A Solving the Recursion

A.1 Independent and Identically Distributed Errors

We start with the case where errors follow an independent and identically distributed (IID) von Mises distribution. For this scenario, we can customise Equation (7) in the following recursive form

$$F_{\text{IID}}(s,\mu) = \max\left(F_{\text{IID}}(s-1,\mu), F_{\text{IID}}(s-1) + \frac{\beta}{\kappa}\right) + \cos(h_s - \mu). \tag{A.1}$$

Given that the sum of scaled and shifted cosine functions can be expressed as a single scaled cosine function with an adjusted phase, $F_{\text{IID}}(s,\mu)$ can be expressed as a piecewise cosine wave. Each segment of this wave can be written in the form $R\cos(\mu - \alpha) + z$. We can illustrate this behaviour by starting at time $s = 1$. At this point, the function is

$$F_{\text{IID}}(1,\mu) = \cos(h_1 - \mu) \tag{A.2}$$

This represents a single cosine wave over the range $I_0^1 = [-\pi, \pi)$. The maximum value of this function occurs when $F_{\text{IID}}(1) = 1$. Now, moving to time $s = 2$ gives us:

$$\begin{aligned} F_{\text{IID}}(2,\mu) &= \max\left\{\cos(h_1 - \mu), 1 + \frac{\beta}{\kappa}\right\} + \cos(h_2 - \mu) \\ &= \begin{cases} \cos(h_2 - \mu) + 1 + \frac{\beta}{\kappa} & \text{if } \cos(h_1 - \mu) \leq 1 + \frac{\beta}{\kappa} \\ \cos(h_1 - \mu) + \cos(h_2 - \mu) & \text{otherwise.} \end{cases} \end{aligned} \tag{A.3}$$

This results in two cases:

1. If $\cos(h_1 - \mu) \leq 1 + \frac{\beta}{\kappa}$, the function becomes a shifted cosine wave by adding $1 + \frac{\beta}{\kappa}$ to the second cosine term.

2. Otherwise, it is the sum of two waves: $\cos(h_1 - \mu) + \cos(h_2 - \mu) = R_{1,2}\cos(\alpha_{1,2} - \mu)$,

where

$$R_{u,s} = \sqrt{\left(\sum_{i=u}^{s} \sin(h_i)\right)^2 + \left(\sum_{i=u}^{s} \cos(h_i)\right)^2} \tag{A.4}$$

$$\alpha_{u,s} = \tan_2^{-1}\left(\sum_{i=u}^{s} \sin(h_i), \sum_{i=u}^{s} \cos(h_i)\right) \tag{A.5}$$

using standard trignometric identities. At this point, $F(2)$, the maximum value at time $s = 2$, is obtained by comparing the maximum of each segment. The key here is that the function evolves as a series of cosine waves over distinct intervals.

As we proceed through the iterations, we must track the coefficients (z, R, and α) along with the intervals

$$I_{s-1,s} = \left\{\mu \mid F(s,\mu) \leq F(s) + \frac{\beta}{\kappa}\right\} \tag{A.6}$$

$$I_{u,s} = I_{u,s-1} \cap (I_{s-1,s})^c \quad \text{for } u < s-1 \tag{A.7}$$

Therefore, the last most probable turning point location is estimated by looking at τ where $\underset{\mu}{\operatorname{argmax}} F_{\text{IID}}(s,\mu) \in I_{\tau,s}$.

By pruning empty intervals and tracking the most recent turning points, we can efficiently determine the best segmentation. To do this, we start at the final time point n and identify the last turn, τ. From τ, we trace back through the sequence of *last* turning points, iterating until we reach the beginning of the sequence at $\tau = 0$. To illustrate, Figure (A.1) shows a simple example of how $F(s,\mu)$ evolves over two iterations using simulated data. This algorithm is an instance of the class of functional pruning algorithms proposed by Maidstone et al. (2017). Pseudo-code for the algorithm is given in Algorithm (1).

Algorithm 1 AFPOP (Angular Functional Pruning Optimal Partitioning (IID))

Initialise the function $F(1, \mu) = \cos(h_1 - \mu)$.

Set $I_0^1 = (-\pi, \pi]$, $T = \{0\}$ and $\alpha = \beta/\kappa$.

for $t = 2$ to n **do**

1. Calculate $F(t, \mu) = \max\{F(t-1, \mu), F(t-1) + \alpha\} + \cos(h_t - \mu)$
2. Calculate $I_{t-1,t} = \{\mu | F(t, \mu) = F(t) + \alpha + \cos(h_t - \mu)\}$
3. Calculate $I_{i,t} = I_{i,t-1} \cap I_{t-1,t}^c$ for each $i < t - 1$
4. Append the values i to T where $\underset{\mu}{\operatorname{argmax}} F_{\text{IID}}(t, \mu) \in I_{i,t}$.

end for

Initialise TPs $= \{n\}$ and $\tau = n$

while $\tau > 0$ **do**

1. set $\tau = T_\tau$.
2. Update TPs = TPs $\cup$ $\{\tau\}$.

end while

The turning points are stored in TPs.

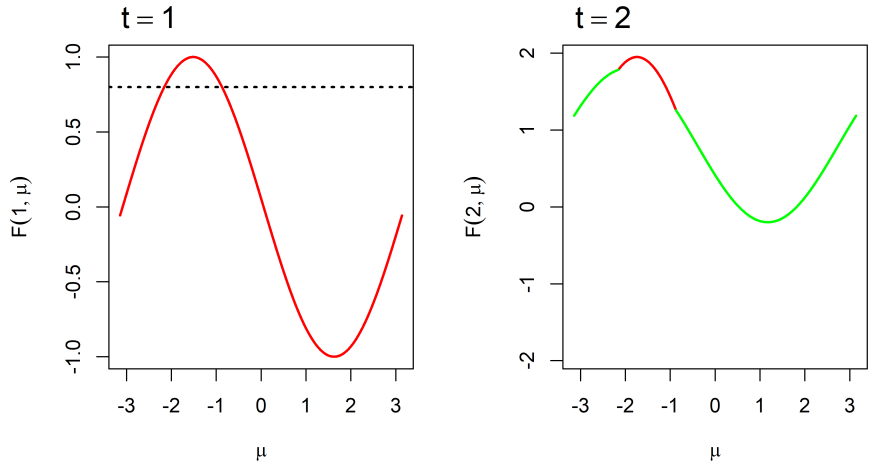

Fig. A.1: A simple example showing the function $F(t, \mu)$ for the first two iterations. The function $F(t, \mu)$ is plotted as a bold line. The colors indicate the time of the most recent turning points (red for $t = 1$ and black for $t = 2$). The dotted horizontal line indicates $F(1) + \frac{\beta}{\kappa}$.

A.2 Estimation of Turning Points with Autocorrelated Errors

In the case of the AR(1) model where $\rho \neq 0$, the function $F(s, \mu)$ can no longer be expressed in a closed form due to the autocorrelation component. This presents a challenge for directly applying the previously discussed methods. However, given that the space of μ is bounded, we can still approximately solve the optimization problem by discretising the space of μ into a finite set with a reasonable resolution.

Discretising the space of μ involves dividing the continuous range $(-\pi, \pi]$ into a finite set of discrete values, making the problem computationally tractable. Once we have discretised the space of possible μ, we perform a forward iteration through the data to calculate

$$F_{\mathrm{AR}}(s, \mu) = \max \left\{ F_{\mathrm{AR}}(s-1, \mu), F_{\mathrm{AR}}(s-1) + \frac{\beta}{\kappa} \right\} + \cos(h_s - E(h_s \mid \mu, h_{s-1}, \rho)) \quad \text{(A.8)}$$

for each value of μ, while keeping track of a vector indicating the most recent turning

points and a vector showing the origin of each point in the space of μ.

B Further Simulations

B.1 Robustness to model mis-specification (set-up)

In addition to testing our inference procedure on data generated by the underlying model, we also tested its robustness to model mis-specification. First, we addressed mis-specification of the error distribution by considering errors from a wrapped normal distribution ($\epsilon_t \sim WN(\mu_t, \gamma)$), where γ is the concentration parameter of the distribution. The values ($\gamma \in \{0.893, 0.966, 0.989, 0.995, 0.997, 0.999\}$) were chosen to closely mimic the concentration parameters used previously when it is assumed that the errors have a von Mises distribution with ($\kappa \in \{5, 15, 45, 45, 100, 200, 400\}$). Again, $\rho \in \{0.03, 0.08, ..., 0.98\}$

Next, we explored mis-specifications in the underlying process. One case focused on an angular AR(3) process, as outlined in Equation (A.9), with the coefficients $\phi_{1,2,3} = [0.5, 0.3, 0.2]$. We also examined the scenario where the true process is a moving average process, an angular MA(10), as described in Equation (A.10).

For the angular AR(3) model, the expected value of h_t conditioned on past values is given by

$$E_1(h_t|\mu_t, h_{(t-1):(t-4)}, \rho) = \tan_2^{-1}\left((1-\rho)\sin(\mu_t) + \rho\sum_{i=t-4}^{t-1}\phi_{t-i}\sin(h_i), \; (1-\rho)\cos(\mu_t) + \rho\sum_{i=t-4}^{t-1}\phi_{t-i}\cos(h_i)\right). \tag{A.9}$$

Similarly, for the angular MA(10) model, we define the expected value of h_t as

$$E_2(h_t|\mu_t, h_{(t-1):(t-11)}, \rho) = \tan_2^{-1}\left((1-\rho)\sin(\mu_t) + \frac{\rho}{10}\sum_{i=t-11}^{t-1}\sin(h_i), \; (1-\rho)\cos(\mu_t) + \frac{\rho}{10}\sum_{i=t-11}^{t-1}\cos(h_i)\right). \tag{A.10}$$

These formulations allow us to capture the effects of misspecification in both autoregressive and moving average structures, ensuring that our path demarcation procedure remains robust under various conditions.

Lastly, given the assumption of a constant concentration parameter across the entire trajectory, we investigated the performance of our methods when this condition is not met. Specifically, we generated errors from different concentration parameters, $\kappa \in \{5, 15, 45\}$, where the first 100 observations were generated using $\kappa = 5$, followed by 100 observations from $\kappa = 15$, and then 100 observations from $\kappa = 45$. This pattern was repeated for the entire length of our artificial trajectories.

B.2 Robustness to model mis-specification (Results)

B.2.1 Case 1: Non-constant κ

In the modelling setup, we assume that the concentration parameter is constant throughout the entire time series. However, in practice, this assumption may not always hold. To assess the robustness of our models under this condition, we analyse how well the methods perform when the concentration parameter varies. Figure A.2 illustrates the accuracy of the demarcation methods across different scenarios. In Scenario A, both the AR(1) and IID models demonstrate nearly identical performance, achieving almost perfect F1-scores for values of $\rho \leq 0.60$. This indicates that both models are robust even when the concentration parameter is not constant, provided the autocorrelation is moderate. A similar trend is observed in Scenario B, where both methods maintain high accuracy as the autocorrelation decrease $\rho \leq 0.70$. In Scenario C, our two proposed methods perform exceptionally well, achieving perfect accuracy across all levels of autocorrelation.

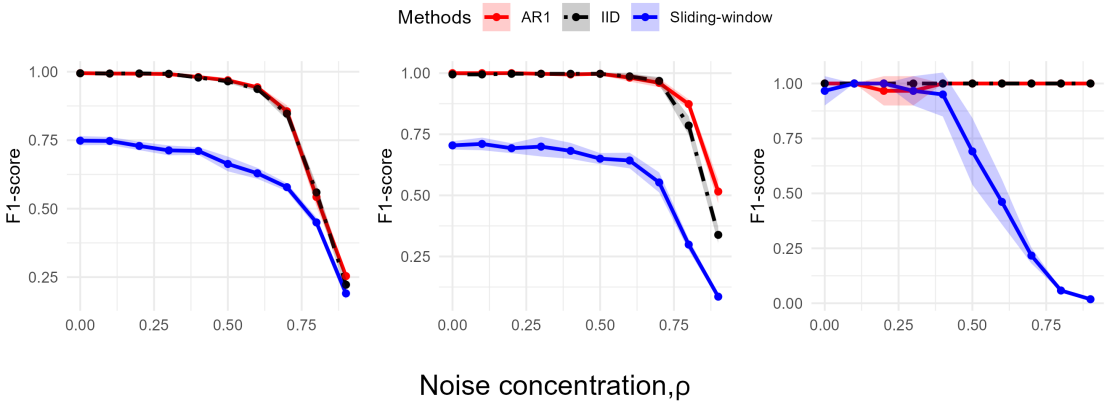

Fig. A.2: Line plots of the F1-score under noise with non-constant κ, as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and sliding-window (blue). The first row shows the results for scenarios A, B, and C (see Section 2.4).

B.2.2 Case 2: Errors generated from wrapped normal

In contrast to the model's assumption of errors generated from a von Mises distribution, Figure A.3 shows the simulation results of errors generated from a wrapped normal distribution with different levels of the variance and the turning points density. The results appear very similar to those in Figure 5, indicating robustness against misspecification of the underlying error-generating distribution.

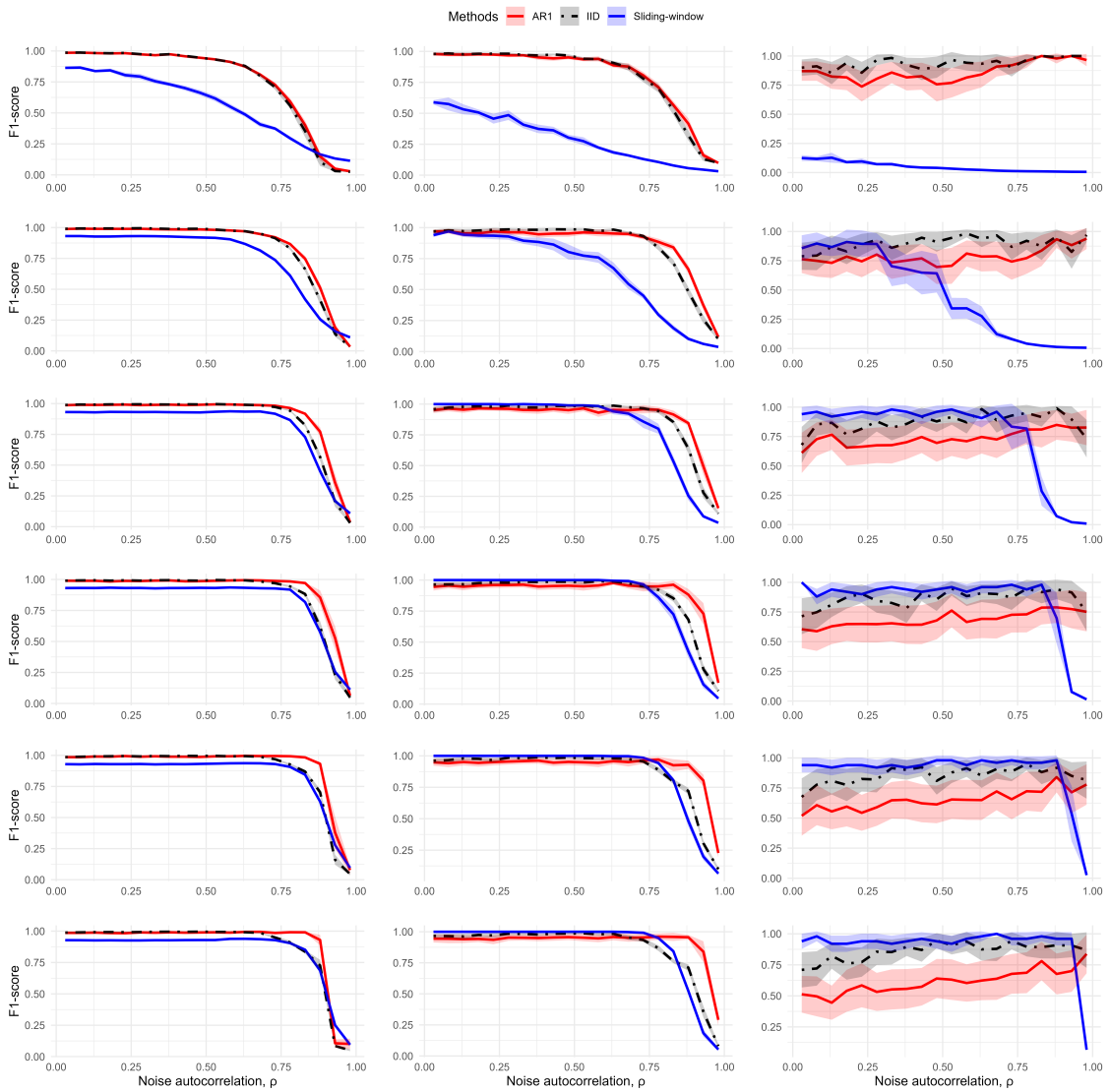

Fig. A.3: Line plots of the F1-score with wrapped normal noise as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and sliding-window (blue). The first row presents the results for scenarios A, B, and C (see Section 2.4) with $\kappa_\rho = 5$, 15, 45, 100, 200, and 400 in each row, respectively.

B.2.3 Case 3: Angular AR(3) and angular MA(10) processes

In this section, we have examined the performance of our turning point identification techniques under scenarios where the underlying processes are misspecified. We first consider the case where the bearings are generated from an angular autoregressive process of order 3 (angular AR(3)), as shown in Figure A.4. Here, the model assumes an angular AR(1) process, but the true data-generating process is angular AR(3). Our methods show very similar and accurate performance for small to moderate levels of autocorrelation. However, as the autocorrelation (ρ) increases, the F1-score begins to decline. This trend is consistent across different scenarios involving the number of turning points and concentration levels, suggesting that higher autocorrelation introduces challenges for turning point detection given this form of model misspecification.

Surprisingly, at very high concentration levels, the performance of our algorithm does not improve as opposed to the trends observed in Figures 5 and A.3 making the sliding-window algorithm comparatively more accurate in these cases. Results from data generated under an angular moving average process of order 10 (MA(10)), shown in Figure A.5, closely mirror those observed in the AR(3) case.

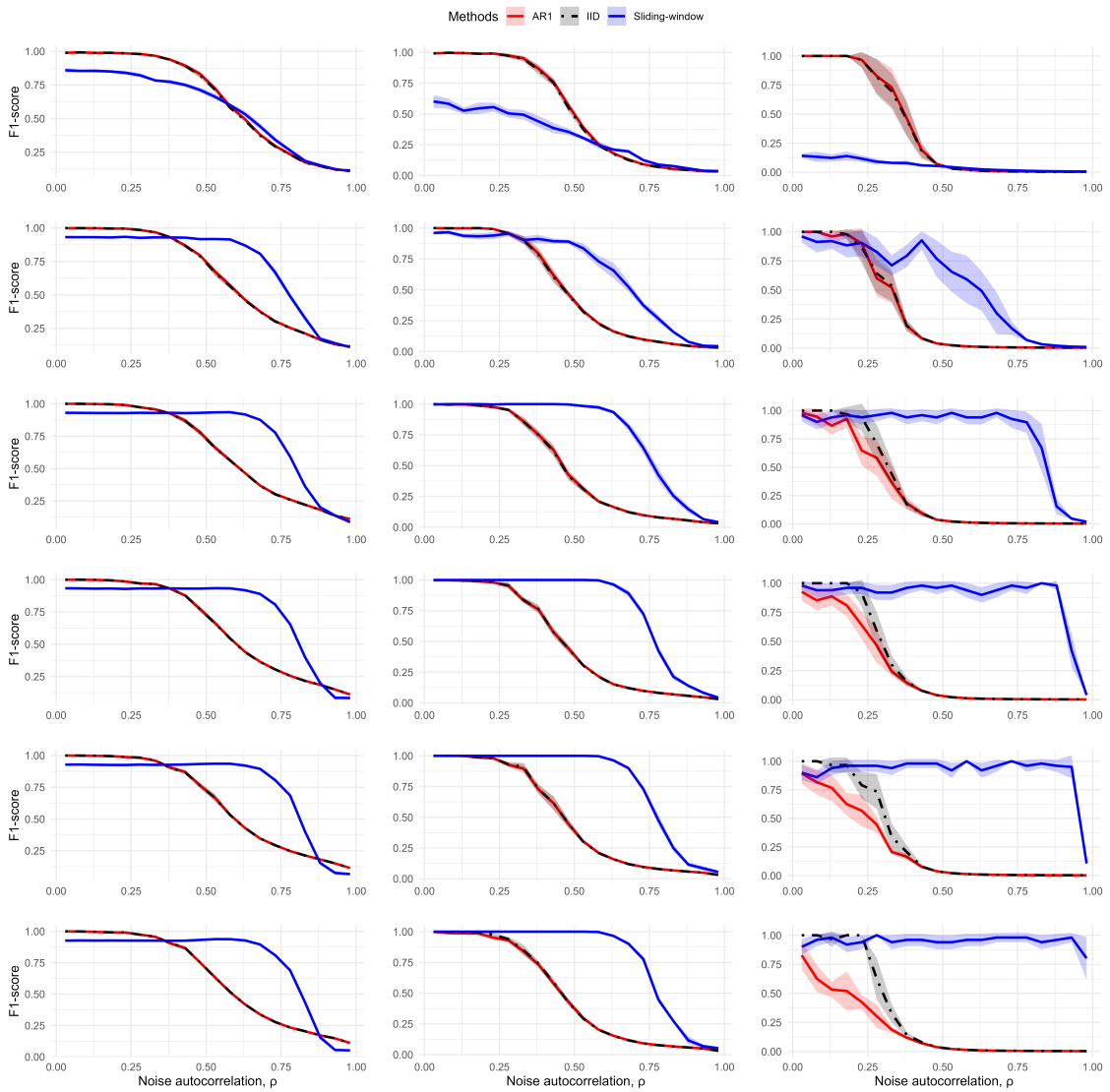

Fig. A.4: Line plots of the F1-score with angular AR(3) noise as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and sliding-window (blue). The first row presents the results for scenarios A, B, and C (see Section 2.4) with $\kappa_\rho = 5$, 15, 45, 100, 200, and 400 in each row, respectively.

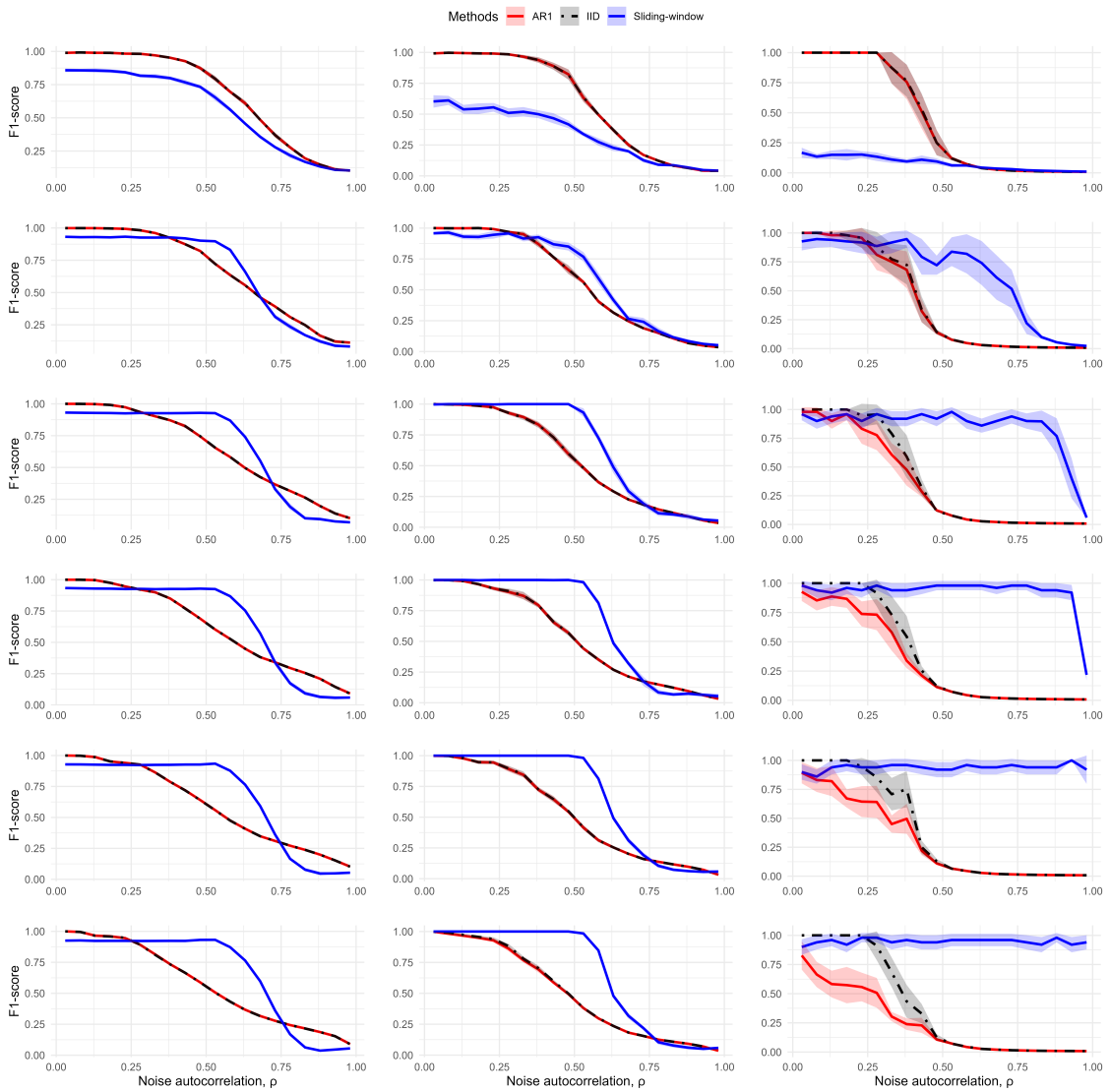

Fig. A.5: Line plots of the F1-score with MA(10) noise as a function of the autocorrelation parameter ρ, comparing the performance of three methods: AR(1) (red), IID (black), and sliding-window (blue). The first row presents the results for scenarios A, B, and C (see Section 2.4) with $\kappa_\rho = 5$, 15, 45, 100, 200, and 400 in each row, respectively.

B.3 Sensitivity to the Acceptance Rule

In our simulation setup, a turning point is considered correctly identified if the inferred location falls within a certain distance of the true turning point. This criterion has been used in prior work, such as Truong et al. (2020). To evaluate how sensitive our conclusions are to the choice of this distance threshold (i.e., the tolerance for matching true and inferred turning points), we conducted a systematic sensitivity analysis.

Specifically, we recomputed the F1-score under three different temporal tolerance thresholds: 5, 10, and 20 time units. Figure A.6,Figure A.7 and Figure A.8 present the results for the three simulated trajectories under an angular AR(1) process. Each panel corresponds to one trajectory, with rows representing different concentration parameters (κ) and columns corresponding to the three tolerance thresholds.

Across all three trajectories, we observe consistent patterns. The AR(1)-based segmentation method consistently outperforms the IID-based and sliding-window approaches across all tolerance levels and noise autocorrelation values. Relaxing the tolerance (moving from left to right across columns) results in slight improvements in F1-scores for all methods, with the most notable gains observed in the sliding-window approach. However, the relative ranking of methods remains largely unchanged.

Importantly, the trends remain qualitatively consistent across all three trajectories and all of our other simulation setups. We find that a moderate tolerance (e.g., 10 time units) provides a reasonable trade-off between strictness and robustness, and we use this value for our main evaluations.

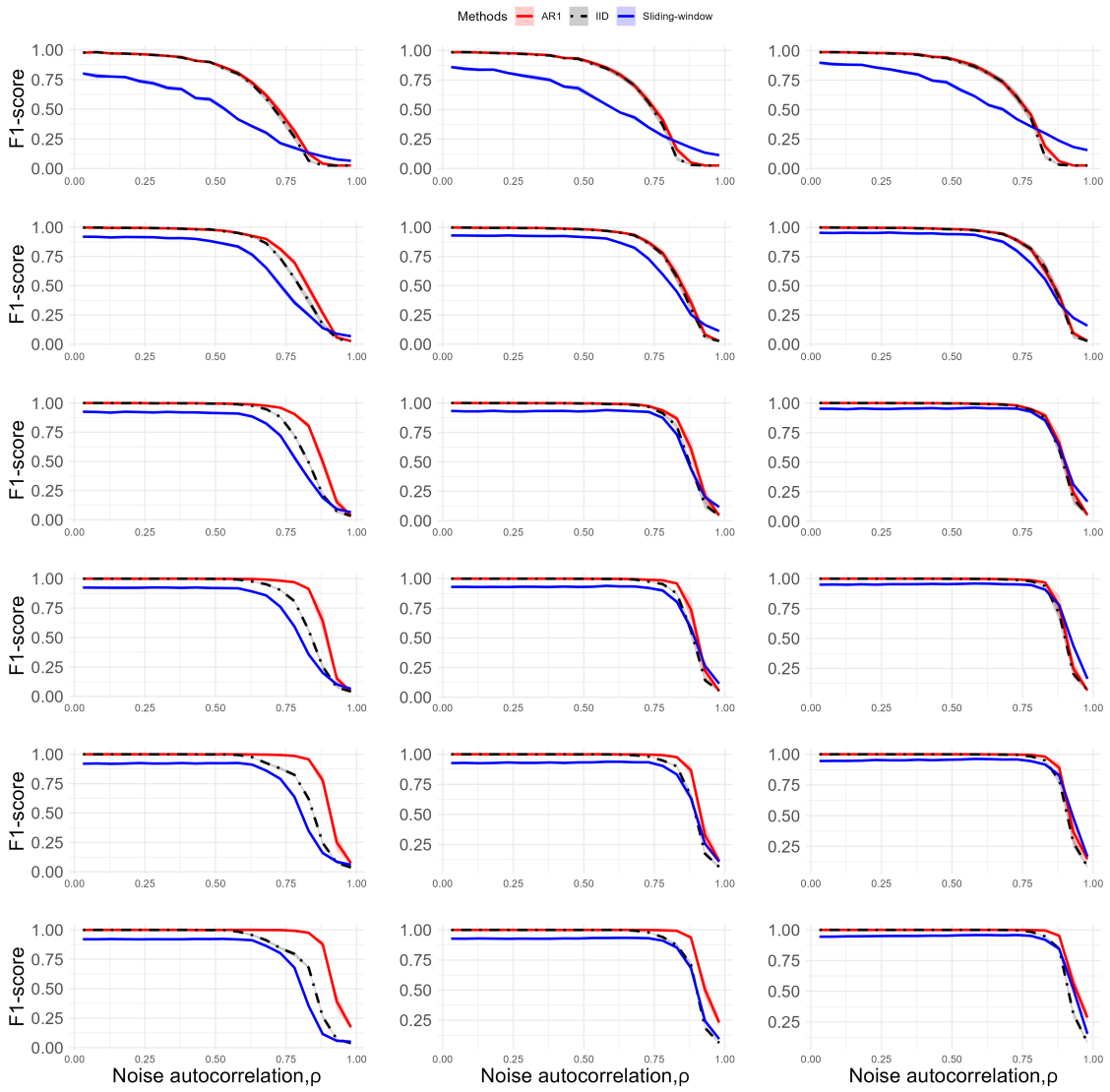

Fig. A.6: Sensitivity analysis of F1-score under varying acceptance rules for scenario A. Each panel shows results from one simulated trajectory under an angular AR(1) process. Rows represent different concentration parameters (κ), and columns represent different temporal tolerance thresholds (from left to right: 5, 10, and 20 time-units). F1-score is plotted against noise autocorrelation ρ for three methods: AR(1)-based segmentation (red), IID-based segmentation (black dashed), and the sliding-window approach (blue).

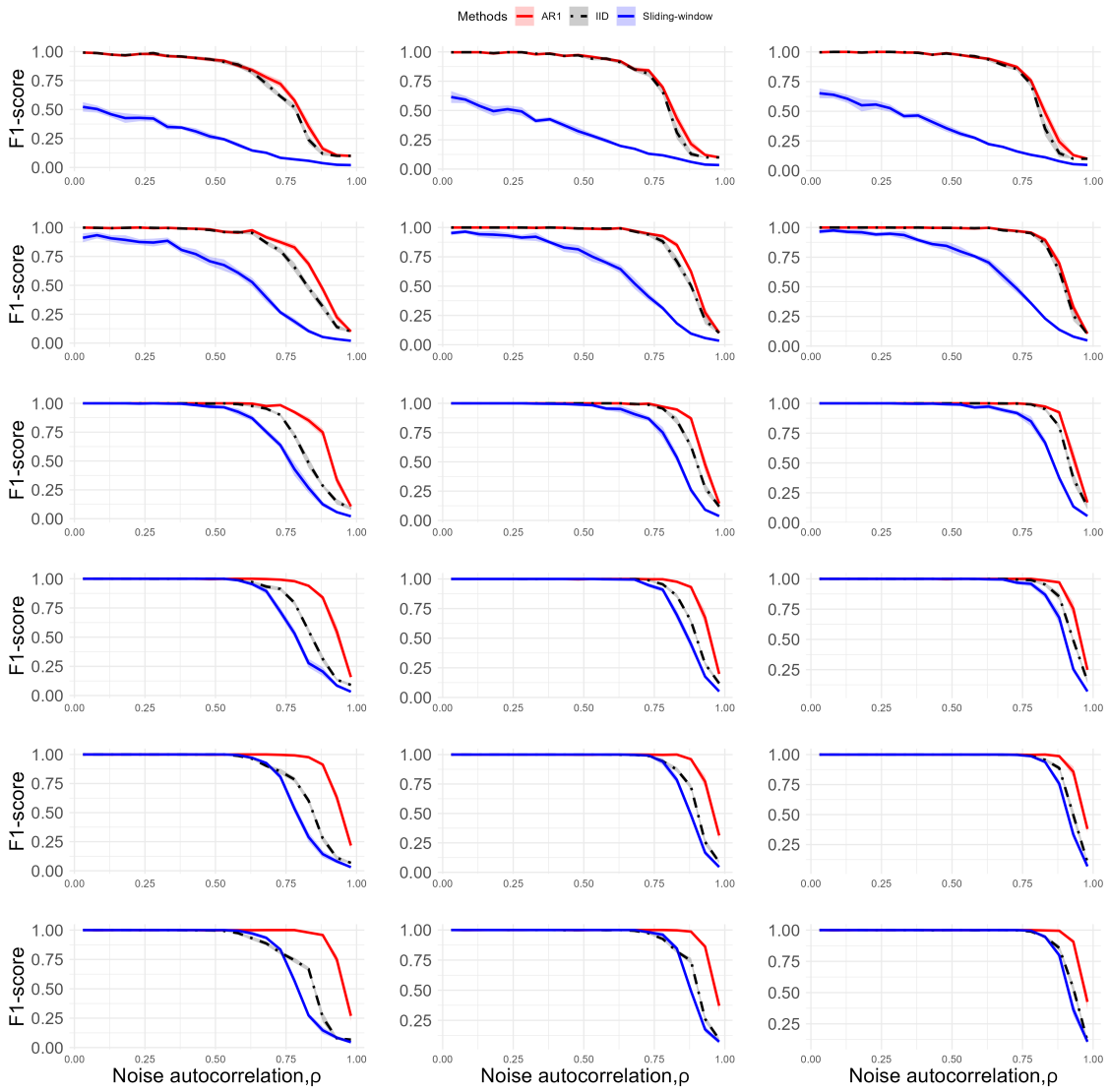

Fig. A.7: Sensitivity analysis of F1-score under varying acceptance rules for scenario B. Each panel shows results from one simulated trajectory under an angular AR(1) process. Rows represent different concentration parameters (κ), and columns represent different temporal tolerance thresholds (from left to right: 5, 10, and 20 time-units). F1-score is plotted against noise autocorrelation ρ for three methods: AR(1)-based segmentation (red), IID-based segmentation (black dashed), and the sliding-window approach (blue).

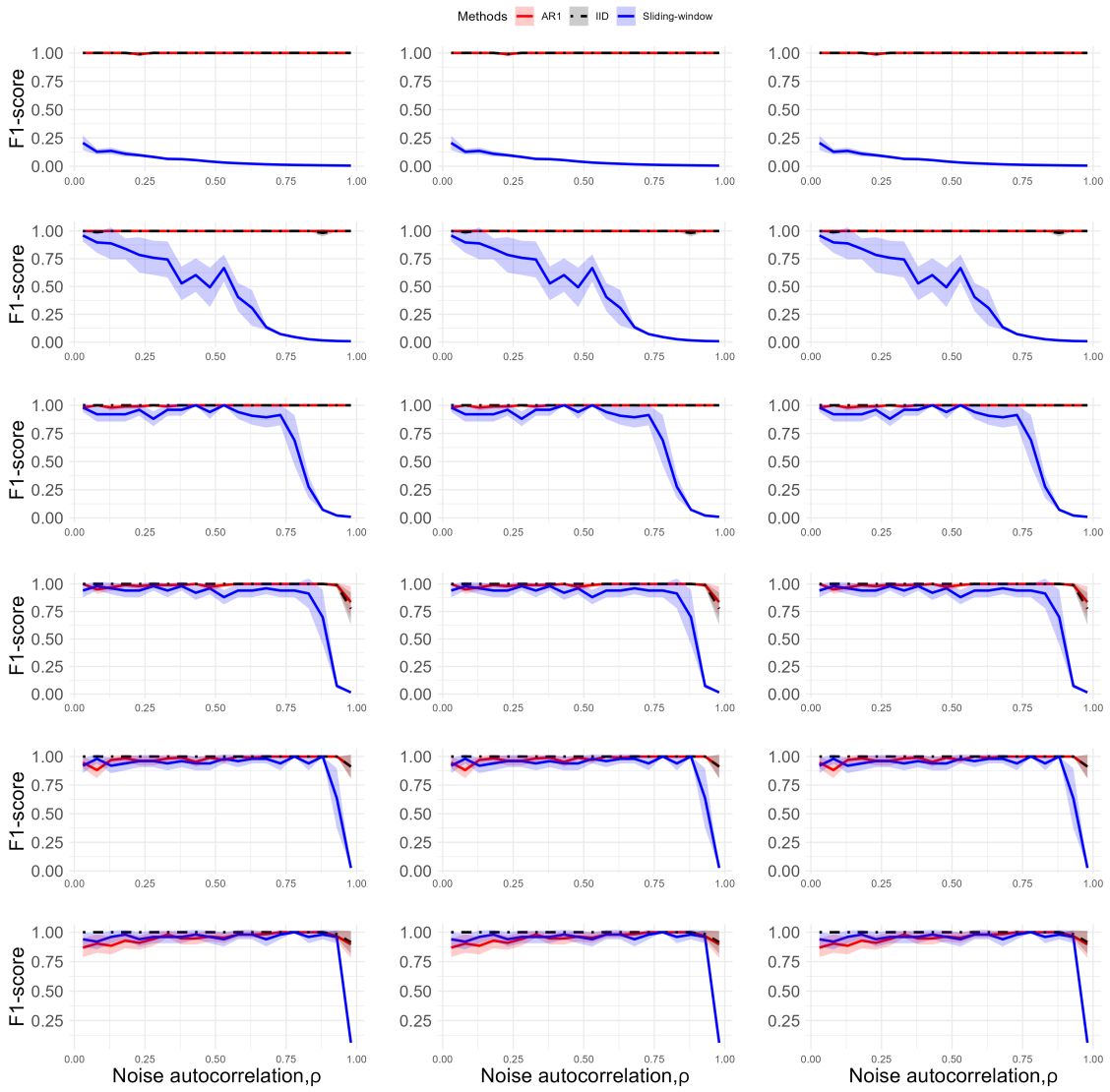

Fig. A.8: Sensitivity analysis of F1-score under varying acceptance rules for scenario C. Each panel shows results from one simulated trajectory under an angular AR(1) process. Rows represent different concentration parameters (κ), and columns represent different temporal tolerance thresholds (from left to right: 5, 10, and 20 time-units). F1-score is plotted against noise autocorrelation ρ for three methods: AR(1)-based segmentation (red), IID-based segmentation (black dashed), and the sliding-window approach (blue).

B.4 Extra Figures

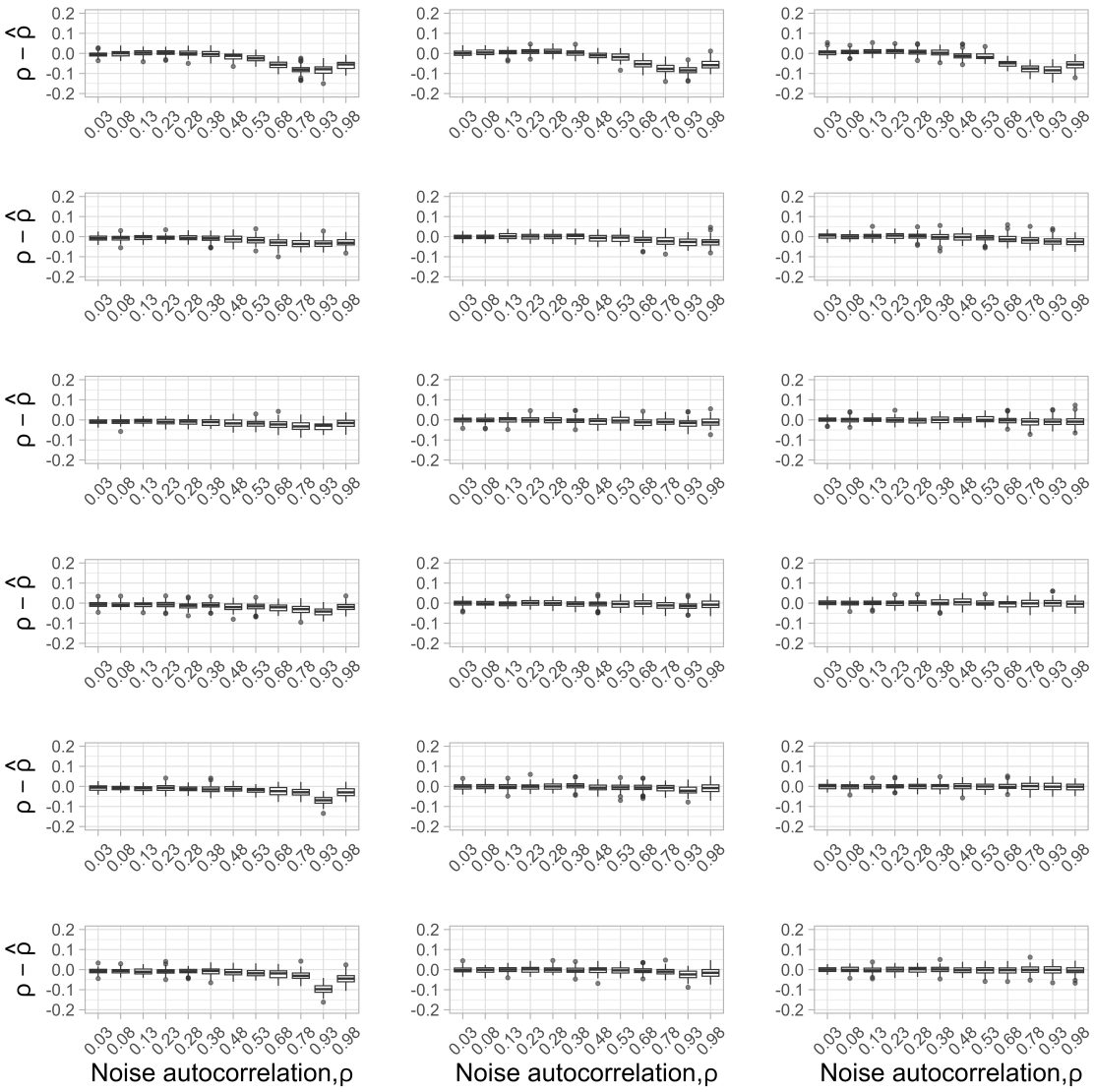

Fig. A.9: Error in estimation of ρ under the different scenarios from Section 2.4. Columns represent scenarios A, B, and C from left to right. Rows represent different κ_ρ values, with $\kappa_\rho = 5$, 15, 45, 100, 200 and 400 from top to bottom.

References

R. Maidstone, T. Hocking, G. Rigaill, and P. Fearnhead. On optimal multiple changepoint algorithms for large data. *Statistics and computing*, 27:519–533, 2017.

C. Truong, L. Oudre, and N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.