Adaptive Conformal Predictions for Time Series

Margaux Zaffran[1,2,3] Olivier Féron[1] Yannig Goude[1] Julie Josse[2] Aymeric Dieuleveut[3]

[1]Electricité De France, Paris, France [2]INRIA, Montpellier, France [3]Ecole Polytechnique, Paris, France

Motivations and setting

Control validity: produce **predictive intervals**, enjoying **theoretical guarantees** on their **coverage** with **few assumptions**.
Optimize efficiency: intervals **as small as possible**.
Setting in time series

- **Data:** T_0 observations $(x_1, y_1), \ldots, (x_{T_0}, y_{T_0})$ in $\mathbb{R}^d \times \mathbb{R}$.
- **Aim:** predict for T_1 subsequent observations $x_{T_0+1}, \ldots, x_{T_0+T_1}$.

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

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

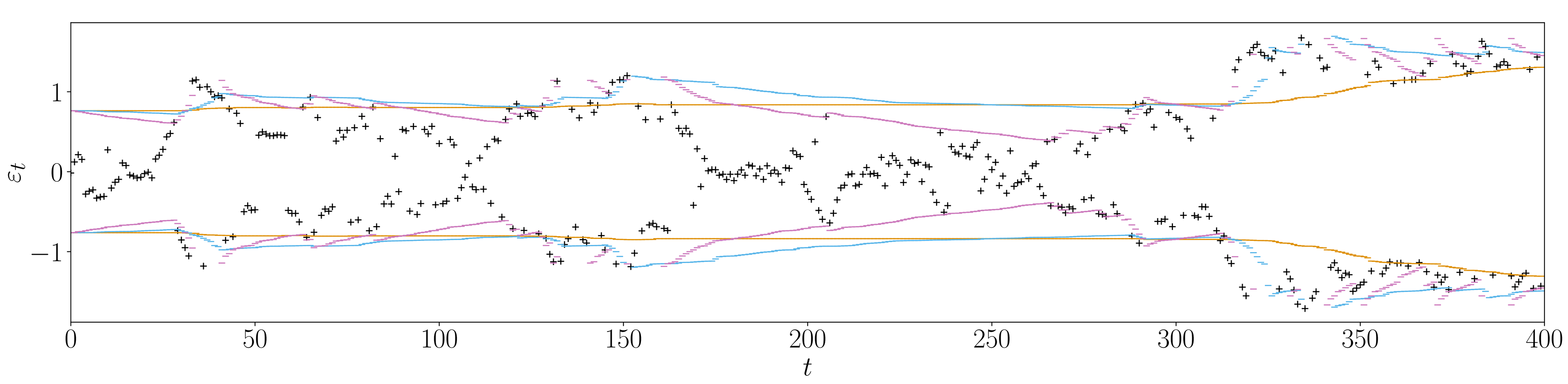
Summary

Conformal prediction gives predictive sets under exchangeability, not time series. **ACI can be used** but require a learning rate γ.
❶ **Theory** on ACI's efficiency depending on the learning rate γ.
❷ **Algorithm** based on expert aggregation, to avoid choosing γ.
❸ **Numerical tests**: synthetic and French electricity prices.

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

Use an **effective quantile level** based on a recursive equation and a **learning rate** γ: $\alpha_{t+1} := \alpha_t + \underbrace{\gamma}_{\geq 0}\left(\alpha - \mathbb{1}\left\{y_t \notin \hat{C}_{\alpha_t}(x_t)\right\}\right)$.

Illustration: ACI with $\gamma = 0$, $\gamma = 0.01$ and $\gamma = 0.05$.



Theory

For any distribution:

$$\underbrace{\left|\frac{1}{T_1}\sum_{t=T_0+1}^{T_0+T_1} \mathbb{1}\left\{y_t \in \hat{C}_{\alpha_t}(x_t)\right\} - (1-\alpha)\right|}_{\text{Average coverage}} \leq \frac{2}{\gamma T_1}$$

❶ Impact of the learning rate γ

Exchangeable case

Theorem 1 (informal)

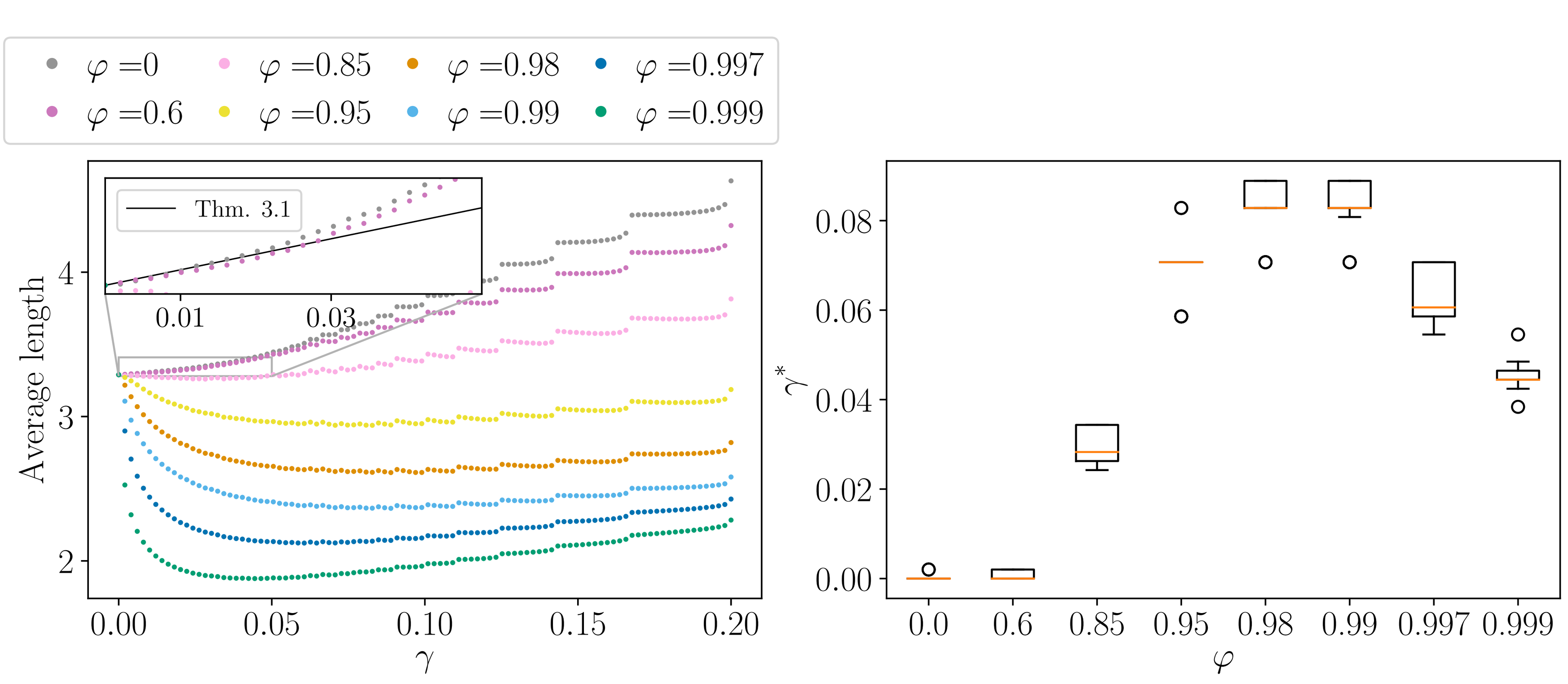
Assume exchangeable scores and perfect calibration. As $\gamma \to 0$:

Average length of intervals from ACI using γ
$=$ Average length of intervals from Split Conformal Prediction
$+\ \gamma \times \underbrace{\mathcal{C}(\alpha, \text{distribution of the data})}_{>0 \text{ in general}}$.

Auto-regressive case: $\varepsilon_{t+1} = \varphi\varepsilon_t + \xi_{t+1}$.

Theorem 2 (informal)

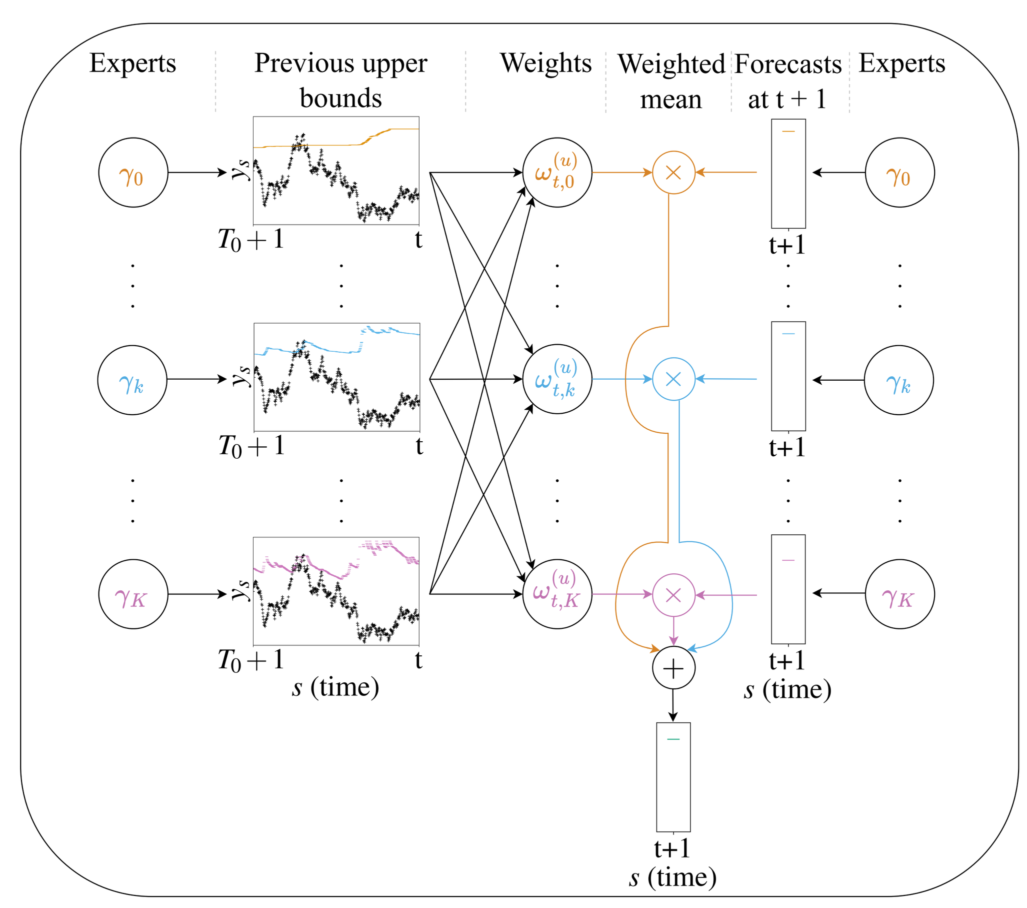
Assume auto-regressive residuals and perfect calibration. There exists an optimal $\gamma^* > 0$ minimizing the average length for high φ.



Conclusion: choosing γ is crucial but difficult.

❷ AgACI

- Experts (ACI with many γ) aggregation (Cesa-Bianchi and Lugosi, 2006).
- One algorithm for the upper bound, another for the lower bound.
- Based on the pinball loss of level $1-\frac{\alpha}{2}$, or of level $\frac{\alpha}{2}$.

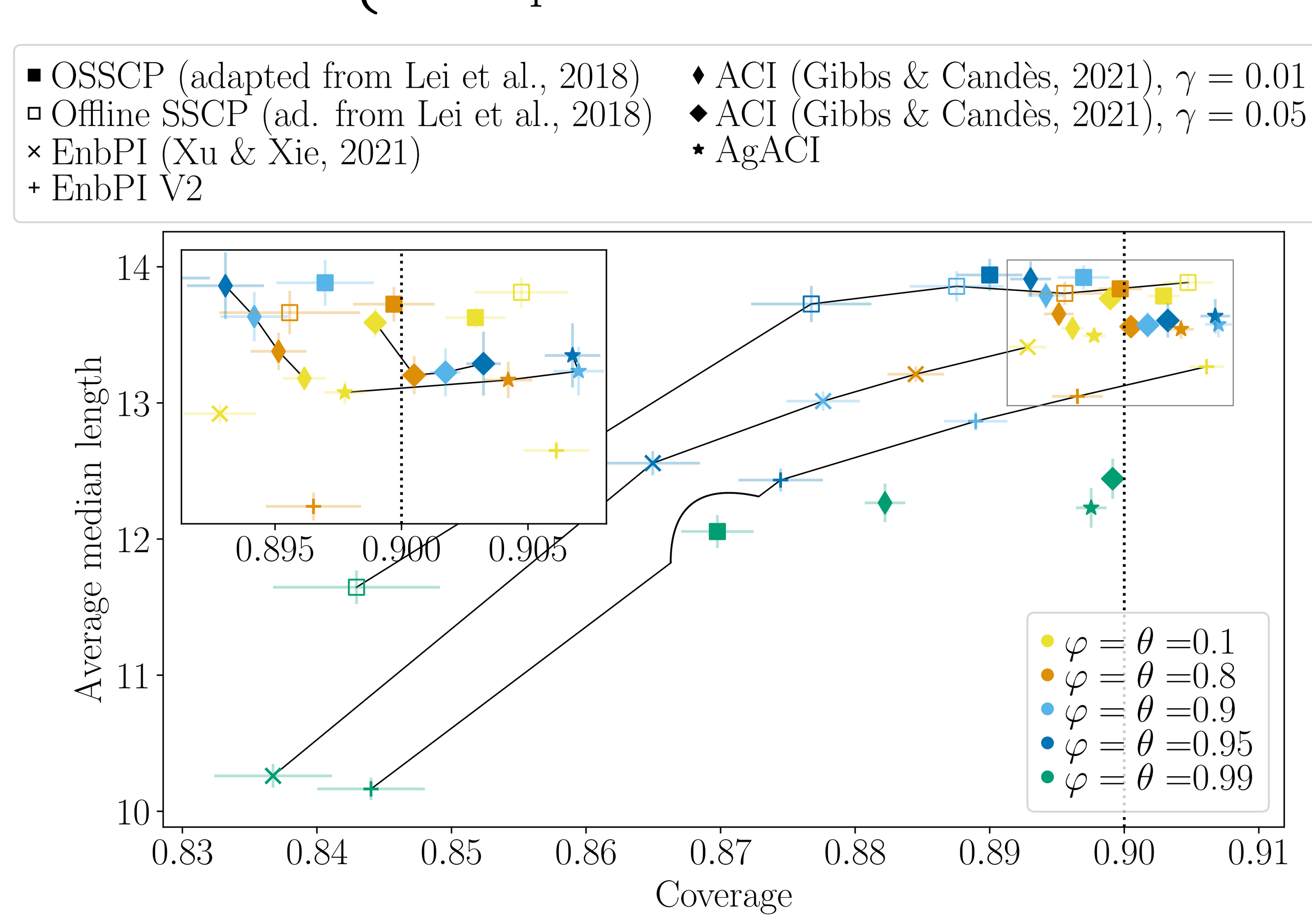


❸ Numerical results

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

with $X_{t,\cdot} \sim \mathcal{U}([0,1])$ and $\varepsilon_{t+1} = \varphi\varepsilon_t + \xi_{t+1} + \theta\xi_t$, with ξ_t a white noise of variance σ^2, such that $\mathrm{Var}(\varepsilon_t)$ is 10.

- Random forest are used as regressor.
- For each setting: $\begin{cases} \text{300 points, last 100 kept for prediction} \\ \text{500 repetitions} \end{cases}$.



Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games.*
Gibbs, I. and Candès, E. (2021). Adaptive Conformal Inference Under Distribution Shift. *NeurIPS.*