Distribution-free uncertainty quantification (for time series)

Margaux Zaffran

28/06/2022

Mathematical Methods of Modern Statistics, best city in the world *(a.k.a. Marseille)*

Aymeric Dieuleveut

Ecole Polytechnique
Paris

Olivier Féron

EDF R&D
FiME
Paris

Yannig Goude

EDF R&D
LMO
Paris

Julie Josse

INRIA
IDESP
Montpellier

Introduction to Split Conformal Prediction

Setting

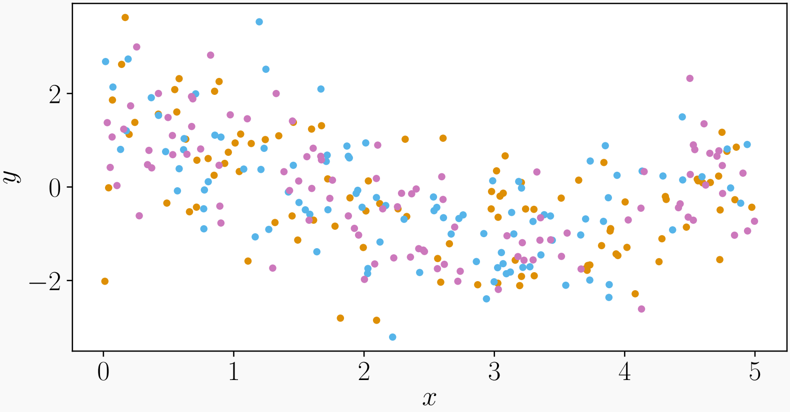
- $(x, y) \in \mathbb{R}^d \times \mathbb{R}$ realization of random variable (X, Y)
- n training samples $(x_i, y_i)_{i=1}^n$
- Goal: predict an unseen point y_{n+1} at x_{n+1} with **confidence**
- Miscoverage level $\alpha \in [0, 1]$

▶ Build a predictive interval $\mathcal{C}_\alpha$ such that:

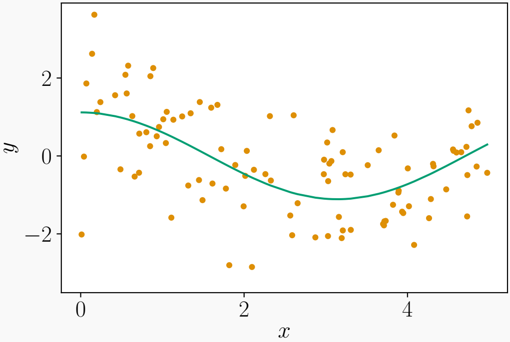
$$\mathbb{P}\left\{Y_{n+1} \in \mathcal{C}_\alpha\left(X_{n+1}\right)\right\} \geq 1 - \alpha, \tag{1}$$

and $\mathcal{C}_\alpha$ should be as small as possible, in order to be informative.

Split conformal prediction: toy example

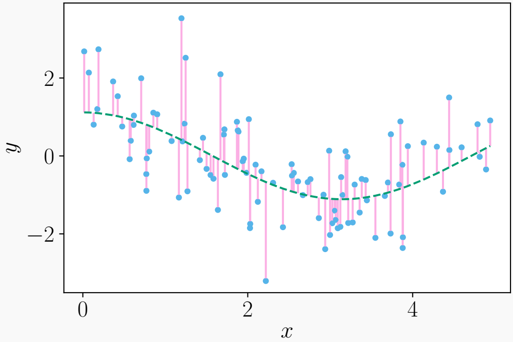


Split conformal prediction: training step



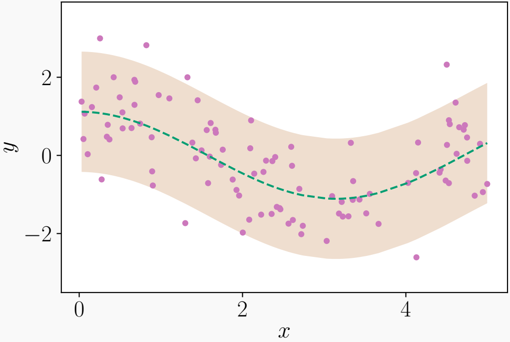
- Learn $\hat{\mu}$

Split conformal prediction: calibration step



- Predict with $\hat{\mu}$
- Get the residuals $\hat{\varepsilon}_i$
- Compute the $(1-\alpha)$ empirical quantile of the $|\hat{\varepsilon}_i|$, noted $q_{1-\alpha}(|\hat{\varepsilon}_i|)$

Split conformal prediction: prediction step



- Predict with $\hat{\mu}$
- Build $\hat{\mathcal{C}}_\alpha(x)$:
  $[\hat{\mu}(x) \pm q_{1-\alpha}(|\hat{\varepsilon}_i|)]$

Split Conformal Prediction: method

1. Split randomly your training data into a proper training set (size n_{train}) and a calibration set (size n_{cal})

Split Conformal Prediction: method

1. Split randomly your training data into a proper training set (size n_{train}) and a calibration set (size n_{cal})
2. Train your algorithm on your proper training set

Split Conformal Prediction: method

1. Split randomly your training data into a proper training set (size n_{train}) and a calibration set (size n_{cal})
2. Train your algorithm on your proper training set
3. Get predictions with this fitted model $\hat{A}$ on the calibration set

Split Conformal Prediction: method

1. Split randomly your training data into a proper training set (size n_{train}) and a calibration set (size n_{cal})
2. Train your algorithm on your proper training set
3. Get predictions with this fitted model $\hat{A}$ on the calibration set
4. Assign a conformity score (the smaller the better) to each prediction ($s(\hat{A}(x_i), y_i)$)

Split Conformal Prediction: method

1. Split randomly your training data into a proper training set (size n_{train}) and a calibration set (size n_{cal})
2. Train your algorithm on your proper training set
3. Get predictions with this fitted model $\hat{A}$ on the calibration set
4. Assign a conformity score (the smaller the better) to each prediction ($s(\hat{A}(x_i), y_i)$) $\Rightarrow$ obtain a set of n_{cal} conformity scores $S = \{s_i = s(\hat{A}(x_i), y_i),\ \text{for } i \text{ such that } (x_i, y_i) \in \text{Cal}\}$

Split Conformal Prediction: method

1. Split randomly your training data into a proper training set (size n_{train}) and a calibration set (size n_{cal})
2. Train your algorithm on your proper training set
3. Get predictions with this fitted model $\hat{A}$ on the calibration set
4. Assign a conformity score (the smaller the better) to each prediction $(s(\hat{A}(x_i), y_i)) \Rightarrow$ obtain a set of n_{cal} conformity scores $S = \{s_i = s(\hat{A}(x_i), y_i),\ \text{for } i \text{ such that } (x_i, y_i) \in \text{Cal}\}$
5. Compute the $1 - \alpha$[1] quantile of these scores, noted $q_{1-\alpha}(S)$

[1] a correction is needed to ensure finite sample guarantees: the $(1 - \alpha)(1 + \frac{1}{n_{\text{cal}}})$ quantile is used instead.

Split Conformal Prediction: method

1. Split randomly your training data into a proper training set (size n_{train}) and a calibration set (size n_{cal})
2. Train your algorithm on your proper training set
3. Get predictions with this fitted model $\hat{A}$ on the calibration set
4. Assign a conformity score (the smaller the better) to each prediction $(s(\hat{A}(x_i), y_i)) \Rightarrow$ obtain a set of n_{cal} conformity scores $S = \{s_i = s(\hat{A}(x_i), y_i), \text{ for } i \text{ such that } (x_i, y_i) \in \text{Cal}\}$
5. Compute the $1 - \alpha$[1] quantile of these scores, noted $q_{1-\alpha}(S)$
6. For a new point x_{n+1}, output $\mathcal{C}_{\alpha}(x_{n+1}) = \{y \text{ such that } s(\hat{A}(x_{n+1}), y) \leq q_{1-\alpha}(S)\}$

[1] a correction is needed to ensure finite sample guarantees: the $(1-\alpha)(1+\frac{1}{n_{\text{cal}}})$ quantile is used instead.

Split Conformal Prediction: beyond the toy example

$$\mathcal{C}_{\alpha}(x_{n+1}) = \{y \text{ such that } s(\hat{A}(x_{n+1}), y) \leq q_{1-\alpha}(S)\}$$

- In regression, if $s(\hat{A}(x_i), y) = |\hat{y}_i - y|$, then $\mathcal{C}_{\alpha}(x_{n+1}) = [\hat{A}(x_{n+1}) \pm q_{1-\alpha}(S)]$

Split Conformal Prediction: beyond the toy example

$$\mathcal{C}_{\alpha}(x_{n+1}) = \{y \text{ such that } s(\hat{A}(x_{n+1}), y) \leq q_{1-\alpha}(S)\}$$

- In regression, if $s(\hat{A}(x_i), y) = |\hat{y}_i - y|$, then $\mathcal{C}_{\alpha}(x_{n+1}) = [\hat{A}(x_{n+1}) \pm q_{1-\alpha}(S)]$
- Extension: $s(\hat{A}(x_i), y) = \dfrac{|\hat{y}_i - y|}{\hat{\sigma}(x_i)}$ leading to $\mathcal{C}_{\alpha}(x_{n+1}) = [\hat{y}_{n+1} \pm \hat{\sigma}(x_{n+1}) \times q_{1-\alpha}(S)]$ Lei et al. (2018)

Split Conformal Prediction: beyond the toy example

$$\mathcal{C}_{\alpha}(x_{n+1}) = \{y \text{ such that } s(\hat{A}(x_{n+1}), y) \leq q_{1-\alpha}(S)\}$$

- In regression, if $s(\hat{A}(x_i), y) = |\hat{y}_i - y|$, then $\mathcal{C}_{\alpha}(x_{n+1}) = [\hat{A}(x_{n+1}) \pm q_{1-\alpha}(S)]$
- Extension: $s(\hat{A}(x_i), y) = \dfrac{|\hat{y}_i - y|}{\hat{\sigma}(x_i)}$ leading to $\mathcal{C}_{\alpha}(x_{n+1}) = [\hat{y}_{n+1} \pm \hat{\sigma}(x_{n+1}) \times q_{1-\alpha}(S)]$ Lei et al. (2018)
- Beyond mean regression: Romano et al. (2019), many others

Split Conformal Prediction: beyond the toy example

$$\mathcal{C}_{\alpha}(x_{n+1}) = \{y \text{ such that } s(\hat{A}(x_{n+1}), y) \leq q_{1-\alpha}(S)\}$$

- In regression, if $s(\hat{A}(x_i), y) = |\hat{y}_i - y|$, then $\mathcal{C}_{\alpha}(x_{n+1}) = [\hat{A}(x_{n+1}) \pm q_{1-\alpha}(S)]$
- Extension: $s(\hat{A}(x_i), y) = \dfrac{|\hat{y}_i - y|}{\hat{\sigma}(x_i)}$ leading to $\mathcal{C}_{\alpha}(x_{n+1}) = [\hat{y}_{n+1} \pm \hat{\sigma}(x_{n+1}) \times q_{1-\alpha}(S)]$
Lei et al. (2018)
- Beyond mean regression: Romano et al. (2019), many others
- Classification

Split Conformal Prediction: beyond the toy example

$$\mathcal{C}_\alpha(x_{n+1}) = \{y \text{ such that } s(\hat{A}(x_{n+1}), y) \leq q_{1-\alpha}(S)\}$$

- In regression, if $s(\hat{A}(x_i), y) = |\hat{y}_i - y|$, then $\mathcal{C}_\alpha(x_{n+1}) = [\hat{A}(x_{n+1}) \pm q_{1-\alpha}(S)]$
- Extension: $s(\hat{A}(x_i), y) = \dfrac{|\hat{y}_i - y|}{\hat{\sigma}(x_i)}$ leading to $\mathcal{C}_\alpha(x_{n+1}) = [\hat{y}_{n+1} \pm \hat{\sigma}(x_{n+1}) \times q_{1-\alpha}(S)]$ Lei et al. (2018)
- Beyond mean regression: Romano et al. (2019), many others
- Classification

$\hookrightarrow$ what is important is the definition of the conformity scores

Conformal prediction: theoretical guarantees

This procedure enjoys finite sample guarantee proposed and proved in Vovk et al. (2005) and Lei et al. (2018).

Theorem
Suppose $(X_i, Y_i)_{i=1}^{n+1}$ *are exchangeable, and we apply split conformal prediction on* $(X_i, Y_i)_{i=1}^{n}$ *to predict an interval on* X_{n+1}, $\hat{\mathcal{C}}_\alpha(X_{n+1})$. *Then we have:*

$$\mathbb{P}\left\{Y_{n+1} \in \hat{\mathcal{C}}_\alpha(X_{n+1})\right\} \geq 1 - \alpha.$$

If, in addition, the scores $\hat{\varepsilon}_j$ *have a continuous joint distribution, we also have an upper bound:*

$$\mathbb{P}\left\{Y_{n+1} \in \hat{\mathcal{C}}_\alpha(X_{n+1})\right\} \leq 1 - \alpha + \frac{1}{n_{cal} + 1}.$$

Conformal prediction: summary

Split conformal prediction is simple to compute and works:

- any regression algorithm (neural nets, random forest...);
- distribution-free as long as the data is exchangeable;

- finite sample.

Conformal prediction: summary

Split conformal prediction is simple to compute and works:

- any regression algorithm (neural nets, random forest...);
- distribution-free as long as the data is exchangeable;
  - $\hookrightarrow$ the scores need to be exchangeable (but then it would not work with any regression algorithm)
- finite sample.

Conformal prediction: summary

Split conformal prediction is simple to compute and works:

- any regression algorithm (neural nets, random forest...);
- distribution-free as long as the data is exchangeable;
  - $\hookrightarrow$ the scores need to be exchangeable (but then it would not work with any regression algorithm)
- finite sample.

Two interests:

- quantify the uncertainty of the underlying model $\hat{\mu}$
- output predictive regions

Conformal prediction and time series, what's the issue?

Framework and notations

- Data: T_0 observations $(x_1, y_1), \ldots, (x_{T_0}, y_{T_0})$ in $\mathbb{R}^d \times \mathbb{R}$
- Aim: predict the response values as well as predictive intervals for T_1 subsequent observations $x_{T_0+1}, \ldots, x_{T_0+T_1}$

$\hookrightarrow$ Build the smallest interval $\mathcal{C}_\alpha^t$ such that:

$$\mathbb{P}\left\{Y_t \in \mathcal{C}_\alpha^t\left(X_t\right)\right\} \geq 1 - \alpha, \text{ for } t \in [\![T_0 + 1, T_0 + T_1]\!].$$

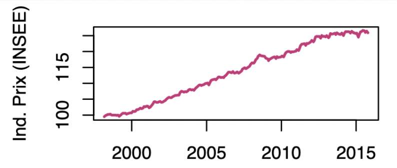


Figure 1: Trend[2]

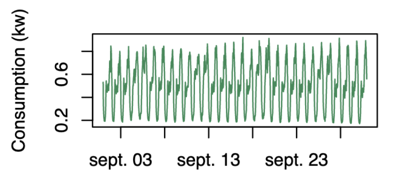


Figure 2: Seasonality[2]

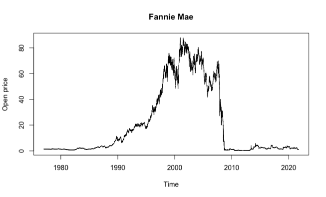


Figure 3: Shift

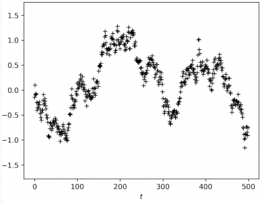


Figure 4: Time dependence

[2]Images from Yannig Goude class material.

Non-exchangeable even if the noise is exchangeable

Assume the following model:

$$Y_t = f_t(X_t) + \varepsilon_t, \text{ for } t \in \mathbb{N}^*,$$

for some function f_t, and some noise ε_t.

Even if the noise ε_t is exchangeable, we can produce dependent residuals.

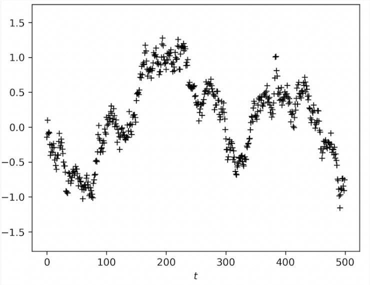


Figure 5: Auto-Regressive residuals

Adaptive Conformal Inference

Online sequential split conformal prediction (OSSCP)

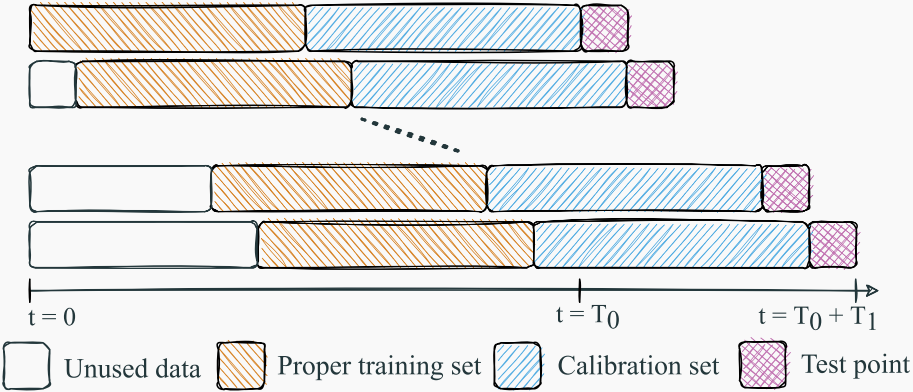


Figure 6: Diagram describing the online sequential split conformal prediction.

Adaptive Conformal Inference (ACI), Gibbs and Candès (2021)

Refitting the model may be insufficient $\Rightarrow$ adapt the quantile level used on the calibration's scores.

Adaptive Conformal Inference (ACI), Gibbs and Candès (2021)

Refitting the model may be insufficient $\Rightarrow$ adapt the quantile level used on the calibration's scores. **(Distribution shift)**

Adaptive Conformal Inference (ACI), Gibbs and Candès (2021)

Refitting the model may be insufficient $\Rightarrow$ adapt the quantile level used on the calibration's scores. **(Distribution shift)**

The proposed update scheme is the following:

$$\alpha_{t+1} := \alpha_t + \gamma \left(\alpha - \mathrm{err}_t \right) \tag{2}$$

with:

$$\mathrm{err}_t := \begin{cases} 1 \text{ if } y_t \notin \hat{\mathcal{C}}_{\alpha_t}(x_t), \\ 0 \text{ otherwise }, \end{cases}$$

and $\alpha_1 = \alpha$, $\gamma \geq 0$.

Adaptive Conformal Inference (ACI), Gibbs and Candès (2021)

Refitting the model may be insufficient ⇒ adapt the quantile level used on the calibration's scores. **(Distribution shift)**

The proposed update scheme is the following:

$$\alpha_{t+1} := \alpha_t + \gamma \left(\alpha - \mathrm{err}_t \right) \tag{2}$$

with:

$$\mathrm{err}_t := \begin{cases} 1 \text{ if } y_t \notin \hat{\mathcal{C}}_{\alpha_t}(x_t), \\ 0 \text{ otherwise }, \end{cases}$$

and $\alpha_1 = \alpha$, $\gamma \geq 0$.

Intuition: if we did make an error, the interval was too small so we want to increase its length by taking a higher quantile (a smaller α_t). Reversely if we included the point.

Adaptive Conformal Inference (ACI), Gibbs and Candès (2021)

Refitting the model may be insufficient $\Rightarrow$ adapt the quantile level used on the calibration's scores. **(Distribution shift)**

The proposed update scheme is the following:

$$\alpha_{t+1} := \alpha_t + \gamma \left(\alpha - \mathrm{err}_t \right) \tag{2}$$

with:

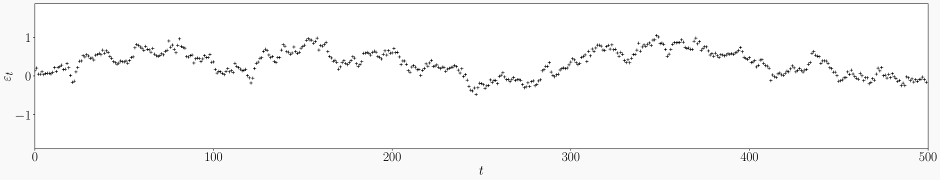
$$\mathrm{err}_t := \begin{cases} 1 \text{ if } y_t \notin \hat{\mathcal{C}}_{\alpha_t}\left(x_t\right), \\ 0 \text{ otherwise }, \end{cases}$$

and $\alpha_1 = \alpha$, $\gamma \geq 0$.

Intuition: if we did make an error, the interval was too small so we want to increase its length by taking a higher quantile (a smaller α_t). Reversely if we included the point.

Gibbs and Candès (2021) provide asymptotic validity result for any distribution.

Visualisation of the procedure



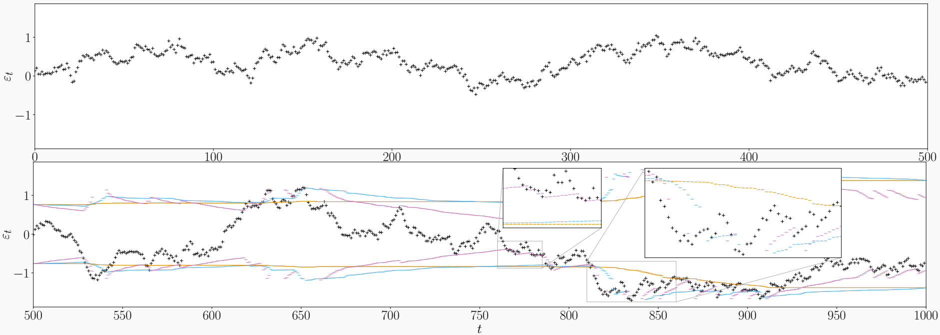


Figure 7: Visualisation of ACI with different values of γ ($\gamma = 0$, $\gamma = 0.01$, $\gamma = 0.05$)

Concluding remarks

Contributions and messages

- Analysis of ACI in the context of time series

Contributions and messages

- Analysis of ACI in the context of time series
- Theoretical and numerical analysis of the impact of γ in the length of the resulting intervals

Contributions and messages

- Analysis of ACI in the context of time series
- Theoretical and numerical analysis of the impact of γ in the length of the resulting intervals
- Empirical proposition of an adaptive choice of γ: `AgACI`

Contributions and messages

- Analysis of ACI in the context of time series
- Theoretical and numerical analysis of the impact of γ in the length of the resulting intervals
- Empirical proposition of an adaptive choice of γ: `AgACI`
- Extensive synthetic experiments on increasing temporal dependence (AgACI, various ACI, benchmark methods)

Contributions and messages

- Analysis of ACI in the context of time series
- Theoretical and numerical analysis of the impact of γ in the length of the resulting intervals
- Empirical proposition of an adaptive choice of γ: `AgACI`
- Extensive synthetic experiments on increasing temporal dependence (AgACI, various ACI, benchmark methods)
- Application to forecasting French electricity spot prices

Contributions and messages

- Analysis of ACI in the context of time series
- Theoretical and numerical analysis of the impact of γ in the length of the resulting intervals
- Empirical proposition of an adaptive choice of γ: `AgACI`
- Extensive synthetic experiments on increasing temporal dependence (AgACI, various ACI, benchmark methods)
- Application to forecasting French electricity spot prices

$\hookrightarrow$ **Perspective:** refined analysis of `AgACI` and expert aggregation

Contributions and messages

- Analysis of ACI in the context of time series
- Theoretical and numerical analysis of the impact of γ in the length of the resulting intervals
- Empirical proposition of an adaptive choice of γ: `AgACI`
- Extensive synthetic experiments on increasing temporal dependence (AgACI, various ACI, benchmark methods)
- Application to forecasting French electricity spot prices

$\hookrightarrow$ **Perspective:** refined analysis of `AgACI` and expert aggregation

- Theoretical guarantees about validity: *what happens to the asymptotic result when aggregated?*
- Analysis of the obtained efficiency
- More data sets

Thank you! Questions?

References i

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge University Press.

Gibbs, I. and Candès, E. (2021). Adaptive Conformal Inference Under Distribution Shift. *Advances in Neural Information Processing Systems*, 34.

Kath, C. and Ziel, F. (2021). Conformal prediction interval estimation and applications to day-ahead and intraday power markets. *International Journal of Forecasting*.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*.

Papadopoulos, H., Proedrou, K., Vovk, V., and Gammerman, A. (2002). Inductive Confidence Machines for Regression. In *Machine Learning: ECML 2002*. Springer.

Romano, Y., Patterson, E., and Candès, E. (2019). Conformalized Quantile Regression. *Advances in Neural Information Processing Systems*, 32.

Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer US.

Wisniewski, W., Lindsay, D., and Lindsay, S. (2020). Application of conformal prediction interval estimations to market makers' net positions. Proceedings of Machine Learning Research. PMLR.

Xu, C. and Xie, Y. (2021). Conformal prediction interval for dynamic time-series. In *Proceedings of the 38th International Conference on Machine Learning*. PMLR.

Zaffran, M., Féron, O., Goude, Y., Josse, J., and Dieuleveut, A. (2022). Adaptive Conformal Predictions for Time Series. arXiv:2202.07282.

Exchangeability

Definition (Exchangeability)

$(Z_i)_{i=1}^n$ are exchangeable if for any permutation σ of $[1, n]$ we have:

$$\mathcal{L}\left(Z_1, \ldots, Z_n\right) = \mathcal{L}\left(Z_{\sigma(1)}, \ldots, Z_{\sigma(n)}\right),$$

where $\mathcal{L}$ designates the joint distribution.

AgACI

AgACI: adaptive wrapper around ACI, setting

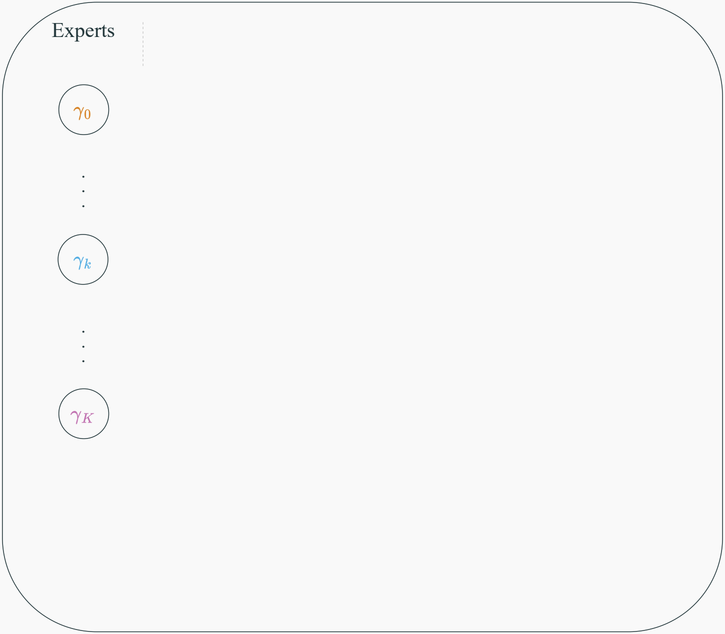
Online aggregation under expert advice (Cesa-Bianchi and Lugosi, 2006) computes an optimal weighted mean of experts.

AgACI: adaptive wrapper around ACI, setting

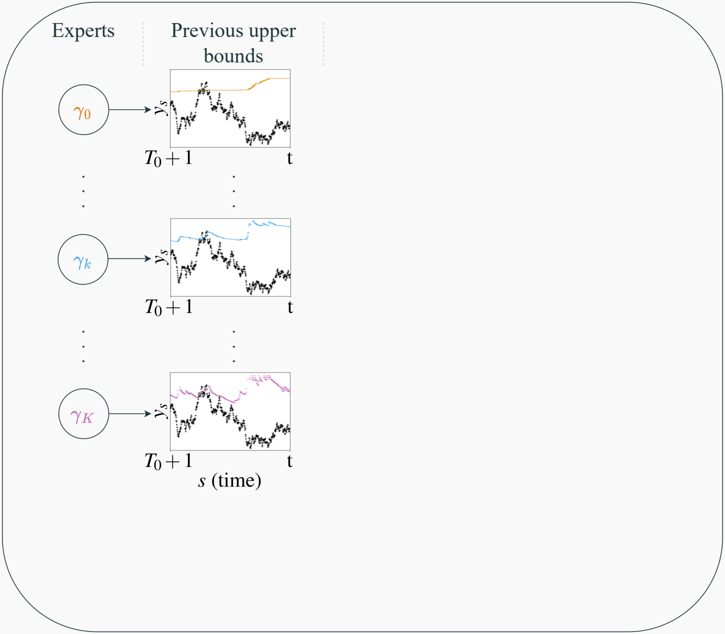
Online aggregation under expert advice (Cesa-Bianchi and Lugosi, 2006) computes an optimal weighted mean of experts.

`AgACI` performs 2 independent aggregations: one for each bound (the upper and lower ones).

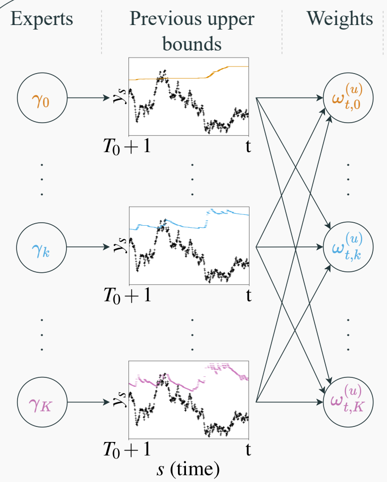
AgACI: adaptive wrapper around ACI



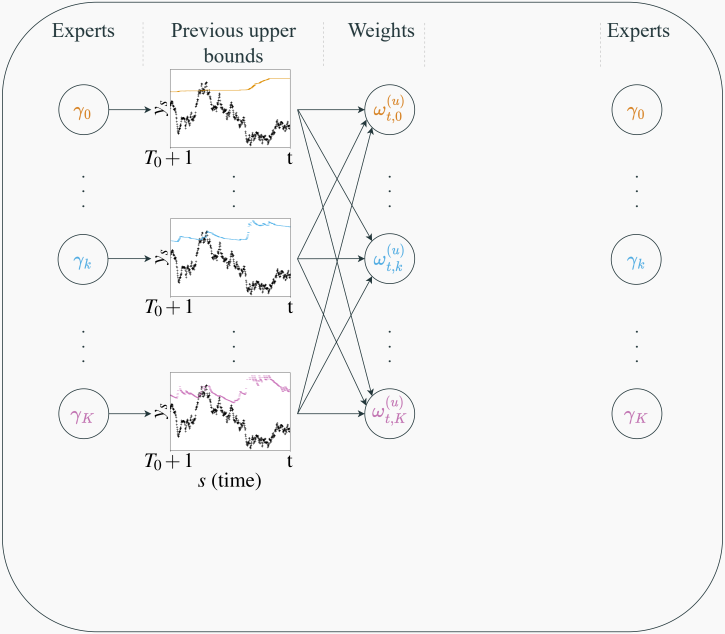
AgACI: adaptive wrapper around ACI



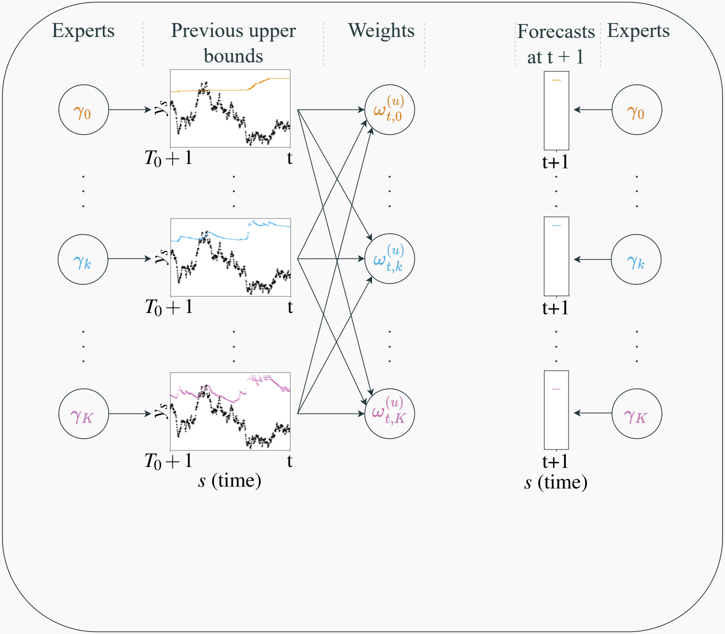
AgACI: adaptive wrapper around ACI



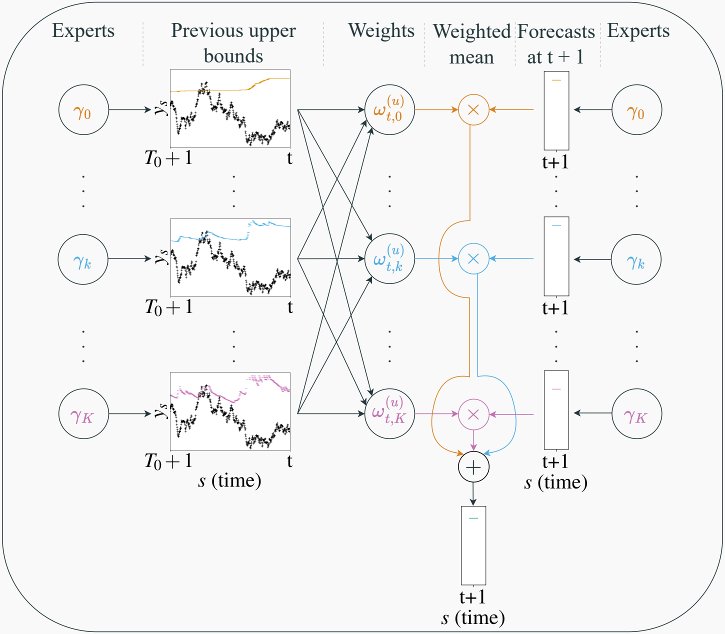
AgACI: adaptive wrapper around ACI



AgACI: adaptive wrapper around ACI



AgACI: adaptive wrapper around ACI



Comparison on simulated data

Data generation and simulation settings

$$Y_t = 10 \sin\left(\pi X_{t,1} X_{t,2}\right) + 20\left(X_{t,3} - 0.5\right)^2 + 10 X_{t,4} + 5 X_{t,5} + \varepsilon_t$$

Data generation and simulation settings

$$Y_t = 10 \sin\left(\pi X_{t,1} X_{t,2}\right) + 20 \left(X_{t,3} - 0.5\right)^2 + 10 X_{t,4} + 5 X_{t,5} + \varepsilon_t$$

where the $X_{t,\cdot} \sim \mathcal{U}([0,1])$

Data generation and simulation settings

$$Y_t = 10 \sin(\pi X_{t,1} X_{t,2}) + 20 (X_{t,3} - 0.5)^2 + 10 X_{t,4} + 5 X_{t,5} + \varepsilon_t$$

where the $X_{t,\cdot} \sim \mathcal{U}([0,1])$ and ε_t is an ARMA(1,1) process:

Data generation and simulation settings

$$Y_t = 10\sin(\pi X_{t,1}X_{t,2}) + 20(X_{t,3} - 0.5)^2 + 10X_{t,4} + 5X_{t,5} + \varepsilon_t$$

where the $X_{t,\cdot} \sim \mathcal{U}([0,1])$ and ε_t is an ARMA(1,1) process:

$$\varepsilon_{t+1} = \varphi\varepsilon_t + \xi_{t+1} + \theta\xi_t,$$

with ξ_t is a white noise of variance σ^2.

Data generation and simulation settings

$$Y_t = 10\sin(\pi X_{t,1} X_{t,2}) + 20(X_{t,3} - 0.5)^2 + 10X_{t,4} + 5X_{t,5} + \varepsilon_t$$

where the $X_{t,\cdot} \sim \mathcal{U}([0,1])$ and ε_t is an ARMA(1,1) process:

$$\varepsilon_{t+1} = \varphi\varepsilon_t + \xi_{t+1} + \theta\xi_t,$$

with ξ_t is a white noise of variance σ^2.

- $\varphi = \theta$ range in $[0.1, 0.8, 0.9, 0.95, 0.99]$.

Data generation and simulation settings

$$Y_t = 10\sin(\pi X_{t,1}X_{t,2}) + 20\,(X_{t,3} - 0.5)^2 + 10X_{t,4} + 5X_{t,5} + \varepsilon_t$$

where the $X_{t,\cdot} \sim \mathcal{U}([0,1])$ and ε_t is an ARMA(1,1) process:

$$\varepsilon_{t+1} = \varphi\varepsilon_t + \xi_{t+1} + \theta\xi_t,$$

with ξ_t is a white noise of variance σ^2.

- $\varphi = \theta$ range in $[0.1, 0.8, 0.9, 0.95, 0.99]$.
- We fix σ to keep the variance $\mathrm{Var}(\varepsilon_t)$ constant to 10 (or 1).

Data generation and simulation settings

$$Y_t = 10\sin\left(\pi X_{t,1} X_{t,2}\right) + 20\left(X_{t,3} - 0.5\right)^2 + 10X_{t,4} + 5X_{t,5} + \varepsilon_t$$

where the $X_{t,\cdot} \sim \mathcal{U}([0,1])$ and ε_t is an ARMA(1,1) process:

$$\varepsilon_{t+1} = \varphi\varepsilon_t + \xi_{t+1} + \theta\xi_t,$$

with ξ_t is a white noise of variance σ^2.

- $\varphi = \theta$ range in $[0.1, 0.8, 0.9, 0.95, 0.99]$.
- We fix σ to keep the variance $\mathrm{Var}(\varepsilon_t)$ constant to 10 (or 1).
- We use random forest as regressor.

Data generation and simulation settings

$$Y_t = 10\sin(\pi X_{t,1}X_{t,2}) + 20(X_{t,3} - 0.5)^2 + 10X_{t,4} + 5X_{t,5} + \varepsilon_t$$

where the $X_{t,\cdot} \sim \mathcal{U}([0,1])$ and ε_t is an ARMA(1,1) process:

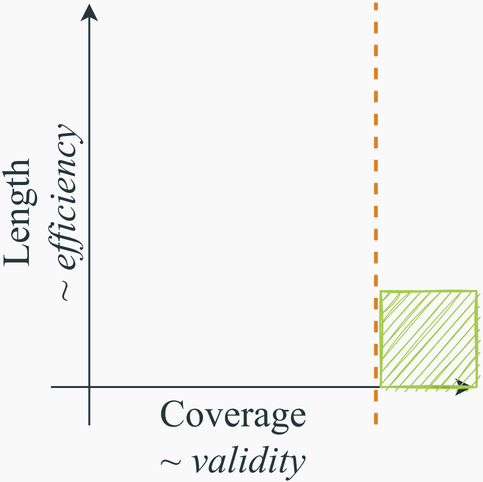
$$\varepsilon_{t+1} = \varphi\varepsilon_t + \xi_{t+1} + \theta\xi_t,$$

with ξ_t is a white noise of variance σ^2.

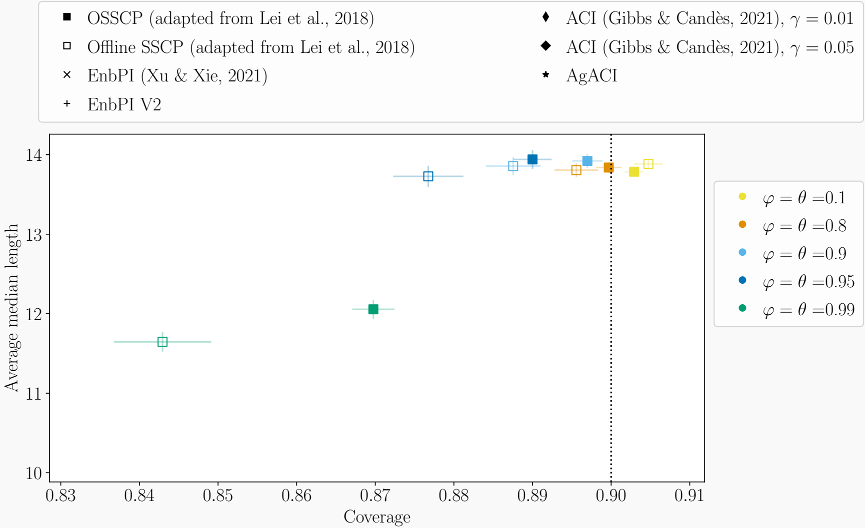
- $\varphi = \theta$ range in $[0.1, 0.8, 0.9, 0.95, 0.99]$.
- We fix σ to keep the variance Var(ε_t) constant to 10 (or 1).
- We use random forest as regressor.
- For each setting (pair variance and φ,θ):
  - 300 points, the last 100 kept for prediction and evaluation,
  - 500 repetitions,

$\Rightarrow$ in total, $100 \times 500 = 50000$ predictions are evaluated.

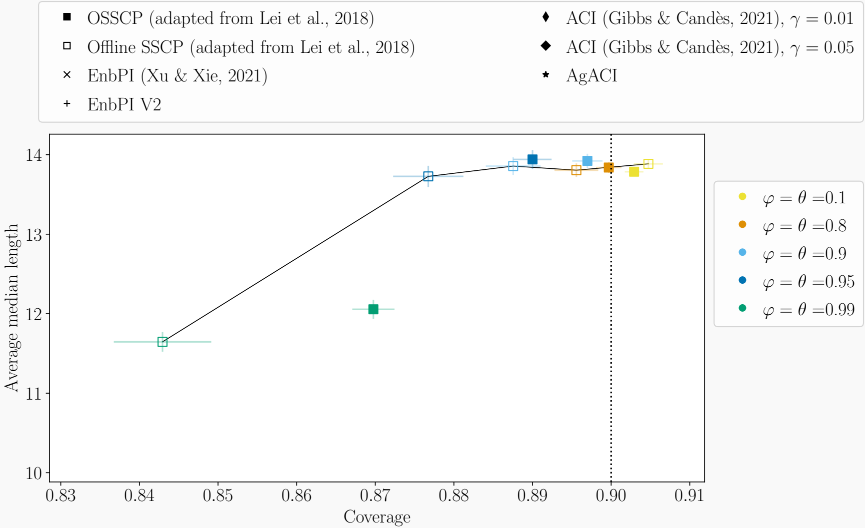
Visualisation of the results



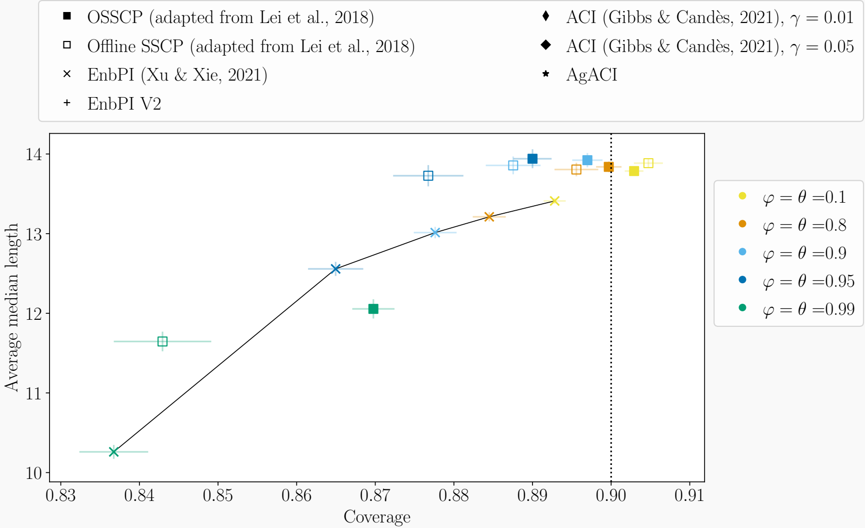
Results: impact of the temporal dependence, ARMA(1,1), variance 10



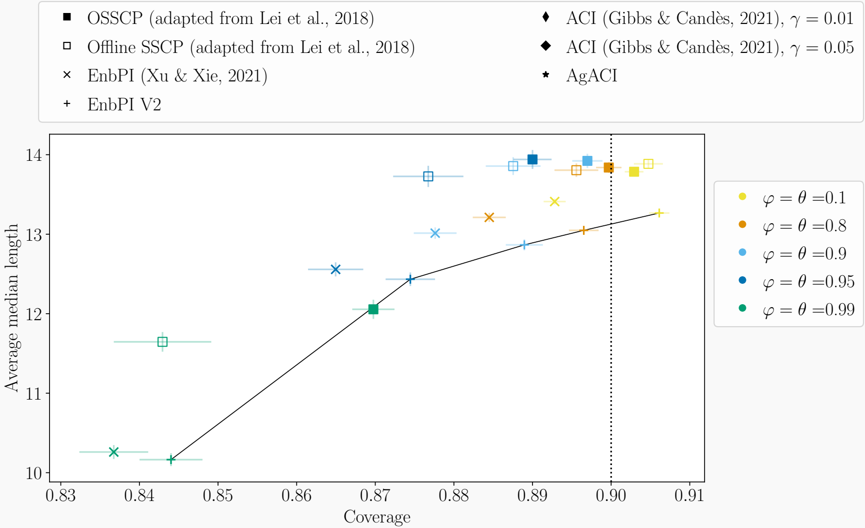
Results: impact of the temporal dependence, ARMA(1,1), variance 10



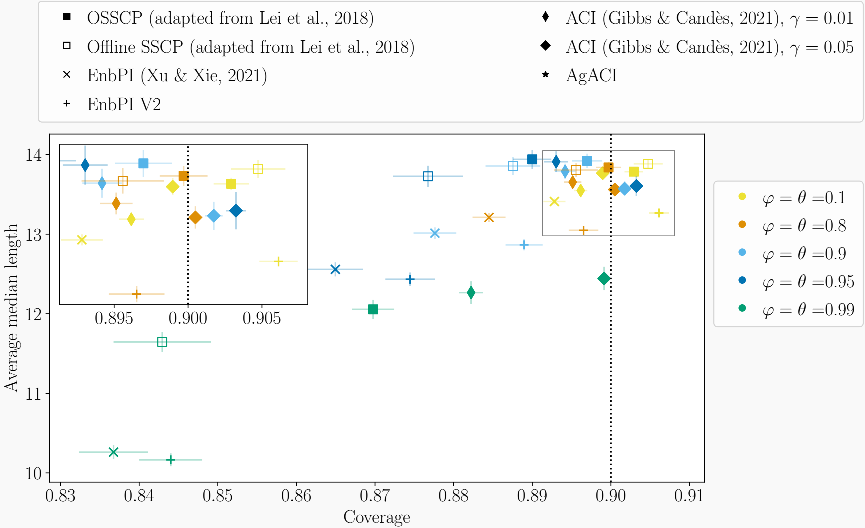
Results: impact of the temporal dependence, ARMA(1,1), variance 10



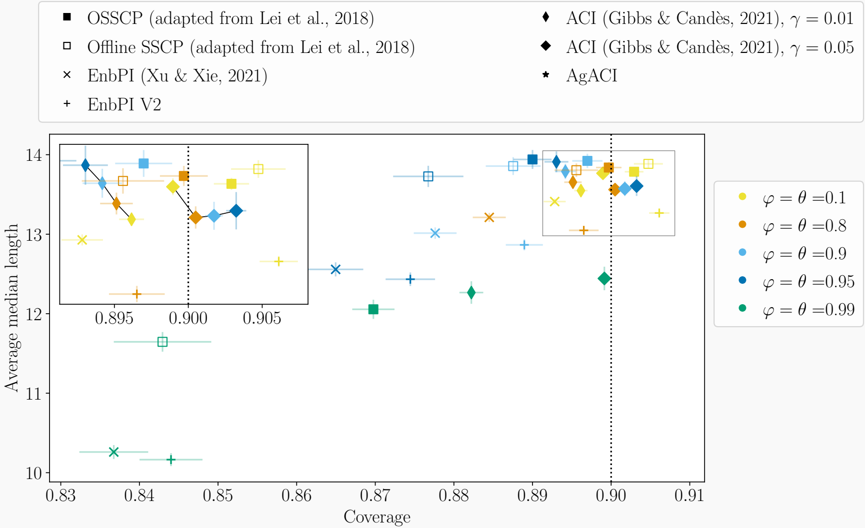
Results: impact of the temporal dependence, ARMA(1,1), variance 10



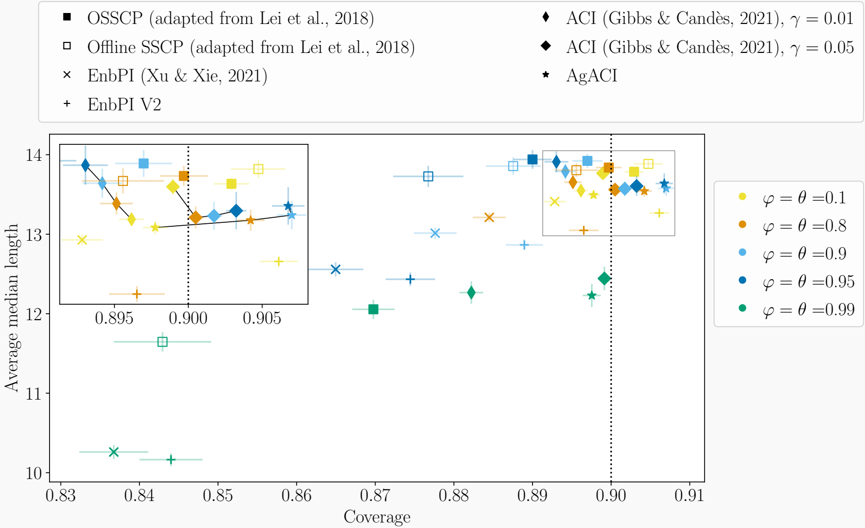
Results: impact of the temporal dependence, ARMA(1,1), variance 10



Results: impact of the temporal dependence, ARMA(1,1), variance 10



Results: impact of the temporal dependence, ARMA(1,1), variance 10



Theoretical analysis of ACI's length

Approach

Aim: derive theoretical results on the **average length** of ACI depending on γ

$\hookrightarrow$ Guideline for choosing γ

Approach

Aim: derive theoretical results on the **average length** of ACI depending on γ

$\hookrightarrow$ Guideline for choosing γ

Approach: consider extreme cases (useful in an adversarial context) with simple theoretical distributions (additional assumptions)

1. exchangeable
2. Auto-Regressive case (AR(1))

Theoretical analysis of ACI's length: exchangeable case

Define $L(\alpha_t) = 2Q(1-\alpha_t)$ the length of the interval predicted by the adaptive algorithm at time t, and $L_0 = 2Q(1-\alpha)$ the length of the interval predicted by the non-adaptive algorithm ($\gamma = 0$).

Theorem

Assume the scores are exchangeable with quantile function Q perfectly estimated at each time, and other assumptions.

Then, for all $\gamma > 0$, $(\alpha_t)_{t>0}$ forms a Markov Chain, that admits a stationary distribution π_γ, and

$$\frac{1}{T}\sum_{t=1}^{T} L(\alpha_t) \xrightarrow[T\to+\infty]{a.s.} \mathbb{E}_{\pi_\gamma}[L] \overset{not.}{=} \mathbb{E}_{\tilde{\alpha}\sim\pi_\gamma}[L(\tilde{\alpha})].$$

Moreover, as $\gamma \to 0$,

$$\mathbb{E}_{\pi_\gamma}[L] = L_0 + Q''(1-\alpha)\frac{\gamma}{2}\alpha(1-\alpha) + O(\gamma^{3/2}).$$

Theoretical analysis of ACI's length: exchangeable case

Define $L(\alpha_t) = 2Q(1-\alpha_t)$ the length of the interval predicted by the adaptive algorithm at time t, and $L_0 = 2Q(1-\alpha)$ the length of the interval predicted by the non-adaptive algorithm ($\gamma = 0$).

Theorem

Assume the scores are exchangeable with quantile function Q perfectly estimated at each time, and other assumptions.

Then, for all $\gamma > 0$, $(\alpha_t)_{t>0}$ forms a Markov Chain, that admits a stationary distribution π_γ, and

$$\frac{1}{T}\sum_{t=1}^{T} L(\alpha_t) \xrightarrow[T\to+\infty]{a.s.} \mathbb{E}_{\pi_\gamma}[L] \overset{not.}{=} \mathbb{E}_{\tilde{\alpha}\sim\pi_\gamma}[L(\tilde{\alpha})].$$

Moreover, as $\gamma \to 0$,

$$\mathbb{E}_{\pi_\gamma}[L] = L_0 + Q''(1-\alpha)\frac{\gamma}{2}\alpha(1-\alpha) + O(\gamma^{3/2}).$$

Theoretical analysis of ACI's length: exchangeable case

Define $L(\alpha_t) = 2Q(1-\alpha_t)$ the length of the interval predicted by the adaptive algorithm at time t, and $L_0 = 2Q(1-\alpha)$ the length of the interval predicted by the non-adaptive algorithm ($\gamma = 0$).

Theorem

Assume the scores are exchangeable with quantile function Q perfectly estimated at each time, and other assumptions.

Then, for all $\gamma > 0$, $(\alpha_t)_{t>0}$ forms a Markov Chain, that admits a stationary distribution π_γ, and

$$\frac{1}{T}\sum_{t=1}^{T} L(\alpha_t) \xrightarrow[T\to+\infty]{a.s.} \mathbb{E}_{\pi_\gamma}[L] \overset{not.}{=} \mathbb{E}_{\tilde{\alpha}\sim\pi_\gamma}[L(\tilde{\alpha})].$$

Moreover, as $\gamma \to 0$,

$$\mathbb{E}_{\pi_\gamma}[L] = L_0 + Q''(1-\alpha)\frac{\gamma}{2}\alpha(1-\alpha) + O(\gamma^{3/2}).$$

Theoretical analysis of ACI's length: exchangeable case

Define $L(\alpha_t) = 2Q(1-\alpha_t)$ the length of the interval predicted by the adaptive algorithm at time t, and $L_0 = 2Q(1-\alpha)$ the length of the interval predicted by the non-adaptive algorithm ($\gamma = 0$).

Theorem

Assume the scores are exchangeable with quantile function Q perfectly estimated at each time, and other assumptions.

Then, for all $\gamma > 0$, $(\alpha_t)_{t>0}$ forms a Markov Chain, that admits a stationary distribution π_γ, and

$$\frac{1}{T}\sum_{t=1}^{T} L(\alpha_t) \xrightarrow[T\to+\infty]{a.s.} \mathbb{E}_{\pi_\gamma}[L] \overset{not.}{=} \mathbb{E}_{\tilde{\alpha}\sim\pi_\gamma}[L(\tilde{\alpha})].$$

Moreover, as $\gamma \to 0$,

$$\mathbb{E}_{\pi_\gamma}[L] = L_0 + Q''(1-\alpha)\frac{\gamma}{2}\alpha(1-\alpha) + O(\gamma^{3/2}).$$

Theoretical analysis of ACI's length: exchangeable case

Define $L(\alpha_t) = 2Q(1-\alpha_t)$ the length of the interval predicted by the adaptive algorithm at time t, and $L_0 = 2Q(1-\alpha)$ the length of the interval predicted by the non-adaptive algorithm ($\gamma = 0$).

Theorem

Assume the scores are exchangeable with quantile function Q perfectly estimated at each time, and other assumptions.

Then, for all $\gamma > 0$, $(\alpha_t)_{t>0}$ forms a Markov Chain, that admits a stationary distribution π_γ, and

$$\frac{1}{T}\sum_{t=1}^{T} L(\alpha_t) \xrightarrow[T\to+\infty]{a.s.} \mathbb{E}_{\pi_\gamma}[L] \overset{not.}{=} \mathbb{E}_{\tilde{\alpha}\sim\pi_\gamma}[L(\tilde{\alpha})].$$

Moreover, as $\gamma \to 0$,

$$\mathbb{E}_{\pi_\gamma}[L] = L_0 + Q''(1-\alpha)\frac{\gamma}{2}\alpha(1-\alpha) + O(\gamma^{3/2}).$$

Theoretical analysis of ACI's length: AR(1) case

Theorem

Assume the residuals follow an AR(1) process: $\hat{\varepsilon}_{t+1} = \varphi \hat{\varepsilon}_t + \xi_{t+1}$ *with* $(\xi_t)_t$ *i.i.d. random variables and other assumptions, we have:*

$$\frac{1}{T} \sum_{t=1}^{T} L(\alpha_t) \underset{T \to +\infty}{\overset{a.s.}{\longrightarrow}} \mathbb{E}_{\pi_{\gamma,\varphi}}[L].$$

Numerical analysis of ACI's length: AR(1) case

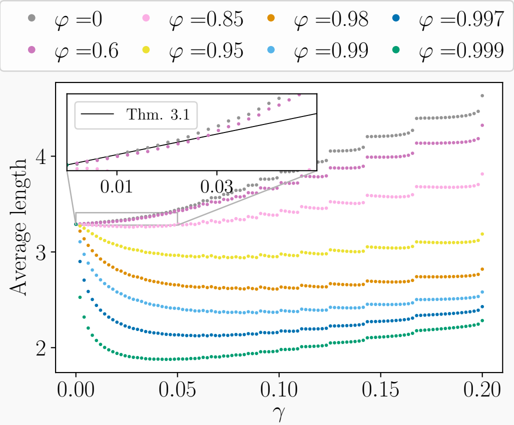


Figure 8: Average length depending on γ for each φ.

Numerical analysis of ACI's length: AR(1) case, cont'd

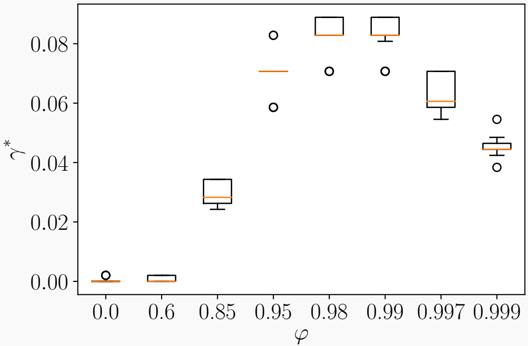


Figure 9: γ^* minimizing the average length for each φ.

Price prediction with confidence in 2019

Electricity Spot prices

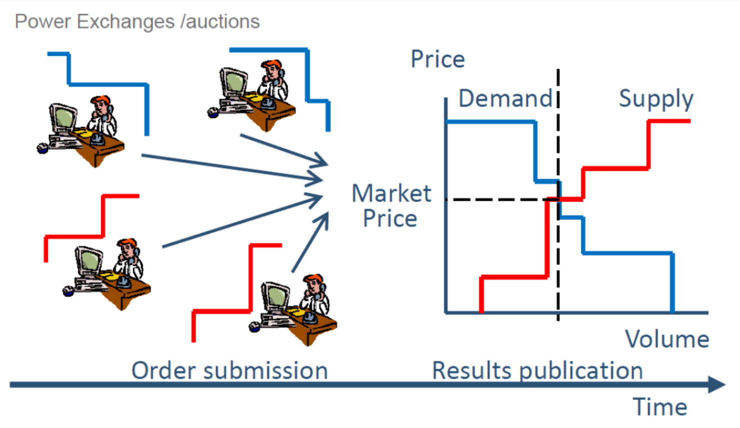


Figure 10: Drawing of spot auctions mechanism

French Electricity Spot prices data set: visualisation

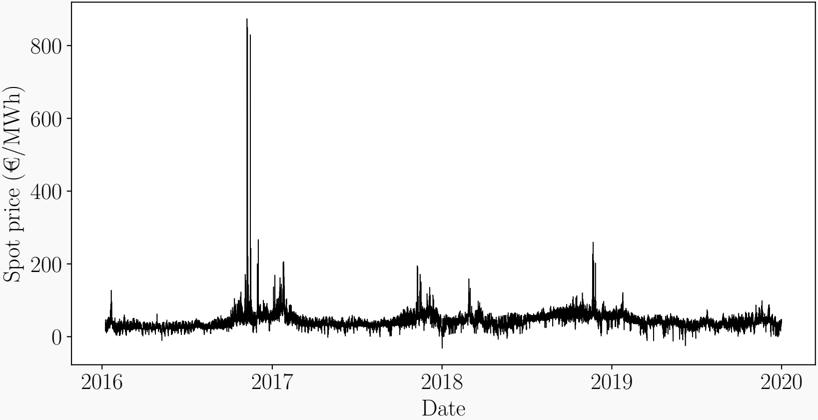


Figure 11: Representation of the French electricity spot price, from 2016 to 2019.

French Electricity Spot prices data set: extract

Date and time	Price	Price D-1	Price D-7	For. cons.	DOW
11/01/16 0PM	21.95	15.58	13.78	58800	Monday
11/01/16 1PM	20.04	19.05	13.44	57600	Monday
⋮	⋮	⋮	⋮	⋮	⋮
12/01/16 0PM	21.51	21.95	25.03	61600	Tuesday
12/01/16 1PM	19.81	20.04	24.42	59800	Tuesday
⋮	⋮	⋮	⋮	⋮	⋮
18/01/16 0PM	38.14	37.86	21.95	70400	Monday
18/01/16 1PM	35.66	34.60	20.04	69500	Monday
⋮	⋮	⋮	⋮	⋮	⋮

Table 1: Extract of the built data set, for French electricity spot price forecasting.

French Electricity Spot prices data set: extract

Date and time	Price	Price D-1	Price D-7	For. cons.	DOW
11/01/16 0PM	21.95	15.58	13.78	58800	Monday
11/01/16 1PM	20.04	19.05	13.44	57600	Monday
⋮	⋮	⋮	⋮	⋮	⋮
12/01/16 0PM	21.51	21.95	25.03	61600	Tuesday
12/01/16 1PM	19.81	20.04	24.42	59800	Tuesday
⋮	⋮	⋮	⋮	⋮	⋮
18/01/16 0PM	38.14	37.86	21.95	70400	Monday
18/01/16 1PM	35.66	34.60	20.04	69500	Monday
⋮	⋮	⋮	⋮	⋮	⋮

Table 1: Extract of the built data set, for French electricity spot price forecasting.

- $y_t \in \mathbb{R}$

French Electricity Spot prices data set: extract

Date and time	Price	Price D-1	Price D-7	For. cons.	DOW
11/01/16 0PM	21.95	15.58	13.78	58800	Monday
11/01/16 1PM	20.04	19.05	13.44	57600	Monday
⋮	⋮	⋮	⋮	⋮	⋮
12/01/16 0PM	21.51	21.95	25.03	61600	Tuesday
12/01/16 1PM	19.81	20.04	24.42	59800	Tuesday
⋮	⋮	⋮	⋮	⋮	⋮
18/01/16 0PM	38.14	37.86	21.95	70400	Monday
18/01/16 1PM	35.66	34.60	20.04	69500	Monday
⋮	⋮	⋮	⋮	⋮	⋮

Table 1: Extract of the built data set, for French electricity spot price forecasting.

- $y_t \in \mathbb{R}$
- $x_t \in \mathbb{R}^d$

Settings

- Forecast for the year 2019.
- Random forest regressor.
- One model per hour, we concatenate the predictions afterwards.

Settings

- Forecast for the year 2019.
- Random forest regressor.
- One model per hour, we concatenate the predictions afterwards.

$\hookrightarrow$ 24 models

Settings

- Forecast for the year 2019.
- Random forest regressor.
- One model per hour, we concatenate the predictions afterwards.

$\hookrightarrow$ 24 models

  - $y_t \in \mathbb{R}$

Settings

- Forecast for the year 2019.
- Random forest regressor.
- One model per hour, we concatenate the predictions afterwards.

$\hookrightarrow$ 24 models

  - $y_t \in \mathbb{R}$
  - $x_t \in \mathbb{R}^d$, with $d = 24 + 24 + 1 + 7 = 56$

24 prices of the day before

Settings

- Forecast for the year 2019.
- Random forest regressor.
- One model per hour, we concatenate the predictions afterwards.

$\hookrightarrow$ 24 models

  - $y_t \in \mathbb{R}$
  - $x_t \in \mathbb{R}^d$, with $d = 24 + 24 + 1 + 7 = 56$

24 prices of the day before

24 prices of the 7 days before

Settings

- Forecast for the year 2019.
- Random forest regressor.
- One model per hour, we concatenate the predictions afterwards.

↪ 24 models

  - $y_t \in \mathbb{R}$
  - $x_t \in \mathbb{R}^d$, with $d = 24 + 24 + 1 + 7 = 56$

24 prices of the day before

24 prices of the 7 days before

Forecasted consumption

Settings

- Forecast for the year 2019.
- Random forest regressor.
- One model per hour, we concatenate the predictions afterwards.

$\hookrightarrow$ 24 models

  - $y_t \in \mathbb{R}$
  - $x_t \in \mathbb{R}^d$, with $d = 24 + 24 + 1 + 7 = 56$

24 prices of the day before

24 prices of the 7 days before

Forecasted consumption

Encoded day of the week

Settings

- Forecast for the year 2019.
- Random forest regressor.
- One model per hour, we concatenate the predictions afterwards.

$\hookrightarrow$ 24 models

  - $y_t \in \mathbb{R}$
  - $x_t \in \mathbb{R}^d$, with $d = 24 + 24 + 1 + 7 = 56$
  - 3 years for training/calibration, i.e. $T_0 = 1096$ observations

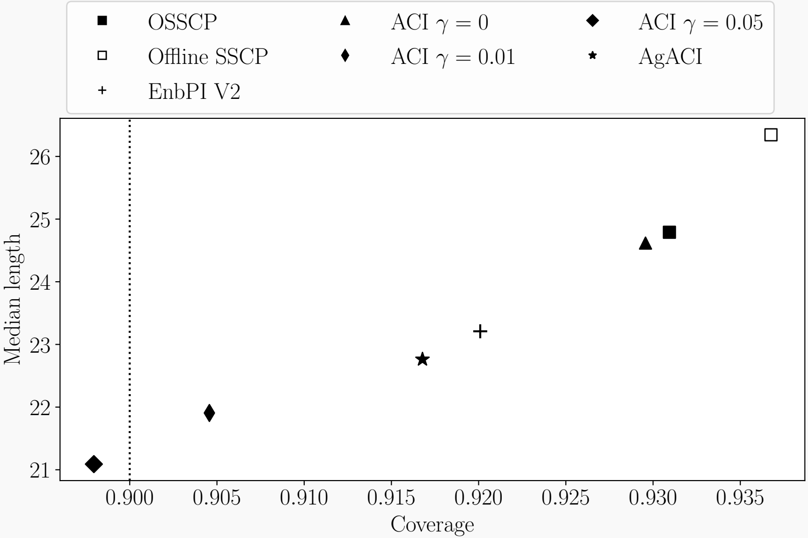
Settings

- Forecast for the year 2019.
- Random forest regressor.
- One model per hour, we concatenate the predictions afterwards.

$\hookrightarrow$ 24 models

  - $y_t \in \mathbb{R}$
  - $x_t \in \mathbb{R}^d$, with $d = 24 + 24 + 1 + 7 = 56$
  - 3 years for training/calibration, i.e. $T_0 = 1096$ observations
  - 1 year to forecast, i.e. $T_1 = 365$ observations

Performance on predicted French electricity Spot price for the year 2019



Performance on predicted French electricity Spot price: visualisation of a day

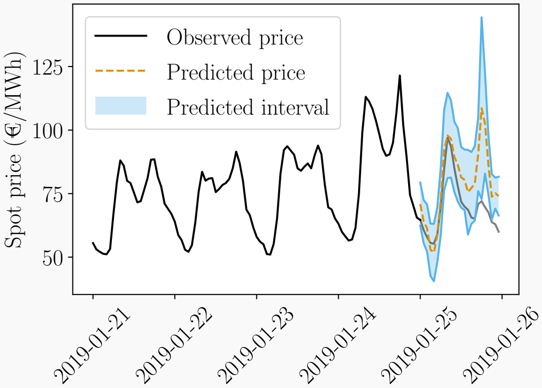


Figure 12: French electricity spot price, its prediction and its uncertainty with AgACI.

Available methods for non-exchangeable data, in the context of time series

How to adapt to time series?

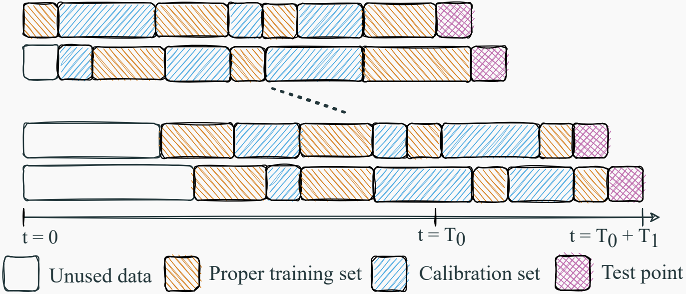
Usual ideas from the time series literature:

- Consider an online procedure (for each new data, re-train and re-calibrate)

How to adapt to time series?

Usual ideas from the time series literature:

- Consider an online procedure (for each new data, re-train and re-calibrate)



How to adapt to time series?

Usual ideas from the time series literature:

- Consider an online procedure (for each new data, re-train and re-calibrate)
  - $\hookrightarrow$ update to recent observations (trend impact, period of the seasonality, dependence...)

How to adapt to time series?

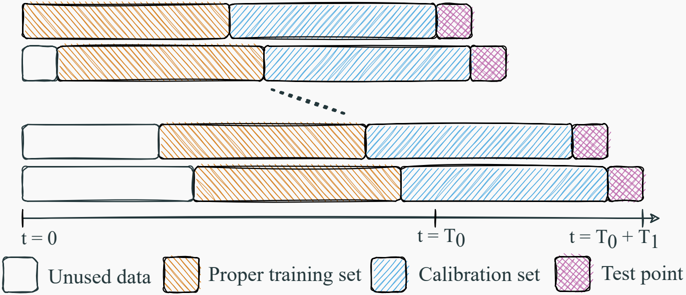
Usual ideas from the time series literature:

- Consider an online procedure (for each new data, re-train and re-calibrate)
  - $\hookrightarrow$ update to recent observations (trend impact, period of the seasonality, dependence...)
- Use a sequential split

How to adapt to time series?

Usual ideas from the time series literature:

- Consider an online procedure (for each new data, re-train and re-calibrate)
  - $\hookrightarrow$ update to recent observations (trend impact, period of the seasonality, dependence...)
- Use a sequential split



How to adapt to time series?

Usual ideas from the time series literature:

- Consider an online procedure (for each new data, re-train and re-calibrate)
  - $\hookrightarrow$ update to recent observations (trend impact, period of the seasonality, dependence...)
- Use a sequential split
  - $\hookrightarrow$ use only the past so as to correctly estimate the variance of the residuals (using the future leads to optimistic residuals and underestimation of their variance)

Online sequential split conformal prediction (OSSCP)

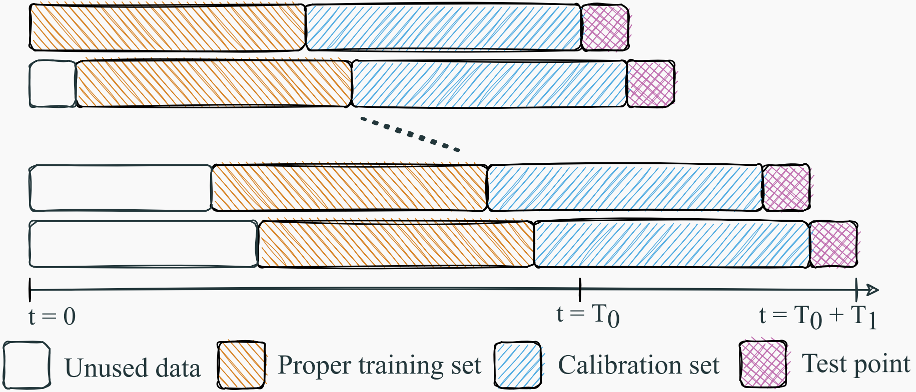


Figure 14: Diagram describing the online sequential split conformal prediction.

Online sequential split conformal prediction (OSSCP)

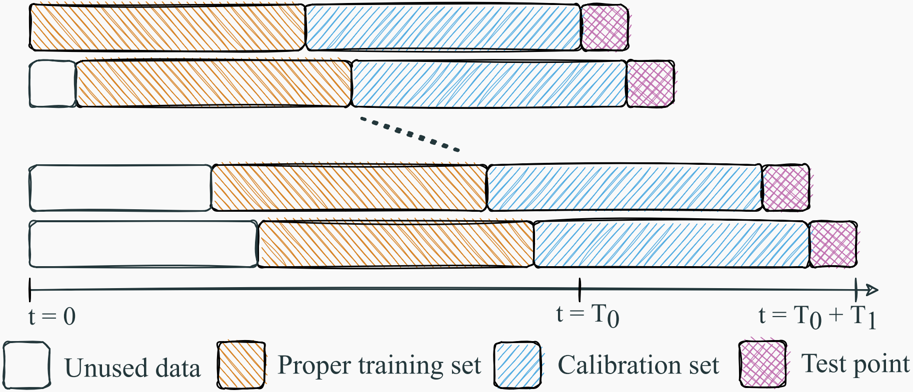


Figure 14: Diagram describing the online sequential split conformal prediction.

Wisniewski et al. (2020); Kath and Ziel (2021); and our study

Online sequential split conformal prediction (OSSCP)

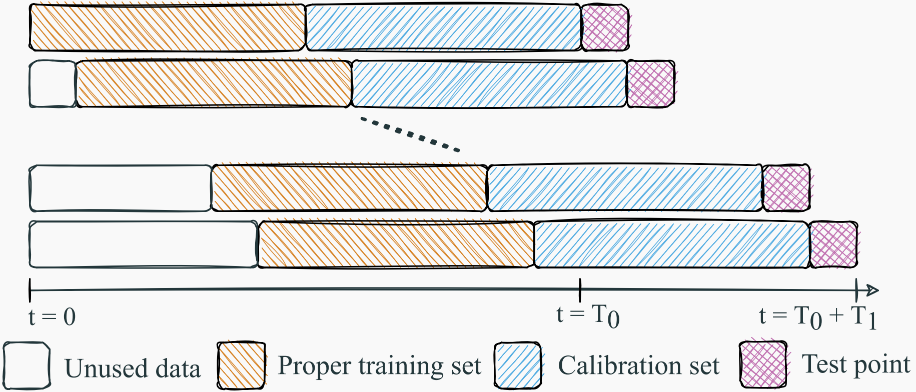


Figure 14: Diagram describing the online sequential split conformal prediction.

Wisniewski et al. (2020); Kath and Ziel (2021); and our study

↪ tested on real time series

EnbPI, Xu and Xie (2021)

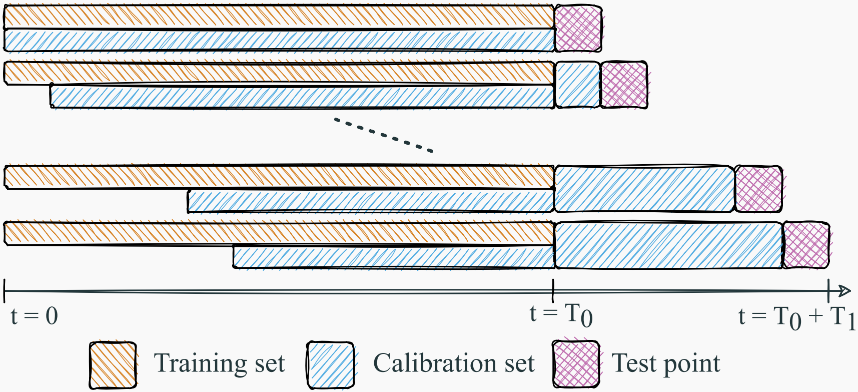


Figure 15: Diagram describing the EnbPI algorithm.

EnbPI, Xu and Xie (2021)

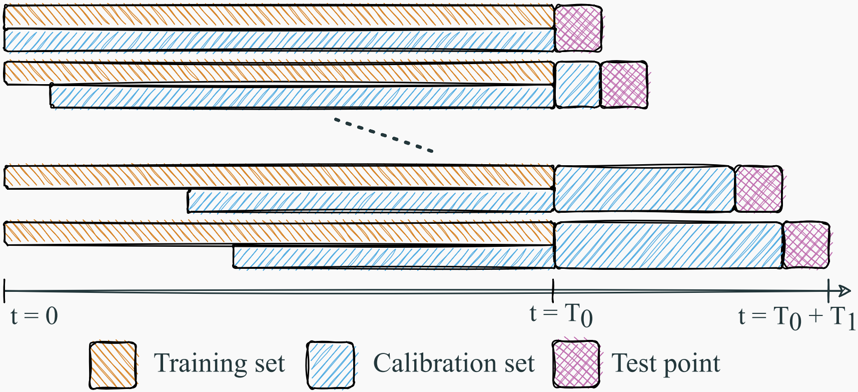


Figure 15: Diagram describing the EnbPI algorithm.

↪ tested on other real time series

↪ compared to offline methods