts can fully address the uncertainty linked to the source and to the propagation of harmonic emissions and can suit the probabilistic assessment recommended by the Standards. However, the PQ parameter forecasting tasks are only examined by a few studies, mostly in a deterministic framework [7-11] and even more rarely in a probabilistic framework [12]. While most of the effort is put into the forecast of the overall PQ indices, such as the THD, there is a lack of consolidated methodologies to predict the PQ parameters at different levels of aggregation, for different time horizons and resolutions, and to get predictions comparable to the recommended harmonic limits for decision-making stages.

This paper contributes to fill this gap by framing the forecasting task based on the time resolution of the target PQ parameter and based on the aggregation scale for the limit comparison. Different methodologies based on Quantile Regression (QR) models are proposed for the disaggregated and aggregated task: disaggregated forecasts at each time interval (e.g., 10 minutes) are obtained through a direct approach, while daily or weekly percentile aggregated forecasts are obtained through either a direct or an indirect approach. In the latter case, a dedicated percentile estimation procedure is proposed to overcome problems linked to the unfeasibility of the statistical integration that is required to get probabilistic forecasts of a percentile from the disaggregated probabilistic forecasts. The proposed methodologies are tested on actual data collected at public Low-Voltage (LV) distribution networks to assess the performance.

This paper begins with a description of the considered data and with some examples of exploratory data analysis in Section II and shows the proposed forecasting methodology in Section III. Section IV presents the indices for the assessment of the forecasting accuracy. Section V shows the results of the method application. The paper is concluded in Section VI.

II. BACKGROUND: DATA AND EXPLORATORY ANALYSIS

PQ parameter forecasting tasks are framed by the availability of historical monitored variables (e.g., current or voltage waveform points, or PQ indices like individual harmonics estimated by Fourier transforms, or the THD), and by the time resolution/aggregation of measurements. The "PQ nomenclature" [2-4] classifies measurements as "very-short-time" (related to a three-second interval) and "short-time" (related to a ten-minute interval). This classification is extended also to the aggregation scale considered for assessing the harmonic limits: limits on the daily 99th percentile values suggested for very-short-time measurements (3 s), and limits on the weekly 95th and 99th percentile values suggested for short-time measurements (10 min). A similar wording, but with a different interpretation related to the prediction lead time, is considered to classify predictions in the "forecasting nomenclature" [13]: "short-term" forecasts are up to two weeks ahead, "medium-term" forecasts are up to three years ahead, and "long-term" forecasts are beyond three years ahead. Words are similar so, to avoid misinterpretation between the PQ and the forecasting nomenclatures, below we refer to "disaggregated forecasting" when considering the short-term forecast of the disaggregated (either very-short-time (3 s) or short-time (10 min)) harmonic values, and to "percentile aggregated forecasting" when considering the short-term forecast of aggregated daily or weekly percentiles.

A. Data Description

The measurement data includes different public LV networks in Germany. The harmonic current emission mainly depends on the number and the type of electronic equipment, which is determined by the type of connected consumers. Different uniform consumer configurations, such as residential areas with single- or multi-family houses, office areas, commercial areas and mixed areas have been distinguished in the site selection. Two sites (INST1: a commercial area with an electronic market; INST2: an office area) with measurement durations of four years from 2012 to 2016 have been chosen for the experimental application within this paper. Fixed PQ monitors complying with IEC 61000-4-30 Class A have been used for the measurements. PQ parameters monitored include the fundamental and harmonics of orders 3, 5 and 11 on the three lines of installations. Typically, the harmonic current decreases in magnitude with increasing order, which makes a reliable evaluation difficult due to the accuracy of the measurement equipment.

B. Exploratory Analysis of the Target Variables

The model development stage of a forecasting task consists of determining the best combination of predictors to explain the future values of the target variable y. The predictors that can properly explain the target variable are often searched among lagged values of the target variable (i.e., endogenous predictors) and among values of other relevant variables (i.e., exogenous predictors). The number of possible combinations grows exponentially with the number of model predictors (e.g., 2^m permutations on m predictors), and it is unreasonable to exhaustively compare all of them. An exploratory analysis reduces the dimension of the problem by evaluating the relevancy of predictors for the target variable and by discarding the uninformative predictors.

The exploratory data analysis is presented as an example of the model development to predict a 3rd current harmonic, with several other current harmonics available as exogenous predictors. Applying the exploratory analysis to other target variables and/or with other available exogenous predictors is similar. For a quantitative evaluation, Fig. 1 illustrates the autocorrelation of the 3rd current harmonic samples up to one-week lag. This is a known shape occurring for time series with daily and weekly seasonality, that can be fruitfully exploited in the model development by adding endogenous predictors (i.e., target variable with one-day and one-week lags) [14].

For a qualitative evaluation, Fig. 2 illustrates 3rd current harmonic samples against 3rd voltage samples (Fig. 2a) and against 3rd voltage samples lagged by one week (Fig. 2b).

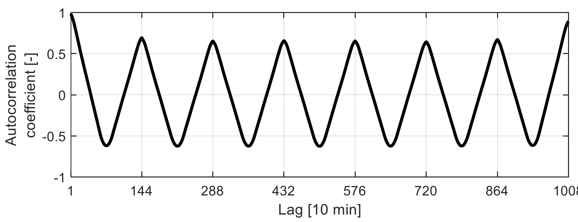

Fig. 1. Autocorrelation of the 3rd current harmonic up to one-week lag.

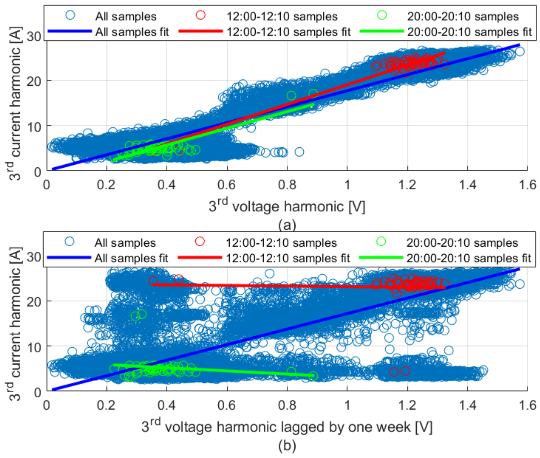

Fig. 2. Scatter plots and best fits of 3rd current harmonic against 3rd voltage harmonic (a) and against 3rd voltage harmonic lagged by one week (b).

The scatter plots and their best fitting lines confirm the existing theoretical relationship between the two variables without lagging, while this relationship is less clear if the one-week lag is applied to the voltage harmonic. Indeed, while the linear fits of the overall samples remain explicative in both figures and the slopes of the fitting lines are very close (17.72 and 17.19), a discrimination among samples referred to different ten-minute intervals of the day highlights the presence of clusters. The best fitting lines for samples at noon or at 20:00 become practically horizontal when considering the lagged 3rd voltage harmonics, suggesting almost no linear relationship between this predictor and the 3rd current harmonic, while the relationship is clear (with substantial slopes of the fitting lines, 21.80 and 18.43) when considering the unlagged 3rd voltage harmonic. This suggests that including the 3rd voltage harmonic lagged by one week as a candidate exogenous predictor in the model development could be useful to forecast the 3rd current harmonic, but only if considering all the samples together. If the model instead discriminates by the interval of the day (e.g., through dummy variables) [12], that predictor can be safely discarded.

III. Probabilistic Forecasting Methodologies for Power Quality Parameters

The proposed probabilistic methodologies to forecast the disaggregated or aggregated PQ parameters are based on underlying QR models that link the target variable to endogenous and exogenous predictors. We hereby indicate the variables in disaggregated (aggregated) forecasting with lower-case (upper-case) letters e.g., the target variable is y (Y in aggregated forecasting). The available PQ variables are initially reduced due to the exploratory data analysis described in Section II.B, keeping only the most informative predictors. The resulting m candidate predictors in disaggregated forecasting are $x_1, x_2, \dots, x_m$ and the M candidate predictors in aggregated forecasting are $X_1, X_2, \dots, X_M$.

A direct approach is proposed to build disaggregated forecasts: the PQ inputs are exploited in a developed QR model to obtain the probabilistic prediction $\hat{y}_{t+k}^{\langle\alpha_1\rangle}, \dots, \hat{y}_{t+k}^{\langle\alpha_Q\rangle}$ issued at the origin time t with lead time k. A direct and an indirect approach are proposed to build percentile aggregated forecasts. In the direct approach, the PQ inputs are processed to get the historical time series of the percentile aggregated variable, and they are exploited in a developed QR model to obtain the probabilistic prediction $\hat{Y}_{p,T+K}^{\langle\alpha_1\rangle}, \dots, \hat{Y}_{p,T+K}^{\langle\alpha_Q\rangle}$ of the p^{th}

percentile of the target variable issued at the origin time T with lead time K. In the indirect approach, intermediate disaggregated forecasts are generated through the whole period of aggregation (i.e., $\hat{y}_{t+k}^{\langle\alpha_1\rangle}, \ldots, \hat{y}_{t+k}^{\langle\alpha_Q\rangle}$ for $k = 1,2, \ldots, n_K$, with origin time t coherent with the aggregated origin time T, and with n_K sufficiently high to cover the aggregated lead time K). Then, a percentile estimation procedure is applied to the intermediate forecasts to get the probabilistic prediction $\hat{Y}_{p,T+K}^{\langle\alpha_1\rangle}, \ldots, \hat{Y}_{p,T+K}^{\langle\alpha_Q\rangle}$ of the p^{th} percentile of the target variable.

Note that time symbols are diversified for aggregated and disaggregated forecasting due to their different resolution (i.e., 3 s/10 min for disaggregated, 1 day/1 week for aggregated), and that all forecasts consist of Q quantiles at coverages $0 \leq \alpha_1 < \cdots < \alpha_Q \leq 1$. The methodologies below are presented focusing, for brevity, on predicting current harmonic distortions, but they can be extended to other PQ parameters.

A. Quantile Regression Model

A QR model links the predictive α_q-quantile $\hat{y}_{t+k}^{\langle\alpha_q\rangle}$ of the target variable y, issued at the origin time t with a lead time k, to the m^* predictors $x_{1_{t+k}}, x_{2_{t+k}}, \ldots, x_{m^*_{t+k}}$ as [15]:

$$\hat{y}_{t+k}^{\langle\alpha_q\rangle} = \hat{\beta}_0^{\langle\alpha_q\rangle} + \hat{\beta}_1^{\langle\alpha_q\rangle} x_{1_{t+k}} + \cdots + \hat{\beta}_{m^*}^{\langle\alpha_q\rangle} x_{m^*_{t+k}}, \quad (1)$$

where $\hat{\beta}_0^{\langle\alpha_q\rangle}, \hat{\beta}_1^{\langle\alpha_q\rangle}, \ldots, \hat{\beta}_{m^*}^{\langle\alpha_q\rangle}$ are $m^* + 1$ parameters estimated by minimizing the Quantile Score (QS) on a training set ω_{tr}. For an individual prediction, the QS is formulated as [16]:

$$QS_{t+k}^{\langle\alpha_q\rangle} = \begin{cases} \alpha_q \left(y_{t+k} - \hat{y}_{t+k}^{\langle\alpha_q\rangle} \right) & \text{if } y_{t+k} > \hat{y}_{t+k}^{\langle\alpha_q\rangle} \\ \left(1 - \alpha_q\right) \left(\hat{y}_{t+k}^{\langle\alpha_q\rangle} - y_{t+k} \right) & \text{if } y_{t+k} \leq \hat{y}_{t+k}^{\langle\alpha_q\rangle}, \end{cases} \quad (2)$$

where y_{t+k} is the true value of the target variable at the forecast horizon $t + k$. If m candidate predictors result from the exploratory data analysis, the QR model is developed by training some (g) models with different combinations of predictors and picking the final model with $m^* \leq m$ predictors as the one with the smallest QS in a validation set ω_{val}. All the above can be particularized for the aggregate forecasting task by replacing lower- with upper-case symbols.

B. Disaggregated Forecasting

Here, disaggregated forecasting refers to the short-term forecast of the disaggregated (either 3 s or 10 min) harmonic values. In the proposed direct (Fig. 3), the PQ inputs are the harmonic currents and voltages and the total current harmonic distortion on all the lines of the installation (e.g., $\boldsymbol{i}_{L_1}, \boldsymbol{i}_{L_2}, \boldsymbol{i}_{L_3}$ are the three datasets for a three-phase installation, with $\boldsymbol{i}_{L_1} = \{\boldsymbol{i}_{L_1|1}, \boldsymbol{i}_{L_1|2}, \ldots, \boldsymbol{i}_{L_1|h}, \ldots, \boldsymbol{i}_{L_1|40}\}$ collecting the forty harmonics in a 50 Hz installation, and with the generic h^{th} harmonic time series being a set of n points collected until the origin time t, i.e., $\boldsymbol{i}_{L_1|h} = \{i_{L_1|h_t}, i_{L_1|h_{t-1}}, \ldots i_{L_1|h_{t-n+1}}\}$. The time resolution of the inputs must match the disaggregated forecast resolution, e.g., 10 mins to predict short-time harmonics.

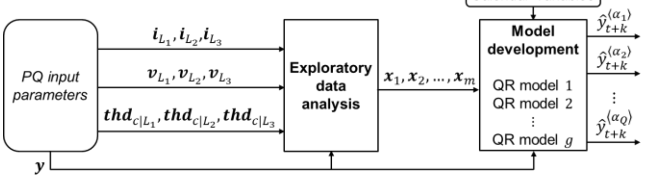

Fig. 3. Flowchart of disaggregated harmonic forecasting.

The m candidate predictors available after the exploratory data analysis are sets of n points, and the generic i^{th} predictor set is $\boldsymbol{x}_i = \{x_{i_t}, x_{i_{t-1}}, \ldots, x_{i_{t-n+1}}\}$. In the model development stage, the candidate predictors and calendar variables to account for the seasonality of the data are shuffled and combined in g different, individually trained QR models. The final picked model is the one among them with the lowest QS on the validation set ω_{val}, and it is used to eventually generate the forecasts $\hat{y}_{t+k}^{\langle\alpha_1\rangle}, \ldots, \hat{y}_{t+k}^{\langle\alpha_Q\rangle}$. The model development is a challenging aspect of disaggregated forecasting, as it is computationally intensive and the exhaustive training and validation of all the possible combinations of predictors is typically infeasible for large m. Moreover, a different model could be theoretically developed for each quantile coverage α_q and for each lead time k. In a real-world case where disaggregated 10-min forecasts are issued one-week-ahead, the forecast horizon consists of 6x24x7=1008 intervals, with Q=99 different models (one for each quantile coverage from 0.01 to 0.99). Developing 1008x99=99792 models is unfeasible and not beneficial, also given the resulting loss of interpretability. To minimize the challenge yet guaranteeing sufficient accuracy, a single model (i.e., a fixed combination of predictors) is developed and applied for multiple lead times (e.g., the 1008 intervals for one-week-ahead disaggregated short-time forecasting) and for all the Q quantile coverages. Note that, even if the combination of predictors is fixed, the parameters in (2) vary as α_q and k varies, so forecasts are diversified by quantile coverage and by lead time.

C. Percentile Aggregated Forecasting

Here, percentile aggregated forecasting, referred to the aggregated daily percentiles (of 3 s harmonic values) or weekly percentiles (of 10 min harmonic values), is solved with a direct or indirect approach. The direct approach for percentile aggregated forecasting is like the one presented for disaggregated forecasting, simply replacing lower- with upper-case letters. The indirect approach (Fig. 4) instead builds the probabilistic prediction $\hat{Y}_{p,T+K}^{\langle\alpha_1\rangle}, \ldots, \hat{Y}_{p,T+K}^{\langle\alpha_Q\rangle}$ of the target variable p^{th} percentile from intermediate disaggregated forecasts $\hat{y}_{t+k}^{\langle\alpha_1\rangle}, \ldots, \hat{y}_{t+k}^{\langle\alpha_Q\rangle}$, for $k = 1,2, \ldots, n_K$. The origin time t must be coherent with the aggregated origin time T, and n_K must be high enough to cover the aggregated lead time K (e.g., in weekly percentile forecasting, n_K=1008 covers the 6x24x7 ten-minute intervals in a week). The core of the indirect approach is the percentile estimation procedure applied on the n_K intermediate forecasts. An exact procedure can be formulated to extract the probabilistic forecasts of the p^{th} percentile of the target variable. Let's first consider that the predictive Cumulative Distribution Function (CDF) is linearly interpolated as in [17] from the Q predictive quantiles, for each disaggregated forecast lead time (i.e., for $k = 1,2, \ldots, n_K$). The matching n_K predictive Probability Density Functions (PDFs) are stepwise constant and can be viewed as PDFs $f_{Z_1}(z_1), \ldots, f_{Z_{n_K}}(z_{n_K})$ of the n_K random variables $Z_1, \ldots, Z_{n_K}$, i.e., the current harmonic in each of the n_K intervals of the aggregate lead time.

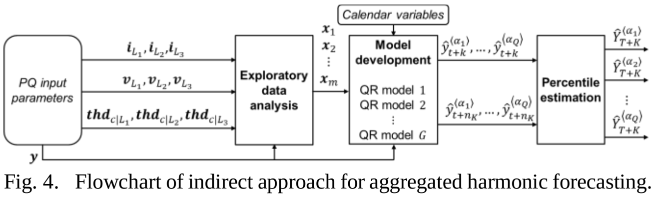

Fig. 4. Flowchart of indirect approach for aggregated harmonic forecasting.

For independent predictions, the joint distribution is:

$$f_{Z_1,\dots,Z_{n_K}}(z_1,\dots,z_{n_K}) = f_{Z_1}(z_1)\cdot\ldots\cdot f_{Z_{n_K}}(z_{n_K}). \quad (3)$$

The calculation of the p th percentile from the n_K realizations $z_1, z_2, \dots, z_{n_K}$ could be implicitly defined as the function $b(z_1, z_2, \dots, z_{n_K})$, and the realization r of the random variable R (i.e., the p^{th} percentile extraction) is:

$$r = b(z_1, z_2, \dots, z_{n_K}). \quad (4)$$

The predictive CDF of the random variable R is [18]:

$$F_R(r) = \int_{D_{b,r}} f_{Z_1,\dots,Z_{n_K}}(z_1,\dots,z_{n_K})\, dz_1 \dots dz_{n_K}, \quad (5)$$

where $D_{b,r} \subset \mathbb{R}^{n_K}$ is the integration domain:

$$D_{b,r} = \{z_1, z_2, \dots, z_{n_K} : b(z_1,\dots,z_{n_K}) \le r\}. \quad (6)$$

The integral (5) is hardly calculated due to the difficulties in representing the explicit form of function $b(z_1, z_2, \dots, z_{n_K})$ as it includes a sorting operation, and due to the large dimension of the domain $D_{b,r}$ (e.g., $n_K = 1008$ for a simple real-world case). The severe limitations of the exact procedure are overcome in this paper by an approximated procedure to estimate the p^{th} percentile. In the approximated procedure, the predictive α_q-quantile $\hat{Y}^{\langle\alpha_q\rangle}_{p,T+K}$ of the p^{th} percentile of the target variable is estimated as the p^{th} percentile of the set of n_K predictive α_q-quantiles $\hat{y}^{\langle\alpha_q\rangle}_{t+1}, \hat{y}^{\langle\alpha_q\rangle}_{t+2}, \dots, \hat{y}^{\langle\alpha_q\rangle}_{t+n_K}$. If S is the sorted set (in ascending order) obtained from $\hat{y}^{\langle\alpha_q\rangle}_{t+1}, \hat{y}^{\langle\alpha_q\rangle}_{t+2}, \dots, \hat{y}^{\langle\alpha_q\rangle}_{t+n_K}$ and its elements are $s^{\langle\alpha_q\rangle}_1 \le s^{\langle\alpha_q\rangle}_2 \le \dots \le s^{\langle\alpha_q\rangle}_{n_K}$, the forecast $\hat{Y}^{\langle\alpha_q\rangle}_{p,T+K}$ is [19]:

$$\hat{Y}^{\langle\alpha_q\rangle}_{p,T+K} = s^{\langle\alpha_q\rangle}_{\lfloor l_p \rfloor} + (l_p - \lfloor l_p \rfloor)\cdot(s^{\langle\alpha_q\rangle}_{\lceil l_p \rceil} - s^{\langle\alpha_q\rangle}_{\lfloor l_p \rfloor}), \quad (7)$$

where $l_p = 1 + (n_K - 1)\cdot p/100$, $\lfloor l_p \rfloor$ is the highest integer that is $\le l_p$, and $\lceil l_p \rceil$ is the lowest integer that is $\ge l_p$.

IV. Probabilistic Error Indices

Probabilistic forecasts are evaluated using a proper score to assess the overall performance of the predictions relative to the actual outcomes, and by considering their reliability (i.e., calibration) separately [20]. The QS in (2), calculated for multiple quantile coverages and forecast issues, is averaged over the Q quantiles and over the horizons of a test set ω_{test}:

$$QS = \frac{1}{N_{\omega_{test}}} \sum_{(t+k)\in\omega_{test}} \frac{1}{Q} \sum_{q=1}^{Q} QS^{\langle\alpha_q\rangle}_{t+k}, \quad (8)$$

where $N_{\omega_{test}}$ is the cardinality of the test set, and it is used to overall evaluate the skill of the forecasts. The QS is also reported in its normalized percentage version (NPQS), with the normalization operated to the maximum spread (max-min) of the considered target variable in the test set ω_{test}:

$$NPQS = 100 \frac{QS}{\max\limits_{(t+k)\in\omega_{test}}\{y_{t+k}\} - \min\limits_{(t+k)\in\omega_{test}}\{y_{t+k}\}}. \quad (9)$$

The forecast reliability, i.e., the consistency between the estimated coverage $\hat{\alpha}_q$ and the nominal coverage α_q of the predictive quantiles [20], is assessed by Absolute Coverage Error (ACE) [12,17] and reliability diagrams [20]. ACE is:

$$ACE^{\langle\alpha_q\rangle} = |\alpha_q - \hat{\alpha}_q| = = \left|\alpha_q - \frac{1}{N_{\omega_{test}}} \sum_{(t+k)\in\omega_{test}} 1\left[y_{t+k} \le \hat{y}^{\langle\alpha_q\rangle}_{t+k}\right]\right|, \quad (10)$$

where function $1[\cdot]$ is equal to 1 if the statement in the brackets is true, and 0 otherwise. The Percentage Absolute ACE ($AACE_{\%}$) is eventually used as the overall index:

$$AACE_{\%} = 100 \frac{1}{Q} \sum_{q=1}^{Q} ACE^{\langle\alpha_q\rangle}. \quad (11)$$

All the indices can be particularized for aggregate forecasting by replacing lower- with upper-case symbols. Reliability diagrams plot estimated coverages against nominal ones for a qualitative evaluation of their consistency [20].

V. Experimental Results

The proposed probabilistic forecasting systems are applied to the data described in Section II. They are evaluated based on the generated disaggregated forecasts of short-time (10 min) current harmonics and the generated forecasts of the weekly 95th percentile aggregated current harmonics. The data of installations INST1 and INST2 are split into:

- training set: 103 weeks (07.01.2012 to 28.12.2014);
- validation set: 27 weeks (29.12.2014 to 05.07.2015);
- test set: 26 weeks (06.07.2015 to 03.01.2016).

All the forecasts are issued one-week-ahead (i.e., $n_K = 1008$ disaggregated forecasts with $k = 1,2,\dots,1008$ and one aggregated forecast with $K = 1$), with predictive quantiles having $Q = 99$ nominal coverages from 0.01 to 0.99. The results are compared with a persistence method (PM), i.e., a naïve benchmark for the assessment of the performance [12].

A. Disaggregated Forecasting Results

The results of the disaggregated forecasts averaged on the three lines of installations are in Table I. The $AACE_{\%}$ is not reported for PM, as PM predictions are not diversified by quantile coverages. The QR model returns a better forecast than the PM, with improvements in the range ~22% (3rd harmonic INST1) to ~43% (fundamental INST2). On average, there is more improvement for the fundamental (~41%) than higher harmonic orders (~27%, 34% and 34% for 3rd, 5th, and 11th harmonics). Fig. 5 exemplarily shows the QR prediction of the L_1 5th current harmonic at INST1 in comparison to the actual values of two weeks through the test period.

TABLE I. Disaggregated Forecast Results in the Test Period

Site	Harmonic order	QS [A]		NPQS [%]		AACE% [%]
		QR	*PM*	*QR*	*PM*	*QR*
INST1	1st	16.62	26.91	1.92	3.11	4.38
	3rd	0.66	0.85	2.57	3.34	6.43
	5th	0.38	0.58	2.65	3.99	4.54
	11th	0.17	0.25	2.53	3.81	2.64
INST2	1st	7.17	12.58	1.67	2.93	2.55
	3rd	0.66	0.96	1.45	2.10	3.44
	5th	0.70	1.04	2.24	3.32	2.07
	11th	0.25	0.39	2.44	3.84	2.05

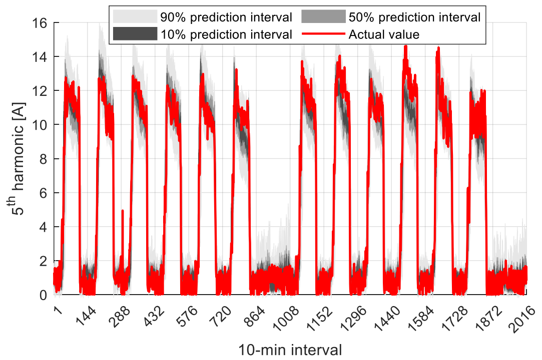

Fig. 5. QR prediction intervals of the L_1 5th current harmonic at the INST1 during two weeks of the test period.

The reliability diagrams of QR forecasts for the three phases of INST1 for the full test period are in Fig. 6. The consistency among nominal and estimated coverages is satisfied, but in some cases (e.g., L_1 and L_3 3rd current harmonics), the forecasts are slightly under-dispersed.

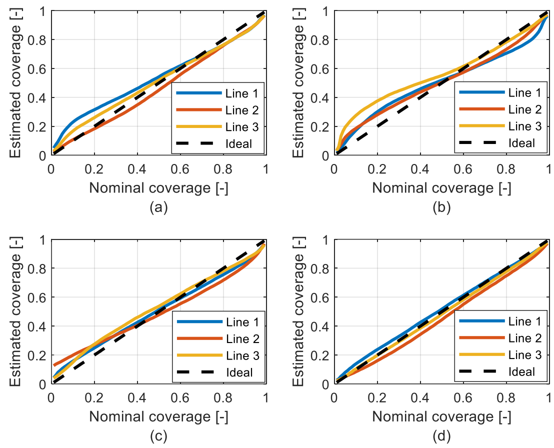

Fig. 6. QR reliability diagrams of the 1st (a), 3rd (b), 5th (c) and 11th (d) current harmonics at the INST1 during the test period.

B. Percentile Aggregated Forecasting Results

The overall results of the aggregated forecasts returned by the direct and indirect approaches, in terms of QS and NPQS indices, are displayed in Table II, where bold values indicate the best forecasts. The $AACE_{\%}$ is not reported in the Table, as the number of forecast issues (26, one for each week) is too small to consistently check the forecast reliability. The QR with direct approach is the best pick in thirteen over twenty-four cases, while the QR with indirect approach is the best pick in twelve cases. The QR always outperforms the PM benchmark. On average, the QR with indirect and direct approaches perform very similarly (~26% improvement over PM), although the indirect approach has the advantage of also providing intermediate disaggregated forecasts. On average, the improvement is higher for INST2 (~29% and ~33%) compared to INST1 (~23% and ~19%), and higher for the fundamental (~35% and ~37%) compared to higher harmonic orders (9% and 20%, 30% and 14%, 30% and 33% for 3rd, 5th, and 11th harmonics, respectively). With respect to the PM, the overall accuracy of the aggregated forecasts has less improvement than the disaggregated forecasts.

TABLE II. PERCENTILE AGGREGATED FORECAST RESULTS IN THE TEST PERIOD. BOLD VALUES INDICATE THE BEST FORECASTS

Site	Harmonic order	QS [A]			NPQS [%]		
		QR Indirect	*QR Direct*	*PM*	*QR Indirect*	*QR Direct*	*PM*
INST1	1st	22.57	**21.22**	33.75	5.19	**4.88**	7.77
	3rd	0.24	**0.22**	0.26	5.99	**5.31**	6.79
	5th	**0.31**	0.41	0.41	**6.37**	7.94	8.06
	11th	**0.09**	0.09	0.12	**6.97**	7.39	9.32
INST2	1st	**5.17**	5.19	8.11	**3.98**	4.00	6.20
	3rd	0.57	**0.47**	0.63	7.09	**5.90**	7.82
	5th	**0.52**	0.58	0.81	**6.73**	7.73	10.51
	11th	0.24	**0.22**	0.37	6.63	**6.17**	10.26

VI. CONCLUSION

This paper frames the PQ parameter forecasting tasks by considering short-term disaggregated and percentile aggregated forecasts of harmonic values. Available PQ input data are used in QR models to generate the probabilistic forecasts, and a dedicated percentile estimation procedure is provided for the indirect approach proposed for the aggregated forecasting task. The evaluation based on actual LV data confirms the accuracy of the proposed methodologies, with improvements from 22% to 43% compared to the naïve benchmark. Future works will address the forecast of other PQ parameters (e.g., voltage harmonics), including different underlying probabilistic models and clustering/decomposition techniques to refine the data pre-processing stage.

REFERENCES

[1] IEC 61000-2-5:2017. Electromagnetic compatibility (EMC) - Part 2-5: Environment - Description and classification of electromagnetic environment. 2017.

[2] IEC 61000-3-6:2008. Electromagnetic compatibility (EMC) - Part 3-6: Limits - Assessment of emission limits for the connection of distorting installations to MV, HV and EHV power systems. 2008.

[3] IEC 61000-3-12:2011. Electromagnetic compatibility (EMC) - Part 3-12: Limits - Limits for harmonic currents produced by equipment connected to public low-voltage systems with input current >16 A and ≤ 75 A per phase.

[4] IEEE Std. 519-2014. IEEE Recommended Practice and Requirements for Harmonic Control in Electric Power Systems. 2014.

[5] C. Noce, L. Di Stasio, P. Varilone, P. Verde, and M. De Santis, "On the Forecast of the Voltage Sags: First Stages of Analysis on Real Systems, in proc. 2020 55th Int. Univ. Power Eng. Conf. (UPEC), pp. 1-6, 2020.

[6] IEEE Std. 1547-2018. IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces. 2018.

[7] T. Vantuch, S. Misak, T. Jezowicz, T. Burianek, and V. Snasel, "The power quality forecasting model for off-grid system supported by multiobjective optimization," IEEE Trans. Industr. Electron., vol. 64, no. 12, pp. 9507-9516, 2017.

[8] L. Zjavka, "Power quality multi-step predictions with the gradually increasing selected input parameters using machine-learning and regression," Sustain. Energy Grids Netw., vol. 26, 100442, 2021.

[9] P. Rodríguez-Pajaron, A. Hernandez Bayo, and J. V. Milanovic, "Forecasting voltage harmonic distortion in residential distribution networks using smart meter data," Int. J. Electr. Power Energy Syst., vol. 136, 107653, 2022.

[10] I. S. Jahan, S. Misak, and V. Snasel, "Smart control system based on power quality parameter short-term forecasting," in proc. 2020 21st Int. Sci. Conf. Electr. Power Engineer. (EPE2020), pp. 1-5, 2020.

[11] J. Rodway, P. Musilek, S. Misak, and L. Prokop, "Prediction of PV power quality: Total harmonic distortion of current," in proc. 2013 IEEE Electr. Power Energy Conf., pp. 1-4, 2013.

[12] A. Bracale, P. Caramia, P. De Falco, and G. Carpinelli, "Probabilistic power quality level forecasting through quantile regression models," in proc. 2022 20th Int. Conf. Harmon. Quality Power (ICHQP2022), pp. 1-6, 2022.

[13] T. Hong, and S. Fan, "Probabilistic electric load forecasting: A tutorial review," Int. J. Forecast., vol. 32, no. 3, 2016.

[14] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, and G.M. Ljung, Time series analysis: forecasting and control, 5th ed., Wiley, 2015.

[15] R. Koenker. Quantile regression. Cambridge University Press, 2005.

[16] T. Gneiting, and A.E. Raftery, "Strictly proper scoring rules, prediction, and estimation," J. Am. Stat. Assoc., vol. 102, no. 477, pp. 359-378, 2007.

[17] L. Alfieri, and P. De Falco, "Wavelet-based decompositions in probabilistic load forecasting," IEEE Trans. Smart Grid, vol. 11, no. 2, pp. 1367-1376, 2020.

[18] A. Papoulis, S.U. Pillai, Probability, random variables and stochastic processess, 4th ed., McGraw-Hill, 2002.

[19] R.J. Hyndman, and Y. Fan, "Sample quantiles in statistical packages," Am. Stat., vol. 50, no. 4, pp. 361-365, 1996.

[20] P. Pinson, H.A. Nielsen, J.K. Moller, H. Madsen, and G.N. Kariniotakis, "Non-parametric probabilistic forecasts of wind power: required properties and evaluation," Wind Energy, vol. 10, no. 6, pp. 497-516, 2007.