Proceedings of Machine Learning Research 204:1–19, 2023 Conformal and Probabilistic Prediction and Applications

Applications of Conformal Regression on Real-world Industrial Use Cases using Crepes and MAPIE

Nasir Uddin NASIR.UDDIN@HUSQVARNAGROUP.COM
Data Scientist, AI Lab, Husqvarna Group, Regeringsgatan 28, 111 53 Stockholm, Sweden

Tuwe Löfström TUWE.LOFSTROM@JU.SE
Assistant Professor, Department of Computing, Jönköping School of Engineering, Jönköping University.

Editor: Harris Papadopoulos, Khuong An Nguyen, Henrik Boström and Lars Carlsson

Abstract

Applying conformal prediction in real-world industrial use cases is rare, and publications are often limited to popular open-source data sets. This paper demonstrates two experimental use cases where the conformal prediction framework was applied to regression problems at Husqvarna Group with the two Python-based open-source platforms MAPIE and Crepes. The paper concludes by discussing lessons learned for the industry and some challenges for the conformal prediction community to address.
Keywords: Crepes, MAPIE, conformal regression, EnbPI, demand prediction, injection molding, manufacturing analytics, supply-chain, conformal predictive system,

1. Introduction

Conformal prediction (CP) uses past experience to determine precise levels of confidence in new predictions (Vovk et al., 2005). The CP framework (Shafer and Vovk, 2008; Murphy, 2023), including a group of distribution-free uncertainty quantification methods, effectively creates prediction intervals or sets with guaranteed frequentist coverage probability from any predictive method. For example, given a user-defined error probability ϵ, together with a method that makes a prediction $\hat{y}$ of a label y, it produces a set of labels, typically containing $\hat{y}$, that also contains y with probability $1 - \epsilon$. For a general input x and output y (not necessarily discrete) the CP method could be outlined in the following steps (Angelopoulos and Bates, 2021):

1. Identify a heuristic notion of uncertainty using a pre-trained model, such as softmax for classification or variance for a regression problem.

2. Use the heuristic notion to define a conformal score $s(x, y) \in \mathbb{R}$ which measures how badly the output y "conforms" to x. The score is also known as the non-conformity score since larger scores encode worse agreement between x and y. For regression, the most common choice is to use the absolute error as a non-conformity score, α.

3. Apply this score to a calibration set (with splitting approaches) of n labeled instances and compute $\hat{y}$ as the $\frac{(n+1)(1-\alpha)}{n}$ quantile of the calibration scores

$$s_1 = s(X_1, Y_1), ..., s_n = s(X_n, Y_n).$$

The user specifies a desired confidence threshold ϵ and then computes the $1-\epsilon$ quantile $\hat{q}$ of S. If enough data is unavailable to adopt the splitting approaches, the full conformal prediction could be applied, which requires fitting the model n times using a leave-one-out procedure.

4. Use this quantile to form the prediction sets for a new example

$$C(X_{test}) = y : s(X_{test}, y) \geq \hat{q}.$$

The conformal coverage guarantee (Vovk et al., 1999) can be stated as

$$P\left(Y_{test} \in C(X_{test})\right) \geq (1-\epsilon),$$

supposing that $(X_i, Y_i)_{i=1\ldots n}$ and (X_{test}, Y_{test}) are exchangeable, and that $\hat{q}$ and $C(X_{test})$ are defined from the above-mentioned steps 3 and 4, respectively. The exchangeability assumption is weaker than the i.i.d. assumption and is fulfilled if the distribution is invariant to the order of the instances.

In regression problems, the full information of an uncertain test label exists in its probability distribution, as it provides any probability of the event related to the label and can be easily transformed to a point prediction or an interval prediction with the corresponding first-order moment or quantiles (Wang et al., 2020). Conformal predictive systems (CPS) provide predictions in the form of complete cumulative distribution functions (conformal predictive distributions), which are more informative as it is possible to generate not only the prediction intervals but also to obtain calibrated point predictions, as well as p-values for given target values (Vovk et al., 2022; Wang et al., 2020; Boström, 2022). CPS also allows control of the error level when making predictions on whether the correct target values are larger (or lower) than a certain threshold value, e.g., we may rule out with 99% confidence that the temperature will be above a critical level (Boström, 2022).

Generating split (also known as inductive) CPS is similar to forming inductive conformal regressors. The most important difference is that the nonconformity scores are calculated by considering actual (signed) errors instead of absolute errors. The prediction for a test instance is the following conformal predictive distribution:

$$Q(y) = \begin{cases} \frac{n+\tau}{q+1}, \text{ if } y \in \left(C_{(n)}, C_{(n+1)}\right), & \text{for } n \in \{0, ..., q\} \\ \frac{n'-1+(n''-n'+2)\tau}{q+1}, \text{if } y = C_{(n)}, & \text{for } n \in \{1, ..., q\} \end{cases} \tag{1}$$

where $C_{(1)}, \ldots, C_{(q)}$ are obtained from the calibration scores $\alpha_1, \ldots, \alpha_q$, sorted in increasing order $C_{(i)} = h\left(x\right) + \alpha_i$ with $C_{(0)} = -\infty$ and $C_{(q+1)} = \infty$. The role of τ (sampled from the uniform distribution $U(0,1)$) is to allow the p-values of target values to be uniformly distributed. To handle ties, n'' and n' are the highest index such that $y = C_{(n'')}$ and the lowest index such that $y = C_{(n')}$, respectively. The function returns the estimated probability $P(Y \leq y)$, for a specific y and where Y is a random variable corresponding to the true target.

The conformal prediction (CP) framework has a number of very useful features[1] such as i) no assumptions about the distribution of the data, ii) can be applied to any predictive

1. https://leanpub.com/conformal-prediction

model, the resulting prediction sets come with guarantees of covering the true outcome with a certain probability, iii) can be used without retraining the model and iv) can be applied for classification, regression, time series forecasting, and many other tasks. It has been used in real-world use cases including medical diagnosis, face recognition, financial risk prediction (Balasubramanian et al., 2014), and recently the framework has seen exponential growth (Fontana et al., 2023; Manokhin, 2022b).

While the CP framework has several desirable properties, the efforts of researchers in adapting this framework have been limited to a few groups of researchers scattered around the world. Furthermore, it has also been slow in being adopted by machine learning practitioners for a long time (Balasubramanian et al., 2014). Moreover, in some industries, including manufacturing organizations, widespread AI or machine learning adoption is lacking due in part to a lack of high-quality *Big* or *small* data[2]. To the best of our knowledge, application of conformal prediction on real-world data sets coming from the power-tools manufacturing industry is also very rare.

Husqvarna Group[3], founded in Sweden in 1689, is currently a world-leading manufacturer of innovative products[4] and solutions for forest, park and garden management including robotic lawnmowers, chainsaws, trimmers, riding lawnmowers, and garden irrigation systems. The group is also a leader in equipment and diamond tools for the construction and stone industries[5]. The Husqvarna Group has, throughout its long history, adapted early to new market demands, business models, and industry trends (Simonsson et al., 2020).

The Husqvarna AI Lab, established in 2018, is a part of the Husqvarna Group that supports all business units by helping different teams to ideate, test, apply, and learn AI and data-driven technologies for product and solution development. Thus, data science and machine learning have been used at AI Lab, both in big and tinyML approaches, as an RnD supporting tool in various use cases, including manufacturing quality control, supply-chain demand prediction, identifying product uses and customer behavior, management of park and green spaces, measuring robotics and handheld product performances, etc. The lab also collaborates closely with external organizations/startups for solution development[6] as well as academia, including hosting industrial Ph.D. positions (Simonsson, 2021; Agarwal, 2022) and MSc student projects (Anam, 2021).

Models developed using machine learning are widely used for various types of inference and decision-making processes in real-world environments. In order to increase trust in machine learning systems, uncertainty quantification (UQ) is very important since the predictions made by such models are subject to noise and model inference errors (Abdar et al., 2021). However, many machine learning models come without uncertainty quantification or output overconfident predictions, resulting in incorrect decisions and technology acceptance issues (Manokhin, 2022a).

The conformal prediction framework can help decision-making by creating statistically rigorous uncertainty sets/intervals for the predictions of machine learning models which are valid in a distribution-free sense, meaning they possess explicit, non-asymptotic guarantees

2. https://mitsloan.mit.edu/ideas-made-to-matter/why-its-time-data-centric-artificial-intelligence
3. https://www.husqvarnagroup.com/en/about
4. https://time.com/collection/best-inventions-2022/6230428/husqvarna-automower-415x/
5. https://www.husqvarnagroup.com/en/husqvarna-construction-division
6. https://hugsi.green/

even without distributional or model assumptions (Angelopoulos and Bates, 2021; Vovk and Bendtsen, 2018). This allows uncertainty quantification to avoid consequential model failures in high-risk settings such as medical diagnostics. In addition, conformal predictive distributions can be used for decision-making by producing sets that are guaranteed to contain the ground truth with a user-specified probability (Vovk and Bendtsen, 2018).

In probabilistic prediction, the critical evaluation criteria used[7], are *validity* (also termed *calibration* or *coverage*, which is necessary or a *must-have*), and *efficiency* (also termed *width* or *sharpness*, which is desirable but not a *must-have*). In other words, calibration refers to the statistical consistency between the distributional forecasts and the observations which is a joint property of the predictions and the events that materialize, whereas sharpness refers to the concentration of the predictive distributions which is a property of the forecasts only (Gneiting et al., 2007). In a previous study among various uncertainty quantification approaches, only conformal prediction methods satisfied the property of *validity* (Dewolf et al., 2023).

Recently, several open-source python-based platforms, such as Crepes[8], MAPIE[9], AWS Fortuna[10] and Neural prophet[11] have been released for quantifying uncertainty with conformal prediction approaches (Manokhin, 2022b). Crepes and MAPIE are probably the most popular among them due to their extensive documentation and availability of APIs for various conformal prediction methods.

Crepes (Boström, 2022) can generate conformal regressors, which transform point predictions of any underlying regression model into prediction intervals for specified levels of confidence. In addition to that, the package also implements conformal predictive systems, which transform the point predictions into cumulative distribution functions implementing standard, normalized, and Mondrian conformal predictive systems.

MAPIE (Taquet et al., 2022) allows easily estimating prediction intervals (or prediction sets) using a scikit-learn-compatible model for single-output regression or multi-class classification settings. It has better support for time-series data, due to time-series specific solutions such as jacknife+ (Barber et al., 2021) and jackknife+-after-bootstrap (Kim et al., 2020).

2. Application Case: Container Prediction

2.1. Background

Husqvarna Construction division's manufacturing and assembly sites are located on different continents in the world including in Asian countries. Usually, standard shipping containers are used to ship products from manufacturing facilities to warehouses or customer premises. In order to secure the shipment from another continent and reduce unnecessary costs demand planning and ordering the number or volume of containers need to be done, as correctly as possible, at least four weeks ahead of shipment from the manufacturing and assembly site.

7. https://valeman.medium.com/how-to-evaluate-probabilistic-forecasts-ace8b7ad3491
8. https://github.com/henrikbostrom/crepes
9. https://mapie.readthedocs.io/en/latest/
10. https://github.com/awslabs/fortuna
11. https://github.com/ourownstory/neural_prophet

However, predicting the demand of containers (usually in TEU, i.e., twenty-foot equivalent unit, standard-sized metal boxes[12]) is a very challenging and complex task since the demand for containers might also depend on the demands, inventory and supply planning of other products inside the containers. A lack of understanding of the relationship between the construction products category and container demands also hinders planning and forecasting.

In addition, co-ordinations and collaboration among various teams, such as sales & marketing, logistics & operation, and inventory & planning are needed to improve the accuracy of the forecast and deliver the product within the expected time. In recent years, multiple disruptions have occurred in the global supply chain due to various factors such as pandemics, the rising cost of living, labor unrest, energy shortages, geopolitical uncertainty, and extreme weather, which could make forecasting more challenging [13].

2.1.1. Goal of the predictive system

The Business problem was defined together with the Logistics and S&OP Process groups of Husqvarna Construction as a regression problem for now. Currently, ordering containers combines historical data (e.g., 2-5 years) and internal communication among various sales, operation, and logistics team members. However, machine learning or any statistical forecasting methods with uncertainty quantification has not yet been applied for container forecasting or demand planning. Therefore, the main goal of this study is to introduce machine learning and statistical forecasting methods supported by uncertainty quantification framework such as conformal regression (Johansson et al., 2021). Following the CP framework would also make it possible to convert the classical point prediction into a confidence predictor with valid prediction intervals (Alvarsson et al., 2021). In addition, it is also important to understand historical trends and seasonality and identify the key product groups or other driving forces which might contribute to container demands.

2.2. Method: Container Prediction

2.2.1. data set

The following data set was collected, aggregated and min-max normalized from various parts of Husqvarna's supply-chain organizations, including factories in anonymous locations in Asia, mentioned as route X.

- Historical data of container demand: Weekly total container demand in TEU for specific route X.
- Product categories: Quantity of product ordered from various product category classes mentioned as CCE, CDE, CSE.
- Date: Week when containers were loaded for shipment from the manufacturing or assembly location

12. https://en.wikipedia.org/wiki/Twenty-foot_equivalent_unit#
13. https://www.weforum.org/agenda/2022/09/5-challenges-global-supply-chains-trade

Raw data were collected as Excel files (from various supply-chain and logistics groups), containing shipped container and product information between 2017 to 2021 from the manufacturing and assembly site in Asia towards various destinations or routes. The raw data consists of about 838 data points (shipping dates) describing container information including quantity in TEU. It also contained about 15669 data points (loading dates) containing product and parts information (e.g., 2833, 2783, 3358, 3156, and 3539 data points for 2017-2021, respectively) for various routes. However, in this study, we only included the cleaned, weekly aggregated, merged, and pre-processed data from one route consisting of the number of shipped containers and some selected product categories mentioned above. Example of products shipped inside the containers is from Husqvarna Constructions[14], including floor grinders, industrial vacuum systems, dust collectors, early entry saw, power packs, dust collectors, tile saws, floor saws, etc.

2.2.2. Error, trend, seasonality (ETS) decomposition and Machine Learning Modelling

Exploratory data analysis (EDA) and machine learning modeling with point prediction were conducted using open-source Python-based tools, including scikit-learn[15], pycaret[16] and statsmodel[17] in a local or cloud Databricks[18] environment.

The ETS decomposition procedures are used in time series analysis to describe the trend and seasonal factors in a time series. The additive model is useful when the seasonal variation is relatively constant over time whereas the multiplicative model is useful when the seasonal variation increases over time[19]. The additive model assumes that the time series can be expressed as the sum of the trend, seasonal, and remainder or residual component (Hyndman and Athanasopoulos, 2018) and can be expressed as follows:

$$Y_t = S_t + T_t + R_t$$

where Y_t is the observed value at time t, T_t is the trend component at time t, S_t is the seasonal component at time t, and R_t is the random or residual component at time t.

In a preliminary screening with AutoML approaches for monthly and weekly container data, it was observed that traditional autoregressive models (e.g., ARIMA, SARIMA) and tree-based machine learning modes (e.g., Random Forests, XGBoost) performed better than other complex models including neural network-based deep learning (data not shown). For simplicity reasons, Random Forest was applied as the base model for conformal prediction with Crepes and MAPIE following examples notebooks in their respective GitHub repositories.

The data set was divided into a training and a test set, and the training set was further split into a proper training set and a calibration set while using Crepes (Boström, 2022). While applying the Ensemble batch prediction interval (EnbPI) methods in MAPIE (Xu and Xie, 2021), the hyper-parameters are optimized with a RandomizedSearchCV using a

14. https://www.husqvarnaconstruction.com/int/
15. https://scikit-learn.org/stable/index.html
16. https://pycaret.org/
17. https://www.statsmodels.org/stable/index.html
18. https://www.databricks.com/
19. https://online.stat.psu.edu/stat510/lesson/5/5.1

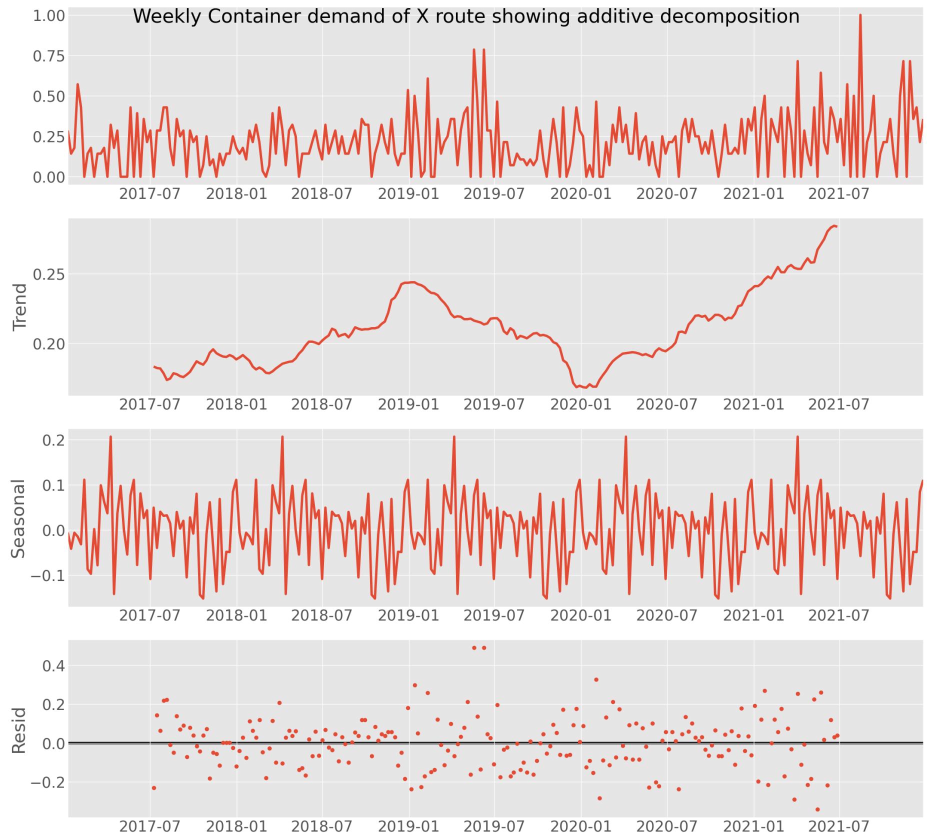

Figure 1: Time series decomposition of Weekly container demand of route X

sequential TimeSeriesSplit cross-validation, in which the training set is prior to the validation set. The best model is then fed into MapieTimeSeriesRegressor to estimate the associated prediction intervals. Finally, we compare two approaches: with or without partial_fit called at every step.

2.3. Results: Container Prediction

2.3.1. ETS decomposition

Figure 1 shows weekly container demand with additive decomposition model (with seasonal_decompose API) including trend, seasonality, and residual components. Figure 1 reveals that demands of containers showed seasonal fluctuation behavior such as increasing demand in every summer (June-July) during the year (indicated in seasonal sub-figure) and an overall slight increasing trend since 2021 (indicated in the trend sub-figure).

The reasons for those periodical seasonal fluctuations are unknown although it is known that during the summer period, the demand for some Husqvarna products used for gardening and green space management increased. However, additional experiments need to be done to see whether climatic change or any weather component could have a causal impact on container demand. In a parallel experiment with the Granger causality test (Granger, 1969), we didn't find significant evidence that temperature or humidity could be useful to forecast container demand for this route for construction products (data not shown). In addition, more experiments need to be done to see whether there is any correlation between the global post-Covid market opening and increasing construction work activities since the beginning of 2021.

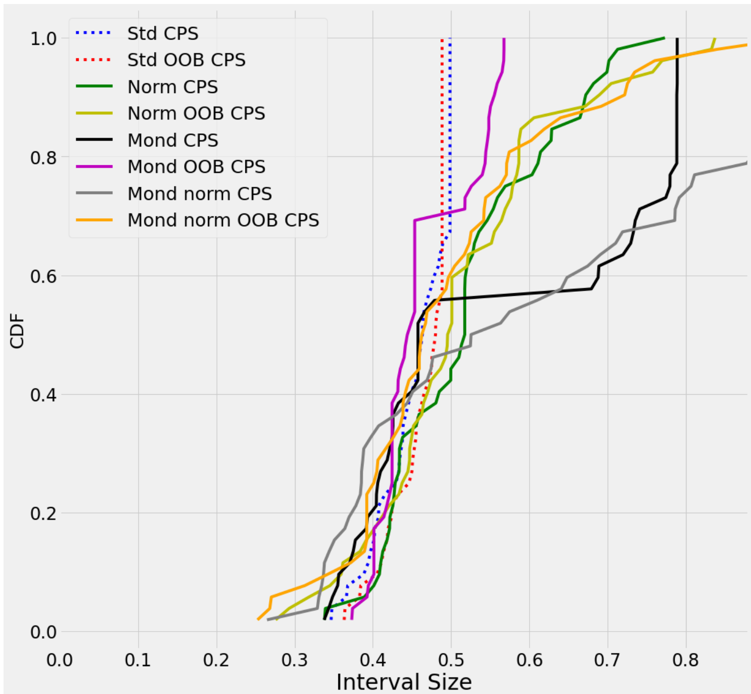

Figure 2: Distribution of interval sizes showing various conformal predictive system (CPS) with Crepes package, applied on container demand prediction data set

2.3.2. Conformal Prediction

The main goal of applying conformal prediction for this study was to quantify uncertainty with prediction intervals, instead of merely a point prediction used in traditional approaches, applicable for any data distribution, data set size, and underlying regression model (e.g., statistical, machine learning, or deep learning).

In order to implement standard and normalized and Mondrian conformal regressors, and predictive systems in crepes we have extracted prediction intervals at the 90% confidence level (lower and upper percentiles = {5, 95}). Figure 2 shows how the interval sizes are distributed for the eight conformal predictive systems when applied to the test set. The graphs showed that the distribution of the interval size varied among various conformal predictive systems. The results of evaluating all CPS on all metrics (error, efficiency, CRPS (continuously ranked probability score), Time to fit, and Evaluation time) and three confidence levels (80%, 85%, and 90%) are demonstrated in Table 1. The first two metrics (error and efficiency), in the table, correspond to the fraction of true targets that are not included in the prediction intervals, which should be close to one minus the chosen confidence level, whereas the last two, time_fit and time_evaluate correspond to the wall clock time in seconds for fitting and making predictions with the conformal regressor (Boström, 2022).

The third metric CRPS (continuous-ranked probability score) has been widely used to evaluate the performance of probabilistic forecasts (Vovk et al., 2020; Werner et al., 2020; Zamo and Naveau, 2018; Gneiting and Katzfuss, 2014). It is measure of the difference between a predicted and observed cumulative distribution function (CDF) [20] and calculated following Crepe's (Boström, 2022) GitHub notebook [21].

For a 90% confidence level, normalized Mondrian CPS showed the lowest CRPS whereas in efficiency metric Standard CPS showed the lowest interval width. In addition, the only result that is close to being valid is normalized Mondrian CPS although in the efficiency metric, it performed worst. It is hard to say whether that is due to that solution being able to address the lack of exchangeability or happened by chance. For this small data set, results from the time_fit and time_evaluate metrics are negligible.

As mentioned previously, different data distributions and trends were observed among different routes and in order to have a better understanding of the future demands it might be necessary to see the whole CDF (cumulative distribution function) which maps a value to its probability of being less than or equal to that value. The CPS gives full and complete quantification of uncertainty tailored to each test object for optimal decision-making. In this particular real-world case, this time series may not satisfy the exchangeability assumption. Following the notebook examples available in the Crepes repository, we were also unable to achieve a confidence interval of more than 90% for this data set with 51 test instances while trying to create the CPS graph (data not shown).

Time series forecasting can be seen as a regression with a time structure. However, due to the time order of the data points, calibration data and new/test data are not exchangeable, as could be seen in Table 1.

MAPIE has various APIs suitable for time series data sets to solve this issue such as EnbPI (Ensemble Prediction Intervals) and doesn't require to exchangeability assumption. Here, EnbPI (Xu and Xie, 2022) method was implemented with the *MapieTimeSeries-Regressor* API which allows updating the residuals during the prediction, each time new observations are available so that the deterioration of predictions, or the increase of noise level, can be dynamically taken into account[22]. The MAPIE API has two options to estimate the prediction intervals such as a) with a regular *.fit* and *.predict* process, limiting the use of training set residuals to create prediction intervals; and b) using *.partial_fit* in addition to *.fit* and *.predict* which allows MAPIE to use new residuals from the test points as new data are becoming available. For example *.partial_fit* method is useful to adjust prediction intervals to high volatility situations such as sudden drop, jump, or change points on the tests that have not been seen by the model during training.

In Figure 3 we observed that the EnbPI with the partial_fit option widened the intervals as time progressed but gave higher coverage than without the partial_fit option. However, in both scenarios the coverage is not satisfactory, as for confidence level 99%, the coverage was 0.80 and 0.90 without and with partial_fit, respectively. Decreasing the confidence level to 95% also decrease the coverage and width.

20. https://www.lokad.com/continuous-ranked-probability-score
21. https://github.com/henrikbostrom/crepes
22. https://mapie.readthedocs.io/en/latest/examples_regression/2-advanced-analysis/plot_timeseries_enbpi.html

	Std CPS			Norm CPS			Mond CPS			Mond norm CPS		
	0.80	0.85	0.90	0.80	0.85	0.90	0.80	0.85	0.90	0.80	0.85	0.90
error	0.3077	0.2500	0.2115	0.2692	0.2115	0.1538	0.3077	0.1731	0.1731	0.2115	0.1346	0.1346
efficiency	0.3769	0.4227	0.4553	0.4091	0.4810	0.5191	0.4128	0.5669	0.5669	0.5203	0.6020	0.6020
CRPS	0.1320	0.1320	0.1320	0.1230	0.1230	0.1230	0.1288	0.1288	0.1288	0.1077	0.1077	0.1077
time_fit	0.0001	0.0001	0.0001	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
time_evaluate	0.0021	0.0014	0.0013	0.0018	0.0015	0.0013	0.0020	0.0015	0.0014	0.0020	0.0015	0.0015

	Std OOB CPS			Norm OOB CPS			Mond OOB CPS			Mond norm OOB CPS		
	0.80	0.85	0.90	0.80	0.85	0.90	0.80	0.85	0.90	0.80	0.85	0.90
error	0.3462	0.3077	0.2308	0.2692	0.2500	0.1923	0.3654	0.3269	0.2885	0.2692	0.2500	0.2308
efficiency	0.3639	0.3995	0.4617	0.4161	0.4550	0.5128	0.4097	0.4306	0.4637	0.4237	0.4486	0.4987
CRPS	0.1289	0.1289	0.1289	0.1285	0.1285	0.1285	0.1421	0.1421	0.1421	0.1341	0.1341	0.1341
time_fit	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
time_evaluate	0.0020	0.0013	0.0014	0.0018	0.0014	0.0014	0.0019	0.0014	0.0015	0.0018	0.0014	0.0015

Table 1: Evaluation of generated CPS using three confidence levels on the test set using Crepes packages

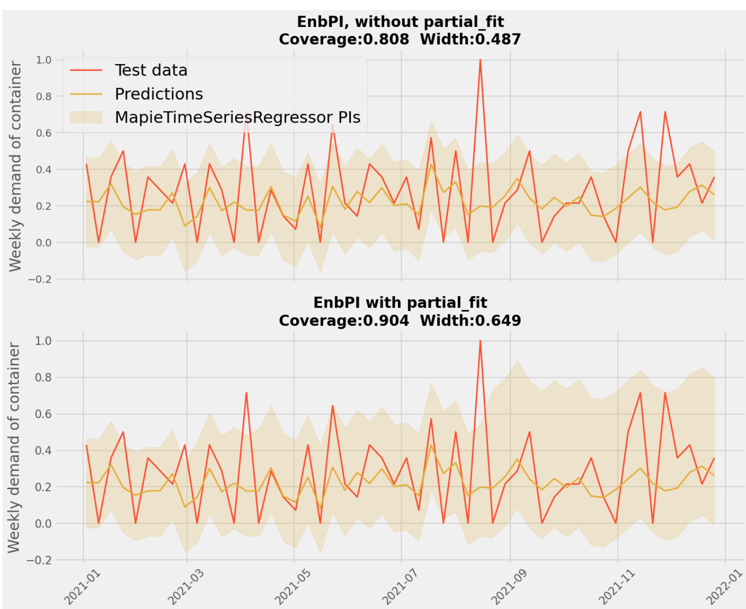

Figure 3: EnbPi with MAPIE showing coverage and width applied on container demand prediction data set

2.4. Discussion: Container Prediction

The main takeaway from these experiments is that achieving good results regarding both validity(calibration/coverage) and efficiency (width/sharpness) with any probabilistic prediction methods including conformal prediction is very challenging. Due to the data set's non-exchangeability and smaller size, only normalized Mondrian CPS gave close to the valid prediction error and lowest CRPS in the Crepes platform. Results from MAPIE with EnbPI also deviate from the expected coverage. However, MAPIE could be a practical option in real-world production use cases since it has a user-friendly API suitable for time series data sets supporting various methods whereas it is easier in Crepes to see the probability distribution at a glance and helpful for decision-making.

For example, in this specific use case, graphs generated by the EnbPI method from MAPIE would be very useful to order the container at least few weeks ahead with close to the coverage guarantee. However, it would be risky to put an order several months ahead if efficiency appears to decrease by widening the interval. It might be possible to keep the order of the container on hold for specific future spikes or anomaly situations arise. MAPIE's partial_fit API could be a good option to implement.

In order to have a successful uncertainty quantification it is not enough to look at the underlying machine learning models or algorithms and probability distribution of the target (e.g., container). Measuring feature importance and feature impact on the final outcome is also necessary. In this case, a parallel study should be conducted on each product group and their impact on the final container volume. In addition, underlying machine learning methods used in CP are also very important as in a practical scenario if the baseline machine learning model gave high error even as mean or median value as point prediction it would be harder to justify spending time on uncertainty quantification alone. Moreover, several predictability factors such as the underlying data-generating process, data size, and adequate repeating patterns must be considered before defining the forecasting problem. In our experience, if the data set is small and lacks enough information, it is worth following simple auto-regressive statistical models and focusing only on the future anomaly detection part for uncertainty quantification. For the supply chain use cases, irregular disruptions usually cost more money than regular disruptions. It would make sense to convert this use case into conformal anomaly detection or classification approaches by setting a certain interval range. The probability of deviation from that range could signal certain business actions. In addition, cloud-based centralized data management and synchronizations are needed in the form of a data lake or lake-house which could accelerate understanding of the data-generating process and the usefulness of historical patterns for the forecasting problem.

3. Application Case: Injection Mold Prediction

3.1. Background

Injection molding developed almost 150 years ago, is a widely used manufacturing technique in which molten material is injected into a mold to produce various machine parts (Fu et al., 2020). More than 30% of the weight of all plastic products is manufactured using the injection molding process (Párizs et al., 2022). A whole process could consist of several stages such as inserting materials inside the mold, closing the mold, filling, holding pressure, cooling, and demolding (Jung et al., 2021). Several important factors including injection time, filling time, cycle time, switchover position and pressure, melting temperature, injection speed, and maximum injection pressure might impact the quality of the final molding products (Jung et al., 2021; Selvaraj et al., 2022).

Modern manufacturing companies have been using AI and machine learning methods in order to optimize production efficiency. However, the application of machine learning in the injection molding industry is a comparatively new phenomenon. The industry has mostly been relying on traditional methods and operators' expertise. The full potential of the fast-growing and changing data in the manufacturing domain has not yet been unlocked (Rousopoulou et al., 2020; Selvaraj et al., 2022). In addition, each injection molding product might need a customized system for optimized performance and quality.

Husqvarna Group has several injection molding facilities in Europe. One of them is located in Newton Aycliffe (UK)[23] producing various parts for robotic lawn mowers[24]. In this study, we have used data from plastic injection molding experiments in order to understand the factors influencing the variation of weight of the molded product (e.g., chassis/case) of the automower. Reducing the final product's weight variation is challenging as it might require predicting and controlling multiple features' impact on the molding process.

3.1.1. Objectives/Benefits

The aim is to develop a system that could consider the important parameters and the product's weight to predict when we will have a problem that could produce a scrap part. We are not expecting to control the weight but the amount of variation in the weight so that it will stay within the expected range. We believe conformal prediction with regression approaches could be useful to monitor the expected variation with valid prediction intervals in this use case.

3.2. Method: Injection Mold Prediction

3.2.1. Data set

About 3438 cycles (each time it makes a part from the molding process) of data points were used for this study using *Weight (Kg)* as target and fifteen features as input variables as shown in Table 2.

3.2.2. Machine Learning Modeling

Data pre-processing and machine learning modeling with conformal prediction were done with the same Python-based platforms mentioned in previous sections. In Crepes, data were split into a train, calibration, and test set, whereas for applying the EnbPI method in MAPIE, the data set was divided into two groups: train plus validation set and test set. Both Crepes and MAPIE used 516 cycles as the test set. Random Forest regression was used as the underlying machine-learning algorithm for conformal prediction, as described in the previous use case.

23. https://youtu.be/k9rFvGVEs_0
24. https://www.husqvarna.com/uk/robotic-lawn-mowers/automower-305/

No.	Feature Variables	Unit
1	Cycle Time	second
2	Efficiency	percentage
3	Average Cycle Time	second
4	Melt Temperature	Celsius
5	Filling Time	second
6	Plasticising Time	second
7	Max. Clamping Force	Kilo Newton
8	Cushion Position	millimeter
9	Switchover Position	millimeter
10	Plasticising Position	millimeter
11	Max. Injection Speed	millimeter per second
12	Max. Screw Rotation Speed	RPM
13	Mean Screw Rotation Speed	RPM
14	Max. Injection Pressure	bar
15	Switchover Pressure	bar

Table 2: List of features used for predicting Weight in injection moulding cycle

3.3. Results: Injection Mold Prediction

Distribution of extracted prediction interval at 95% confidence level (lower and upper percentiles = {2.5, 97.5}) and coverage size for various conformal predictive systems (CPS) showed in Figure 4.

The results of evaluating all CPS on all metrics (error, efficiency, CRPS, Time fit, and time evaluation) and three confidence levels (80%, 90%, and 95%) are demonstrated in Table 3. From the results, it appeared that in efficiency metrics, Std CPS and Std OOB CPS performed well (at 95% confidence level) as in the case of conformal predictors, efficiency means that the prediction regions should be as narrow as possible. In error or coverage metric, all the CPS methods gave valid coverage at a 95% confidence level showing the lowest value for normalized Mondrian CPS. In the CRPS metric, normalized Mondrian CPS also appeared to have the lowest value.

EnbPI method showing coverage and width in regular and partial fit approaches in Figure 5 demonstrated that with an update of residual approaches, both coverage and width increase in the test set. Since we didn't test or simulate the method by applying change point, there is no way to justify whether `partial_fit` approaches will decrease the width and increase the coverages.

3.4. Discussion: Injection Mold Prediction

Unlike the container demand prediction uses case, all of the conformal prediction methods applied to injection molding data gave expected valid coverage applied in Crepes and MAPIE platforms. We believe that the injection molding use case might be a good example where conformal prediction could be applied for monitoring quality control by providing prediction intervals or sets that have guaranteed coverage under certain assumptions. In addition, the conformal risk control concept might be useful to set up the ideal or expected

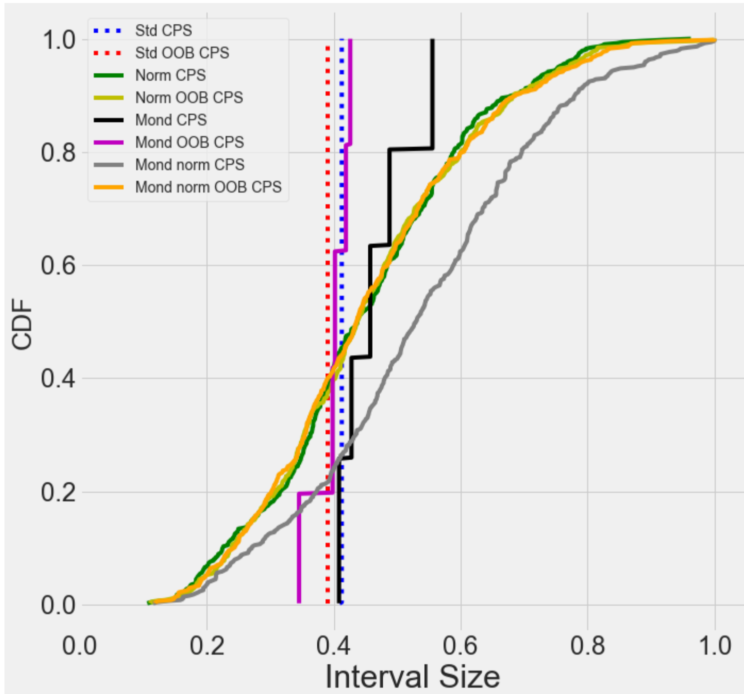

Figure 4: Distribution of interval sizes showing various conformal predictive system (CPS) with Crepes package, applied on injection molding data set

	Std CPS			Norm CPS			Mond CPS			Mond norm CPS		
	0.80	**0.90**	**0.95**	**0.80**	**0.90**	**0.95**	**0.80**	**0.90**	**0.95**	**0.80**	**0.90**	**0.95**
error	0.1899	0.0853	0.0349	0.1996	0.0717	0.0407	0.1647	0.0795	0.0368	0.1996	0.0756	0.0329
efficiency	0.2401	0.3228	0.4127	0.2617	0.3719	0.4522	0.2469	0.3373	0.4639	0.2673	0.3917	0.5360
CRPS	0.0541	0.0541	0.0541	0.0536	0.0536	0.0536	0.0543	0.0543	0.0543	0.0526	0.0526	0.0526
time_fit	0.0001	0.0001	0.0001	0.0003	0.0003	0.0003	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
time_evaluate	0.0142	0.0134	0.0128	0.0144	0.0138	0.0134	0.0128	0.0122	0.0120	0.0130	0.0126	0.0126

	Std OOB CPS			Norm OOB CPS			Mond OOB CPS			Mond norm OOB CPS		
	0.80	**0.90**	**0.95**	**0.80**	**0.90**	**0.95**	**0.80**	**0.90**	**0.95**	**0.80**	**0.90**	**0.95**
error	0.1919	0.0872	0.0465	0.1977	0.0872	0.0465	0.1899	0.0969	0.0446	0.1744	0.0853	0.0504
efficiency	0.2386	0.3165	0.3899	0.2557	0.3539	0.4548	0.2379	0.3116	0.3977	0.2596	0.3590	0.4543
CRPS	0.0540	0.0540	0.0540	0.0540	0.0540	0.0540	0.0538	0.0538	0.0538	0.0532	0.0532	0.0532
time_fit	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0003	0.0003	0.0003
time_evaluate	0.0184	0.0186	0.0174	0.0201	0.0185	0.0182	0.0149	0.0143	0.0146	0.0153	0.0150	0.0147

Table 3: Evaluation of generated CPS using three confidence levels on the test set using Crepes packages applied on injection molding data set

Figure 5: Estimating prediction intervals with EnbPI applied on injection molding data set

range for various process parameters (e.g., mentioned features in this experiment) which could impact the final molded plastics products. As discussed earlier, in addition to validity (which is a must-have criterion), the efficiency or sharpness parameter must also be taken into consideration. This is particularly important in this use case, where the final plastic weight must *not* deviate too much (e.g., maximum 0.29% variation). It is hard to determine whether the sharpness parameter showed in Figure 5 is within the expected range as more injection molding experiments need to be conducted by the factory to set up more precise values and ranges with more certainty.

In addition, it would be interesting to see which parameters are easier to control and optimize as well as which will have the maximum impact to stabilize the process. Some variations that happen during the molding process might not be cost-effective to control if they are due to the simple *warm-up* effect or the occasional situation happening once the system starts to run after a pause. From the time series ETS decomposition, it was easy to see the seasonal fluctuations happening every few hours (data not shown). In those situations, it would be more useful to have UQ such as conformal prediction for those specific process cycles rather than the whole experiment duration. Platform-wise, both MAPIE and Crepes would be useful as injection molding could be an ideal fit both as a time series and a machine-learning regression problem.

4. Concluding discussion

Here we have addressed two industrial use cases with conformal prediction approaches where data sets represent real-world applications. We demonstrated the application of conformal prediction in these cases and discussed the challenges encountered during experimentation. The main takeaways were that we data scarcity and lack of high-quality data sets are two key issues hindering applying advanced analytics and machine learning in an industrial setup. Some use cases, such as container demand prediction, require data to be aggregated into a weekly or monthly level, limiting the data set's size even further. In addition, relevant product data need to be gathered and synchronizations are needed with various internal supply chain groups.

Unsurprisingly there are some challenges remaining in making conformal prediction approaches into mainstream tools. Unlike traditional machine learning point prediction, there is a lack of industrial-scale mature tools for conformal prediction. Most tools were built for academic purposes during the last few years.

Some concrete improvements needed to enable widespread application of those platforms in real-world scenarios include additional APIs for handling special events, like holidays (similar to NeuralProphet and other libraries). Furthermore, more documentation will be needed to be widely applicable, including tutorials and use case examples suitable for non-academic practitioners, such as engineers and data scientists.

Finally, since conformal prediction has not yet become a mainstream tool among machine learning practitioners, it is hard to set standard guidelines for everyone. Consequently, more academic and industrial collaborations, adopting lifelong learning practices, are needed.

Acknowledgments

This work was based on a collaboration between Husqvarna Group and Jönköping University. The authors acknowledge Husqvarna for providing data sets and time for Nasir Uddin's work. The authors also acknowledge the Swedish Knowledge Foundation, Jönköping University, and other industrial partners for financially supporting the research and education environment on Knowledge Intensive Product Realization SPARK at Jönköping University, Sweden. Project: AFAIR with agreement number 20200223 and PREMACOP with agreement number 20220187.

References

Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297, 2021. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2021.05.008. URL https://www.sciencedirect.com/science/article/pii/S1566253521001081.

Girish Agarwal. *Revising Business Model Innovation: Towards a value process framework for AI-based Offerings*. PhD thesis, KTH Royal Institute of Technology, 2022.

Jonathan Alvarsson, Staffan Arvidsson McShane, Ulf Norinder, and Ola Spjuth. Predicting with confidence: using conformal prediction in drug discovery. *Journal of Pharmaceutical Sciences*, 110(1):42–49, 2021.

Md Tahseen Anam. Evaluate machine learning model to better understand cutting in wood, 2021.

Anastasios N. Angelopoulos and Stephen Bates. A gentle introduction to conformal prediction and distribution-free uncertainty quantification, 2021.

Vineeth Balasubramanian, Shen-Shyang Ho, and Vladimir Vovk. *Conformal prediction for reliable machine learning: theory, adaptations and applications*. Newnes, 2014.

Rina Foygel Barber, Emmanuel J Candes, Aaditya Ramdas, and Ryan J Tibshirani. Predictive inference with the jackknife+. 2021.

Henrik Boström. Crepes: a python package for generating conformal regressors and predictive systems. In Ulf Johansson, Henrik Boström, Khuong An Nguyen, Zhiyuan Luo, and Lars Carlsson, editors, *Proceedings of the Eleventh Symposium on Conformal and Probabilistic Prediction and Applications*, volume 179 of *Proceedings of Machine Learning Research*. PMLR, 2022.

Nicolas Dewolf, Bernard De Baets, and Willem Waegeman. Valid prediction intervals for regression problems. *Artificial Intelligence Review*, 56(1):577–613, 2023.

Matteo Fontana, Gianluca Zeni, and Simone Vantini. Conformal prediction: a unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.

Hongbo Fu, Hong Xu, Ying Liu, Zhaogang Yang, S Kormakov, Daming Wu, and Jingyao Sun. Overview of injection molding technology for processing polymers and their composites. *ES Materials & Manufacturing*, 8(4):3–23, 2020.

Tilmann Gneiting and Matthias Katzfuss. Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1:125–151, 2014.

Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007.

Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

Ulf Johansson, Henrik Boström, and Tuwe Löfström. Investigating normalized conformal regressors. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 01–08. IEEE, 2021.

Hail Jung, Jinsu Jeon, Dahui Choi, and Jung-Ywn Park. Application of machine learning techniques in injection molding quality prediction: Implications on sustainable manufacturing industry. *Sustainability*, 13(8):4120, 2021.

Byol Kim, Chen Xu, and Rina Barber. Predictive inference is free with the jackknife+-after-bootstrap. *Advances in Neural Information Processing Systems*, 33:4138–4149, 2020.

Valery Manokhin. *Machine Learning for Probabilistic Prediction*. PhD thesis, Royal Holloway, University of London, 2022a.

Valery Manokhin. Awesome conformal prediction, April 2022b. URL https://doi.org/10.5281/zenodo.6467205. "If you use Awesome Conformal Prediction, please cite it as below.".

Kevin P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023. URL http://probml.github.io/book2.

Richárd Dominik Párizs, Dániel Török, Tatyana Ageyeva, and József Gábor Kovács. Machine learning in injection molding: an industry 4.0 method of quality prediction. *Sensors*, 22(7):2704, 2022.

Vaia Rousopoulou, Alexandros Nizamis, Thanasis Vafeiadis, Dimosthenis Ioannidis, and Dimitrios Tzovaras. Predictive maintenance for injection molding machines enabled by cognitive analytics for industry 4.0. *Frontiers in Artificial Intelligence*, 3:578152, 2020.

Senthil Kumaran Selvaraj, Aditya Raj, R Rishikesh Mahadevan, Utkarsh Chadha, and Velmurugan Paramasivam. A review on machine learning models in injection molding machines. *Advances in Materials Science and Engineering*, 2022:1–28, 2022.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

Johan Simonsson. *Towards Digital Product-Service Platforms in Manufacturing Firms: Exploring Organizational and Managerial Aspects*. PhD thesis, KTH Royal Institute of Technology, 2021.

Johan Simonsson, Mats Magnusson, and Anders Johanson. Organizing the development of digital product-service platforms. *Technology Innovation Management Review*, 10(3), 2020.

Vianney Taquet, Vincent Blot, Thomas Morzadec, Louis Lacombe, and Nicolas Brunel. Mapie: an open-source library for distribution-free uncertainty quantification, 2022. URL https://arxiv.org/abs/2207.12274.

V Vovk, A Gammerman, and C Saunders. Machine-learning applications of algorithmic randomness. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pages 444–453. Morgan Kaufmann, 1999.

Vladimir Vovk and Claus Bendtsen. Conformal predictive decision making. In *Conformal and Probabilistic Prediction and Applications*, pages 52–62. PMLR, 2018.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.

Vladimir Vovk, Ivan Petej, Ilia Nouretdinov, Valery Manokhin, and Alexander Gammerman. Computationally efficient versions of conformal predictive distributions. *Neurocomputing*, 397:292–308, 2020.

Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. Probabilistic regression: Conformal predictive systems. In *Algorithmic Learning in a Random World*, pages 181–224. Springer, 2022.

Di Wang, Ping Wang, Yue Yuan, Pingping Wang, and Junzhi Shi. A fast conformal predictive system with regularized extreme learning machine. *Neural Networks*, 126:347–361, 2020.

Hugo Werner, Lars Carlsson, Ernst Ahlberg, and Henrik Boström. Evaluating different approaches to calibrating conformal predictive systems. In *Conformal and Probabilistic Prediction and Applications*, pages 134–150. PMLR, 2020.

Chen Xu and Yao Xie. Conformal prediction interval for dynamic time-series. In *International Conference on Machine Learning*, pages 11559–11569. PMLR, 2021.

Chen Xu and Yao Xie. Conformal prediction for time-series, 2022.

Michaël Zamo and Philippe Naveau. Estimation of the continuous ranked probability score with limited information and applications to ensemble weather forecasts. *Mathematical Geosciences*, 50(2):209–234, 2018.