

Trend or persistence? Misclassification of long-term dependence in annual low-flow time series

Laaha Gregor

gregor.laaha@boku.ac.at

BOKU University

Research Article

Keywords: annual low flow, trend analysis, long-term persistence, short-memory autocorrelation, fractional differencing, Mann–Kendall

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11189154/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Trend or persistence? Misclassification of long-term dependence in annual low-flow time series

Laaha Gregor

BOKU University, Institute of Statistics, Vienna, Austria.

Contributing authors: gregor.laaha@boku.ac.at;

Abstract

Annual low-flow trends are commonly assessed using methods that account for serial dependence. However, long-term persistence (LTP) may be difficult to distinguish from deterministic trend and short-memory autocorrelation in hydrological records. We examined annual 7-day low-flow series from Austrian and European catchments using autoregressive and fractional-persistence models, supplemented by generic and gauge-calibrated simulations with known dependence structures.

Hydrological-year aggregation produced simpler AIC-selected autoregressive structures than calendar-year aggregation. In long observed records, autoregressive models were selected more frequently than fractional FARIMAX Type 1 models when the two dependence classes were compared. Generic AR(1) simulations with true fractional differencing parameter $d = 0$ showed that FARIMAX Type 1 and HKprocess could nevertheless diagnose statistically and practically relevant LTP when autoregressive terms were omitted. False practical-LTP diagnoses increased strongly with AR(1) dependence and persisted at longer simulated record lengths. Gauge-calibrated AR-only simulations showed the same behaviour for realistic low-flow dependence structures.

The results do not imply that LTP is absent from hydrological processes. Instead, they show that positive fractional-persistence estimates are not, by themselves, robust evidence of genuine long memory. LTP-based low-flow trend inference should therefore include explicit comparison with plausible short-memory autoregressive models and transparent reporting of dependence model uncertainty.

Keywords: annual low flow, trend analysis, long-term persistence, short-memory autocorrelation, fractional differencing, Mann–Kendall

Acknowledgements

The author thanks the members of the DWA working group HW 3.1 “Niedrigwasser” for inspiring discussions and constructive exchange on autocorrelation, long-term persistence, and trend analysis in low-flow series. The author particularly thanks Svenja Fischer for inspiring discussions during the initial development of the study. This study contributes to the UNESCO Intergovernmental Hydrological Programme FRIEND-Water initiative.

1 Introduction

Low flows are sensitive indicators of changes in climate, catchment storage, groundwater recharge, and water use. Detecting trends in annual low-flow indices is therefore important for water-resources planning and drought-risk assessment. Trend inference is complicated, however, when annual observations are serially dependent. Conventional Mann–Kendall and regression-based tests assume independent residuals; positive dependence can otherwise affect uncertainty estimates and trend-test performance.

Most hydrological applications address this problem as one of trend testing under short-memory autocorrelation. Proposed approaches include variance corrections, prewhitening, and block-bootstrap procedures (Hamed and Rao 1998; Yue et al. 2002; Yue and Wang 2004; Hamed 2009; Önöz and Bayazit 2012). These methods are intended to account for dependence that decays over relatively short lags. In contrast, long-term persistence (LTP), commonly represented by a Hurst parameter $H > 0.5$ or a positive fractional differencing parameter d , implies dependence over a broad range of time scales. LTP-aware trend methods may consequently produce substantially wider uncertainty bounds and fewer declared trends than methods that account only for short-memory dependence.

Long-term persistence has a long history in hydrology, beginning with the Hurst phenomenon observed in reservoir and river-flow records (Hurst 1951). Fractional time-series models have subsequently been proposed to represent slowly decaying dependence in hydrological records, including models that combine fractional, short-memory, and seasonal components (Montanari et al. 1997, 2000). For a hydrological discussion of long-term persistence, see also Szolgayová et al. (2014). However, distinguishing long-memory dependence from deterministic non-stationarity, structural change, and finite-sample short-memory behaviour remains difficult (Lins and Cohn 2011).

A central but less often examined issue is the inverse diagnostic problem: whether apparent LTP can be distinguished reliably from deterministic trend, short-memory autoregressive dependence, and mixed short- and long-memory structures in hydrological records of typical length. This distinction is difficult because both fractional and autoregressive processes can generate slowly varying sequences and substantial low-frequency variability. In particular, a fractional model fitted without autoregressive terms may represent short-memory dependence through a positive fractional differencing estimate. Thus, detection of $H > 0.5$ or $d > 0$ alone does not establish that the underlying process has genuine long memory.

Annual-series definition provides an additional practical consideration. For low flows, calendar-year aggregation can split a hydrological recession or recharge season between adjacent annual values. A hydrological year aligned with the seasonal cycle may therefore reduce apparent dependence between successive annual low-flow indices. Whether this affects the selected short-memory structure and subsequent interpretation of persistence has received limited attention.

This study examines trend, short-memory dependence, and apparent LTP in annual 7-day low-flow series using complementary observed-data and simulation analyses. First, we compare AIC-selected autoregressive structures for calendar-year and April-to-March hydrological-year series in Austrian and European station datasets. Second, we compare mutually exclusive fractional and autoregressive model structures in Austrian and European long records. Third, a factorial generic simulation assesses false LTP detection under known AR(1) dependence with true $d = 0$. Finally, gauge-calibrated simulations test whether the same mechanism occurs for fitted low-flow dependence structures from long observed records.

The study does not seek to determine whether LTP is absent from hydrological systems. Instead, it evaluates whether LTP-only procedures provide robust evidence for fractional persistence when plausible short-memory alternatives are considered. We show that annual-series definition affects selected short-memory complexity, that autoregressive structures are frequently preferred over fractional alternatives in observed low-flow records, and that LTP-only methods can diagnose practically relevant persistence in AR-only simulations. These findings support cautious use of LTP-based trend inference unless competing short-memory structures have been evaluated explicitly.

2 Data

2.1 Annual low-flow records

The study used annual 7-day low-flow indices derived from daily discharge records. The index represents the minimum moving 7-day mean discharge within each annual aggregation period. Austrian short-record data comprised 258 stations with complete or near-complete annual series for 1977–2019 ($n = 43$), based on the station selection and data compilation described by [Laaha et al. \(2025\)](#). The European short-record dataset comprised 782 stations with 40 annual observations, based on the station selection of [Laaha et al. \(2017\)](#). Both datasets were restricted to stations selected to be largely free from major anthropogenic disturbances of the low-flow regime, such as substantial regulation, water abstraction, or discharge augmentation.

For the Austrian dataset, annual indices were defined both by calendar year and by hydrological year starting in April. The comparison was designed to assess whether an annual definition aligned with the regional low-flow and recharge cycle alters the apparent short-memory structure. Corresponding calendar-year and hydrological-year series were also compared for the European dataset.

2.2 Long-record data

Long-record analyses used 92 Austrian stations with record lengths of 60–70 years and 28 European stations selected from the screened large-river dataset of Szolgayová et al. (2014). The underlying European discharge records were provided by the Global Runoff Data Centre (Global Runoff Data Centre 2025) and the UNESCO FRIEND database. These records provided the basis for comparing fractional-persistence and short-memory autoregressive representations. They also supplied the fitted dependence structures used in the gauge-calibrated simulation experiment.

2.3 Study design

The analyses were organised as complementary diagnostic steps. First, the short-record datasets were used to examine whether annual-series definition affects AIC-selected short-memory ARX structures. Second, the long-record datasets were used to compare mutually exclusive fractional and autoregressive dependence models. Third, generic simulations with known AR(1) dependence and true fractional differencing parameter $d = 0$ assessed the general potential for false LTP diagnoses. Finally, gauge-calibrated simulations examined this mechanism using fitted dependence structures and trends from long observed low-flow records.

3 Methods

3.1 Trend and dependence representations

We represented annual low-flow series as a deterministic linear trend plus a serially dependent residual process. Short-memory dependence was represented by autoregressive models, for which correlation declines geometrically with lag. Long-term persistence was represented by fractional differencing in FARIMAX Type 1 models. The fractional differencing parameter $d = 0$ corresponds to no fractional persistence, whereas positive d indicates persistent dependence. For simple fractional-noise processes, d is related approximately to the Hurst parameter by $H = d + 0.5$; this relation does not generally apply when additional autoregressive or moving-average structure is present.

All parametric models included a linear time regressor. Short-memory ARX models can be written as

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad \varepsilon_t = \sum_{j=1}^k \phi_j \varepsilon_{t-j} + u_t, \quad (1)$$

where β_1 is the deterministic trend coefficient, k is the AR order, and u_t is an innovation process. FARIMAX Type 1 models instead represented serial dependence through fractional differencing and contained no AR terms. Consequently, FARIMAX Type 1 and ARX models represent competing, rather than joint, descriptions of residual dependence in the principal comparisons.

3.2 Short-memory trend analyses

For each annual low-flow series, autoregressive models with a linear time regressor were fitted for candidate orders AR(0)–AR(4). Model order was selected using the Akaike information criterion (AIC; Akaike (1974)). The analysis was conducted separately for calendar-year and hydrological-year series. The resulting model orders represent AIC-selected ARX structures; they do not, by themselves, imply statistical significance of individual AR coefficients.

3.3 Long-term-persistence analyses

For the long-record datasets, FARIMAX Type 1 models containing a fractional differencing parameter d were compared with short-memory AR(1)–AR(3) models. All candidate models included a linear time regressor. FARIMAX Type 1 models were fitted with the `arfima` R package (Veenstra 2012), using Whittle-likelihood estimation of fractional differencing and no ARMA terms. The candidate structures were mutually exclusive: models contained either fractional dependence or autoregressive dependence, but not both. The AIC-selected model was retained for each station.

Evidence for positive long-term persistence required a positive fractional differencing estimate with $p < 0.05$. Autoregressive dependence was evaluated using the significance test of the corresponding AR coefficient. Trend significance was evaluated for the linear time regressor at $\alpha = 0.05$.

3.4 Generic simulation experiment

A generic simulation experiment assessed whether short-memory dependence can be misclassified as fractional persistence when long-term-persistence (LTP) methods are fitted without autoregressive terms. Simulated series consisted of an additive deterministic linear trend and stationary AR(1) noise,

$$Y_t = \beta_0 + \beta_1 t + Z_t, \quad Z_t = \phi Z_{t-1} + \varepsilon_t, \quad (2)$$

where the true fractional differencing parameter was $d = 0$ in every scenario.

The factorial design comprised record lengths of 40, 70, 100, and 200 years, AR(1) coefficients $\phi = 0, 0.2, 0.4$, and 0.6 , and linear trends of 0, 1, and 3% per decade. Each of the 48 scenarios was replicated 500 times, resulting in 24,000 simulated series. The reference flow was set to 100 and the marginal coefficient of variation of the AR noise was set to 20%.

For each simulated series, ARX(0)–ARX(3) models with a linear time regressor were compared using the corrected Akaike information criterion (AICc; Hurvich and Tsai (1989)). FARIMAX Type 1 and HKprocess (Tyralis and Koutsoyiannis 2014) were then evaluated as LTP-only procedures, i.e. without autoregressive terms. LTP-adjusted nonparametric trend procedures are commonly derived under fractional Gaussian-noise or scaling assumptions (Hamed 2008). Statistical positive LTP was declared for FARIMAX when $d > 0$ with $p < 0.05$, and for HKprocess when $H > 0.5$

with $p < 0.05$. Practical positive LTP was declared when the lower bound of a two-sided 95% Wald confidence interval for d exceeded 0.10, or when $H > 0.6$ with $p < 0.05$, respectively.

3.5 Gauge-calibrated simulation experiment

We conducted gauge-calibrated simulations for four long annual low-flow records: Würzburg on the Main, Děčín on the Elbe, Bratislava on the Danube, and Sala on the Váh. For each record, FARIMAX Type 1 and AR(1)–AR(3) models, each including a linear time regressor, were fitted and compared using AIC. The selected generating models for Würzburg, Děčín, and Bratislava were, respectively, AR(3), AR(1), and AR(1) models with linear trend and fractional differencing fixed at $d = 0$. These three scenarios therefore provide gauge-calibrated controls for false LTP detection. Sala was generated from a fitted fractional FARIMAX Type 1 model and was retained only as a sensitivity case; it was not interpreted as evidence of false LTP detection.

Simulations retained the fitted intercept, innovation variance, dependence parameters, and linear trend coefficient. One hundred replicates were generated for each station and record length. FARIMAX Type 1 was evaluated at the observed record length and at simulated lengths of 1,000, 5,000, and 10,000 years. HKprocess was evaluated only for simulations with observed record length. All FARIMAX fits converged successfully.

For FARIMAX Type 1, statistical LTP was defined as a positive fractional differencing estimate with $p < 0.05$, and practical LTP as a lower 95% Wald confidence bound for d above 0.10. For HKprocess, statistical LTP required $H > 0.5$ with $p(H) < 0.05$, and practical LTP required $H > 0.6$ with $p(H) < 0.05$. Practical positive trend detection was defined as a lower 95% trend bound above 1% per decade.

HKprocess analyses for extended simulated record lengths were not performed because the available implementation was computationally infeasible at these lengths. Modifying or parallelising the underlying HKprocess C routine was outside the scope of this study.

4 Results

4.1 Short-memory dependence in observed low-flow records

Hydrological-year series exhibited simpler AIC-selected short-memory structures than calendar-year series in both datasets (Fig. 1). In Austria, the proportion of AR(0) models increased from 65.1% for calendar-year series to 78.7% for hydrological-year series. The proportion of selected higher-order models, AR(2)–AR(4), decreased from 14.4% to 5.5%.

A similar pattern occurred in the European dataset. AR(0) models accounted for 71.7% of hydrological-year series, compared with 61.5% of calendar-year series. Higher-order AR models decreased from 10.6% for calendar-year series to 6.2% for hydrological-year series. Thus, defining annual low-flow indices using the hydrological year reduced the apparent short-memory complexity.

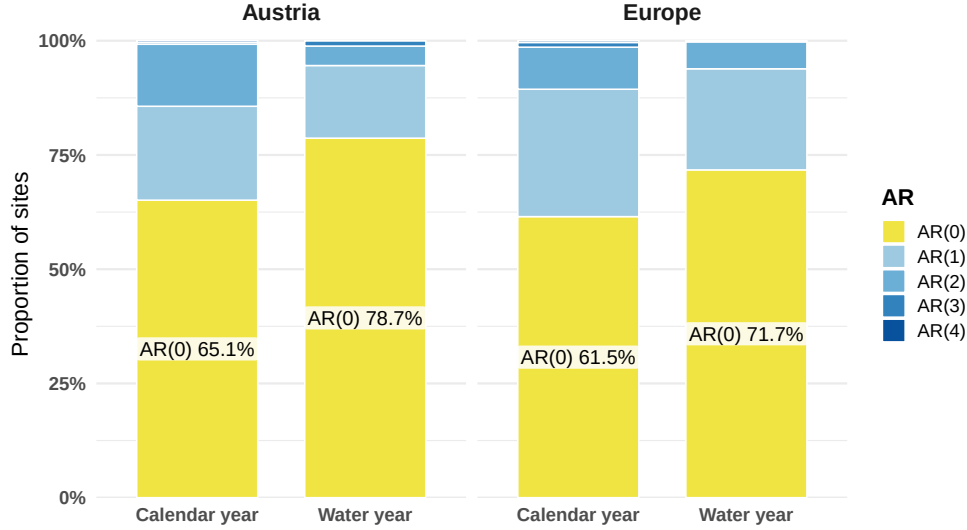


Fig. 1 Distribution of AIC-selected ARX model orders for annual 7-day low-flow series defined by calendar year and hydrological year. Results are shown for 258 Austrian records and 782 European records. Hydrological-year series had a higher proportion of AR(0) models and fewer selected higher-order AR models in both datasets.

4.2 Observed evidence for LTP versus short-memory dependence

In the AIC-based alternative model comparison, short-memory AR structures were selected more frequently than FARIMAX Type 1 models in both long-record datasets (Table 1). In Austria, AR models were selected for 59 of 92 records (64.1%), including 46 AR(1) models, whereas FARIMAX Type 1 was selected for 33 records (35.9%). Only two Austrian records showed significant positive fractional differencing, whereas significant AR(1) dependence occurred in 29 records.

For the European long-record dataset, AR structures were selected for 22 of 28 records (78.6%), including 17 AR(1) models. FARIMAX Type 1 was selected for six records (21.4%), of which two showed significant positive fractional differencing. Significant AR(1) dependence occurred in 14 GRDC records.

Significant linear trends occurred in 35 Austrian and 15 GRDC records. In Austria, 13 significant trends occurred in records for which FARIMAX Type 1 was AIC-selected, but none coincided with significant positive fractional differencing. Twenty-two significant trends occurred in AR(1) models, including seven records with both significant AR(1) dependence and trend. In the GRDC dataset, all 15 significant trends occurred in AIC-selected AR(1) models; 14 also had significant AR(1) dependence.

Table 1 AIC-selected dependence structures in long annual low-flow records. Positive fractional dependence denotes $d > 0$ with $p < 0.05$. The final column counts records for which both the selected dependence parameter and the linear trend term were significant.

Dataset	Model class	Selected	Sig. dependence	Sig. trend	Sig. dependence + trend
Austria	FARIMAX Type 1	33	2	13	0
	AR(1)	46	29	22	7
	AR(2)–AR(3)	13	0	0	0
GRDC	FARIMAX Type 1	6	2	0	0
	AR(1)	17	14	15	14
	AR(2)–AR(3)	5	0	0	0

4.3 Generic simulation: false LTP under AR(1) dependence

The generic simulations showed that short-memory AR(1) dependence can be misclassified as practically relevant LTP when LTP-only procedures are fitted to series with true $d = 0$ (Fig. 2). False practical-LTP diagnoses were negligible or low for independent noise ($\phi = 0$), but increased strongly with autoregressive dependence.

For $\phi = 0.4$, FARIMAX Type 1 classified 10.6–81.6% of the trend-free simulations as practically persistent, depending on record length. The corresponding HKprocess rates were 64.4–99.2%. For $\phi = 0.6$, FARIMAX practical-LTP rates ranged from 51.2% at $n = 40$ to 98.8% at $n = 200$; HKprocess rates ranged from 90.8% to 100%. These diagnoses were false by construction because all simulated processes had $d = 0$.

The pattern was similar in the simulations containing deterministic trends. Thus, the false LTP diagnoses were principally associated with unmodelled short-memory dependence rather than requiring a deterministic trend. Increasing record length did not eliminate the problem; instead, practical LTP diagnoses generally became more frequent as the AR(1) signal was estimated more precisely under the misspecified LTP-only models.

AICc-based ARX selection provided a contrasting short-memory representation. ARX(0) was selected most frequently for $\phi = 0$, whereas selection of ARX(1) increased with autoregressive dependence (Supplementary Fig. A2). Higher-order ARX models were selected less frequently. The trend-detection comparisons are reported as descriptive simulation results because inference after AICc model selection was not formally calibrated as post-selection inference.

4.4 Gauge-calibrated simulations

The gauge-calibrated simulations reproduced the generic finding that short-memory dependence can be diagnosed as fractional persistence when FARIMAX Type 1 is fitted without AR terms. In the AR-only control scenarios, the observed-length simulations for Würzburg and Děčín yielded median FARIMAX Type 1 estimates of $d = 0.220$ and $d = 0.285$, respectively, despite true $d = 0$ (Fig. 3). Practical LTP was falsely detected in 41% of Würzburg simulations and 63% of Děčín simulations. Corresponding false

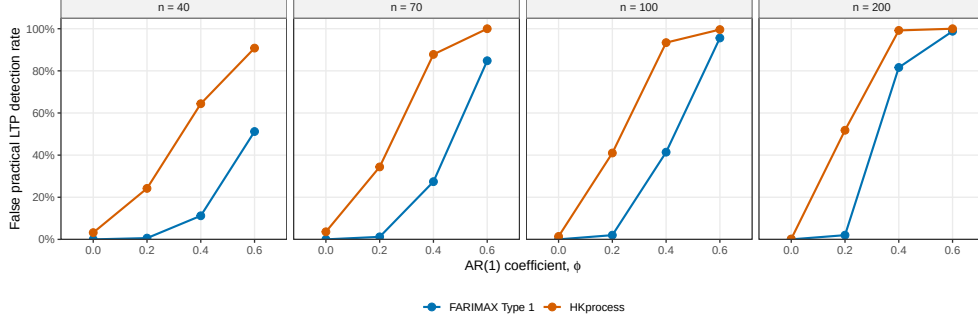


Fig. 2 False detection of practically relevant long-term persistence in trend-free AR(1) simulations. All series had true fractional differencing parameter $d = 0$ and no deterministic trend. Panels show record lengths of 40, 70, 100, and 200 years; each point represents 500 simulations. Practical positive LTP was declared when the lower bound of a two-sided 95% Wald confidence interval for the FARIMAX Type 1 fractional differencing estimate exceeded $d = 0.1$, or when HKprocess estimated $H > 0.6$ with $p < 0.05$.

practical-LTP rates from the observed-length HKprocess analyses were 98% and 100%, respectively.

The Bratislava AR(1) control represented weaker dependence. FARIMAX Type 1 produced a median $d = 0.030$ and no practical LTP diagnoses, whereas observed-length HKprocess analyses produced a median $H = 0.538$ and classified 21% of simulations as practically persistent. Thus, false practical LTP was most pronounced in the stronger AR-only scenarios, but was not confined to them.

Trend detection at observed record lengths was incomplete despite retained positive generating trends. Practical positive trends were detected by FARIMAX Type 1 in 52%, 43%, and 23% of simulations for Würzburg, Děčín, and Bratislava, respectively. HKprocess practical-trend detection rates were 28%, 21%, and 14%, respectively. These results describe method-specific detection frequencies under the simulated conditions; they do not constitute paired method comparisons.

HKprocess was evaluated only at observed record length; the extended-length results reported below are for FARIMAX Type 1 only.

Increasing record length did not resolve the FARIMAX Type 1 false-LTP diagnosis in the stronger AR-only controls (Fig. A3). For Würzburg, median d remained near 0.26 and practical LTP was detected in all simulations at 1,000, 5,000, and 10,000 years. For Děčín, median d remained near 0.31, with practical LTP likewise detected in all long-record simulations. In contrast, Bratislava retained median d values below 0.10 and no practical-LTP detections at all long record lengths. Median FARIMAX trend estimates remained close to the generating trends and their empirical uncertainty narrowed strongly with record length (Fig. A4).

Sala was a fractional-persistence sensitivity scenario rather than an AR-only control. Its results are therefore not interpreted as false-LTP evidence. Although statistical positive- d detections became frequent in long records, practical LTP remained absent at 1,000 and 5,000 years and occurred in 4% of simulations at 10,000 years.

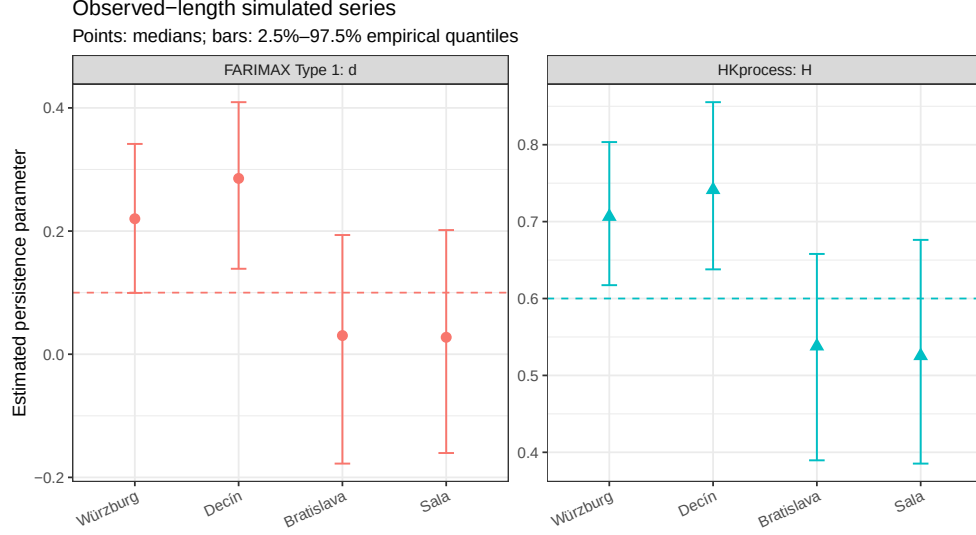


Fig. 3 Observed-length gauge-calibrated simulations. Points show median persistence estimates and bars show empirical 2.5%–97.5% quantiles across 100 valid simulations per station. FARIMAX Type 1 estimates fractional differencing d ; HKprocess estimates the Hurst parameter H . Horizontal lines mark the point-estimate reference values $d = 0.1$ and $H = 0.6$. Practical FARIMAX LTP required the lower bound of the two-sided 95% Wald confidence interval for d to exceed 0.1; practical HKprocess LTP required $H > 0.6$ with $p(H) < 0.05$. Würzburg, Děčín, and Bratislava were simulated from AR-only models with true $d = 0$; Sala was simulated from a fitted fractional-persistence model and is shown as a sensitivity case.

5 Discussion

5.1 Confounding of short- and long-memory dependence

The results show that apparent fractional persistence in annual low-flow series may arise from short-memory dependence when autoregressive alternatives are not considered. In the long-record datasets, AIC more frequently selected AR structures than FARIMAX Type 1 models, and statistically significant positive fractional differencing was rare. This does not demonstrate that LTP is absent from low-flow processes. Rather, it shows that, for these records, a parsimonious short-memory representation was often preferred when the two dependence classes were allowed to compete.

The generic simulations identify a mechanism underlying this result. All simulated series had true $d = 0$, yet FARIMAX Type 1 and HKprocess frequently diagnosed positive and practically relevant LTP as AR(1) dependence increased. The false diagnoses did not require a deterministic trend, as they occurred in the trend-free scenarios. They also did not disappear with increasing record length. This pattern is consistent with structural misspecification: an LTP-only model can represent persistent short-memory residual dependence through a positive fractional parameter when AR terms are excluded.

The gauge-calibrated simulations provide an observed-data analogue of this mechanism. The fitted AR-only controls for Würzburg, Děčín, and Bratislava were simulated with $d = 0$, but the stronger AR scenarios produced substantial positive fractional estimates under FARIMAX Type 1 and high persistence estimates under HKprocess. The results therefore concern identifiability of dependence structures under restricted model specifications, rather than a claim that fractional persistence cannot occur in hydrological systems.

5.2 Implications for low-flow trend inference

The choice of annual aggregation period affected the apparent short-memory structure. Hydrological-year series showed more frequent AR(0) selection and less frequent higher-order AR selection than calendar-year series in both the Austrian and European datasets. For Austria, the hydrological year begins in April and is intended to align annual low-flow indices with the seasonal hydrological cycle. Splitting a recession or recharge period across adjacent calendar years may otherwise induce additional dependence between annual indices. The appropriate hydrological-year definition will vary by region and catchment regime, but the results show that annual-series definition should be considered part of dependence diagnosis rather than merely a data-processing choice.

The results also have consequences for trend uncertainty. LTP-aware methods are often used because persistent residuals can imply wider uncertainty bounds than independent-error models. However, if short-memory dependence is misclassified as fractional persistence, this adjustment may be overly conservative and reduce trend-detection frequency. In the gauge-calibrated simulations, practical positive trends were not detected in all observed-length replicates despite retaining positive generating trends.

A practical implication is that positive estimates of d or H should not be interpreted in isolation. At minimum, LTP-based analyses should compare fractional models with plausible AR structures, report uncertainty in both the dependence and trend parameters, and distinguish statistical from practically relevant persistence. Model selection cannot establish the true physical dependence mechanism, but it can reveal whether a fractional interpretation is robust to parsimonious short-memory alternatives.

5.3 Limitations and practical recommendations

This study considered annual 7-day low-flow indices and a restricted set of dependence representations. The alternative comparison deliberately excluded models containing both fractional and autoregressive terms, because the aim was to examine competition between the two explanations and to avoid unstable finite-sample identification of jointly estimated dependence parameters. This design does not imply that mixed short- and long-memory processes are physically implausible. Fractional ARIMA-type models have been proposed for hydrological time series precisely to represent combined short-memory and long-memory dependence (Montanari et al. 1997), including seasonal dependence in monthly flow series (Montanari et al. 2000). More flexible ARFIMA

structures, nonlinear trends, regime shifts, moving-average terms, and covariate-based models may be appropriate in individual catchments and warrant further assessment.

The gauge-calibrated experiment used four fitted station-specific generating models and 100 replicates per scenario; it is therefore illustrative rather than a comprehensive calibration study. The simulated series retained fitted linear trends, so the experiment does not assess false positive trend detection under a strict no-trend null. HKprocess was evaluated only for observed-length simulations. Extended simulated record lengths were analysed with FARIMAX Type 1 only because the available HKprocess implementation was computationally infeasible at those lengths. Consequently, no direct FARIMAX–HKprocess comparison is made for the extended-length simulations.

For practical low-flow studies, we recommend first selecting an annual definition consistent with the catchment’s hydrological regime, then examining short-memory residual structure before attributing persistence to LTP. A fractional-persistence model should be compared with AR alternatives rather than fitted as the only dependence representation. Where LTP is retained, conclusions should state the competing models considered, the confidence interval for d or H , and the sensitivity of trend inference to the dependence specification.

6 Conclusions

This study examined whether annual low-flow records can distinguish short-memory autocorrelation from long-term persistence. Hydrological-year aggregation produced simpler selected ARX structures than calendar-year aggregation. In long observed records, autoregressive models were selected more frequently than fractional FARIMAX Type 1 models, and significant positive fractional dependence was uncommon.

Generic and gauge-calibrated simulations showed that LTP-only procedures can diagnose practically relevant persistence in series generated with short-memory autoregressive dependence and true $d = 0$. In the stronger AR-only controls, this false LTP diagnosis persisted even for very long simulated records. LTP-based trend inference should therefore be accompanied by explicit comparison with plausible short-memory models and by transparent reporting of dependence-model uncertainty.

Statements and Declarations

Funding. This work was supported by the Climate and Energy Fund under the Austrian Climate Research Programme (ACRP; grant no. C265154).

Competing interests. The author has no relevant financial or non-financial interests to disclose.

Data availability. This study used publicly available European streamflow records obtained from the Global Runoff Data Centre (GRDC; https://www.bafg.de/GRDC/EN/Home/homepage_node.html) and Austrian streamflow records obtained from the Austrian Hydrographic Service through the eHYD portal (<https://ehyd.gv.at/>). Data are available from the respective data providers, subject to their terms of access and use.

Code availability. HKprocess analyses used the archived R package `HKprocess`, version 0.1-1. Code used for data processing, analysis, and simulations can be made available by the author on reasonable request.

Author contributions. Gregor Laaha conceived and designed the study; compiled, processed, and analysed the data; developed and conducted the simulations; interpreted the results; prepared the figures; wrote the first draft of the manuscript; and reviewed and approved the final manuscript.

References

- Akaike, H.: A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**(6), 716–723 (1974) <https://doi.org/10.1109/TAC.1974.1100705>
- Global Runoff Data Centre: The Global Runoff Data Centre. Federal Institute of Hydrology (BfG), Koblenz, Germany. Accessed 8 March 2025 (2025). <https://grdc.bafg.de/>
- Hamed, K.H.: Trend detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. *Journal of Hydrology* **349**(3–4), 350–363 (2008) <https://doi.org/10.1016/j.jhydrol.2007.11.009>
- Hamed, K.H.: Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. *Journal of Hydrology* **368**(1–4), 143–155 (2009) <https://doi.org/10.1016/j.jhydrol.2009.00675>
- Hamed, K.H., Rao, A.R.: A modified Mann–Kendall trend test for autocorrelated data. *Journal of Hydrology* **204**(1–4), 182–196 (1998) [https://doi.org/10.1016/S0022-1694\(97\)00125-X](https://doi.org/10.1016/S0022-1694(97)00125-X)
- Hurvich, C.M., Tsai, C.-L.: Regression and time series model selection in small samples. *Biometrika* **76**(2), 297–307 (1989) <https://doi.org/10.1093/biomet/76.2.297>
- Hurst, H.E.: Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* **116**, 770–799 (1951)
- Lins, H.F., Cohn, T.A.: Stationarity: Wanted dead or alive? *Journal of the American Water Resources Association* **47**(3), 475–480 (2011) <https://doi.org/10.1111/j.1752-1688.2011.00542.x>
- Laaha, G., Gauster, T., Tallaksen, L.M., Vidal, J.-P., Stahl, K., Prudhomme, C., Heudorfer, B., Vlnas, R., Ionita, M., Van Lanen, H.A.J., Adler, M.-J., Caillouet, L., Delus, C., Fendekova, M., Gailliez, S., Hannaford, J., Kingston, D., Van Loon, A.F., Mediero, L., Osuch, M., Romanowicz, R., Sauquet, E., Stagge, J.H.,

- Wong, W.K.: The European 2015 drought from a hydrological perspective. *Hydrology and Earth System Sciences* **21**(6), 3001–3024 (2017) <https://doi.org/10.5194/hess-21-3001-2017> . Accessed 2017-09-08
- Laaha, G., Laimighofer, J., Parajka, J., Bertola, M., Blöschl, G.: Abflusstrends in Österreichs Fließgewässern 1977–2020. *Österreichische Wasser- und Abfallwirtschaft* **77**(5), 338–346 (2025) <https://doi.org/10.1007/s00506-024-01106-8>
- Montanari, A., Rosso, R., Taqqu, M.S.: Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation. *Water Resources Research* **33**(5), 1035–1044 (1997) <https://doi.org/10.1029/97WR00043>
- Montanari, A., Rosso, R., Taqqu, M.S.: A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan. *Water Resources Research* **36**(5), 1249–1259 (2000) <https://doi.org/10.1029/2000WR900012>
- Szolgayová, E., Laaha, G., Blöschl, G., Bucher, C.: Factors influencing long range dependence in streamflow of european rivers. *Hydrological Processes* **28**(4), 1573–1586 (2014) <https://doi.org/10.1002/hyp.9694>
- Tyralis, H., Koutsoyiannis, D.: A bayesian statistical model for deriving the predictive distribution of hydroclimatic variables. *Climate Dynamics* **42**, 2867–2883 (2014) <https://doi.org/10.1007/s00382-013-1804-y>
- Veenstra, J.Q.: Persistence and anti-persistence: Theory and software. PhD thesis, Western University (2012)
- Yue, S., Pilon, P., Phinney, B., Cavadias, G.: The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes* **16**(9), 1807–1829 (2002) <https://doi.org/10.1002/hyp.1095>
- Yue, S., Wang, C.: The Mann–Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water Resources Management* **18**(3), 201–218 (2004) <https://doi.org/10.1023/B:WARM.0000043140.61082.60>
- Önöz, B., Bayazit, M.: Block bootstrap for Mann–Kendall trend test of serially dependent data. *Hydrological Processes* **26**(23), 3552–3560 (2012) <https://doi.org/10.1002/hyp.8438>

Appendix A Supplement

Supplementary information. Supplementary material contains additional results, model diagnostics, sensitivity analyses at $\alpha = 0.05$, and station-level results.

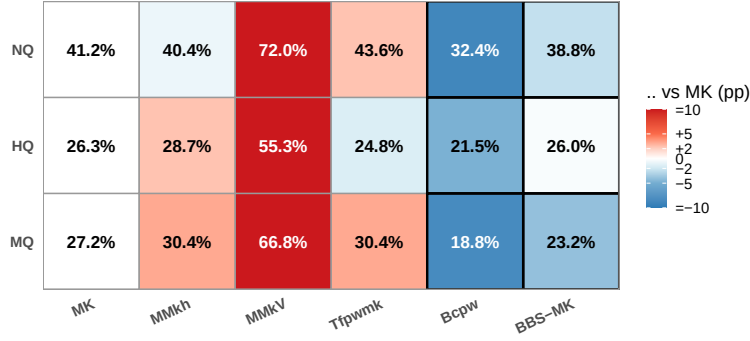


Fig. A1 Significant-trend rates obtained with Mann–Kendall (MK) and modified MK procedures for Austrian annual mean-flow (MQ), high-flow (HQ), and low-flow (NQ) series. Labels show absolute rates; colours show differences from the conventional MK result in percentage points. Results are reproduced from Laaha et al. [Laaha et al. \(2025\)](#).

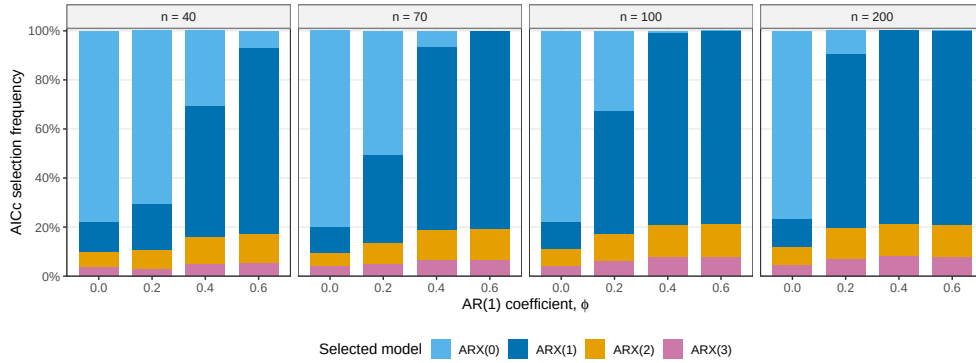


Fig. A2 AICc-based selection of short-memory ARX models under AR(1) dependence. Bars show the frequency with which candidate ARX(0)–ARX(3) models were selected using AICc. Each bar pools 1,500 simulations across three deterministic-trend scenarios. ARX(0) is the true model for $\phi = 0$, whereas ARX(1) is the true generating structure for $\phi > 0$.

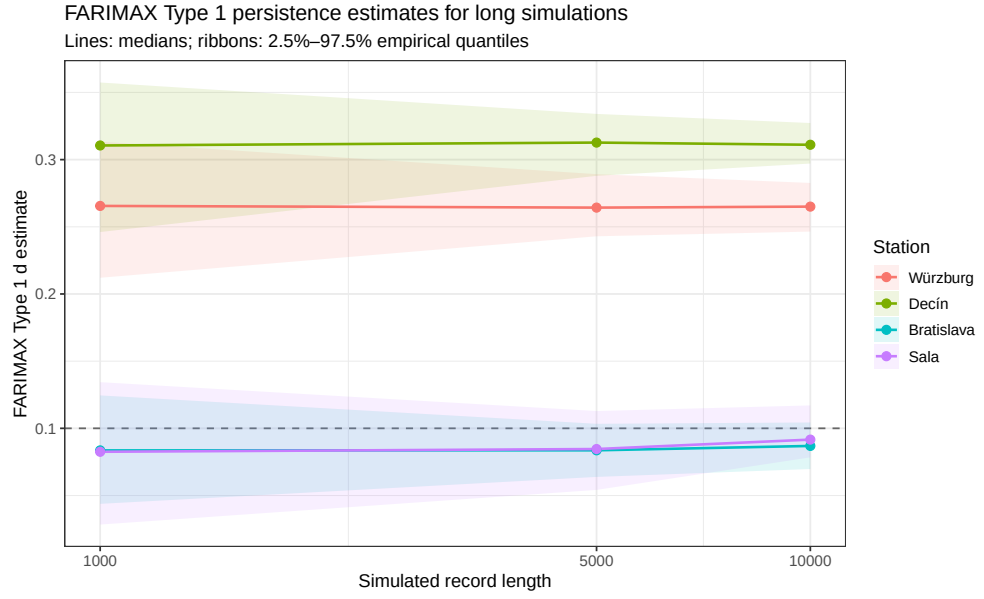


Fig. A3 FARIMAX Type 1 fractional-differencing estimates in long gauge-calibrated simulations. Lines show medians and ribbons show empirical 2.5%–97.5% quantiles across 100 valid simulations per station and record length. The dashed line marks the point-estimate reference value $d = 0.1$. Würzburg, Děčín, and Bratislava are AR-only generating processes with true $d = 0$; Sala is a fractional-persistence sensitivity scenario.

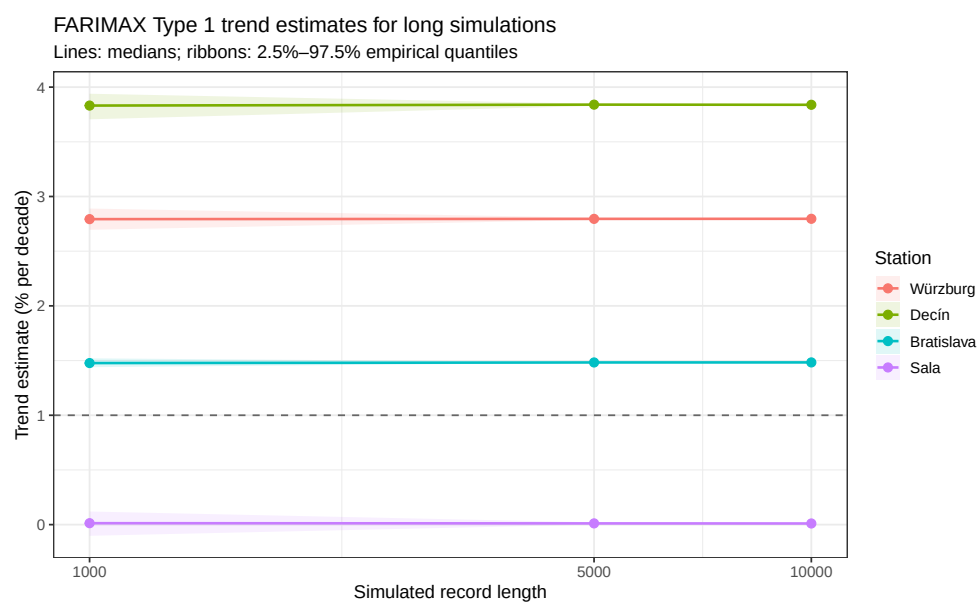


Fig. A4 FARIMAX Type 1 trend estimates in long gauge-calibrated simulations. Lines show median trends and ribbons show empirical 2.5%–97.5% quantiles across 100 valid simulations per station and record length. The dashed line marks the point-estimate reference value of 1% per decade.