Dear Editors and Reviewers,

Thank you for your comments and the opportunity to revise our manuscript. This document contains our point-by-point responses (in blue) to your comments (in black). We believe the manuscript has improved significantly and hope that it addressed the review team’s concerns. The major changes to the manuscript are as follows:

Major Changes to Manuscript

1. **Repositioning of the paper**. To target a wider audience, we have re-positioned the paper as proposing a machine learning-based time series feature selection method paired with a swapping mechanism for data privacy. The proposed method (k-nTS+) is flexible because it can be applied to multiple forecast horizons, forecasting models, forecast accuracy metrics, and correlated time series features. Furthermore, our paper demonstrates an emerging research theme by incorporating the usefulness (forecast accuracy) of the protected data into the data protection process itself. which is also applicable to other domains such as missing data imputation, nowcasting, and outlier replacement.
2. **New datasets.** To increase the generalizability of our results, we have included all of the M3 competition data in the empirical application. Also, we applied our methodology to a subset of the hierarchical M5 competition data which represents real-world retail data. Using the new data, we find that…
3. **Readability of the paper.** We have rewritten the paper in LaTeX and revised the algorithms and figures within the paper. Specifically, we have revised the Introduction on pages XXX to XXX, Figures 1 and 2 on pages XXX and XXX, and the Conclusion on pages XXXX to XXX. We have uploaded all code to Github for reproducibility and posted the revised draft on a pre-print server.
4. **Contributions of the paper**. We have carefully rewritten to the claimed contributions of the paper over the prior literature to include:
   1. Created a data protection method that enables organizations to share protected time series data with good forecast accuracy and useful time series features. Compared to benchmark methods, our proposed method balances the trade-off between forecast accuracy (14% worse) and data privacy well (3.3% reidentification risk). Furthermore, we show that differential privacy with a high value of epsilon (10) can result in a reidentification risk (50%) that is not protective.
   2. Applied a machine learning-based feature selection approach to improve forecast accuracy of swapped data (14% worse than original) over manual feature selection (40% worse). We show that the most useful features for predicting forecast accuracy (Makridakis et al., 2018; Spiliotis et al., 2020) are not necessarily the most useful for swapping time series values.
   3. We demonstrate how time series features change for different forms of data protection and how changes in these features affect multiple forecasting models.

In full disclosure, an indexing error in the code caused only k-1 nearest neighbor time series to be used for swapping instead of k nearest neighbor time series. This changed the main results of XXX to XXX…

**Editor Comments:**      
  
The manuscript concentrates on a timely and relevant topic. The approach is original and the point is well supported. The reviewers and I see that the manuscript potentially brings some interesting novel ideas that could be worth of publication. However, at this stage, there seems to be quite a lot of work to be done for the paper to get there.

Thank you for your helpful feedback. We hope that you find the proposed changes satisfactory.

Besides the point of the reviewers, I would like to insist on the fact the presentation of the work needs a serious upgrade. For instance, A figure like Figure 1 cannot end up in the final version of the paper. I would encourage the authors to be careful in the way to design and produce that figure.

* Removed figure 1. Focused on the series with desirable/undesirable features.
* General discussion – features like trend appear important, we want to consider many possible features

Similarly, in general, the paper looks like a draft, which makes it a bit tricky at stages (e.g., when I checked the validity of equations...). Using a more "profession" text editor (possibly Latex) could help a lot in improving the readability and make the work easier for the reviewers and I. Finally, for the maths, it would be good to be really thorough, e.g., one does not need to use "\*" for multiplication, the flow of information around eq. (2) is quite confusing (possibly introduce relevant notation first, for diag, 1, among others), etc.. Similarly, algorithms could be better presented.

We apologize for the submitted presentation of the paper…. Thank you for the feedback. Per the “Major Changes to the Manuscript,” we have repositioned the paper (Point 1) and re-written the draft (Point 3) in LaTeX including the figures and algorithms.

**Reviewer 1 Comments:**

The authors propose a matrix-based privacy method called k-nearest time series + (k-nTS+) swapping that preserves time series features to maintain forecast accuracy. The proposed privacy method has been applied to a forecasting competition data set and proven its advantages through a series of empirical studies. Overall, the paper is well-structured and written, while its contribution is clearly explained and justified.

Thank you for your constructive feedback on our manuscript. Please see the “Major Changes to the Manuscript” and our comments directly to you below. We hope that the revised version is much improved.

Below you may find some comments that could help further improve the current work.  
  
1. Page 5: When first introducing the k-nTS+ swapping method in Figure 2, the authors should provide more details on how it works. The framework now is primitive.

Thank you for this suggestion, we have increased the amount of detail in Figure X on page XX to better illustrate our proposed k-nTS+ method. We have also included the full details of the algorithm on page XXXX. To reduce confusion, we have removed Figure XXX on page XXXX and point the readers to Section XXX where the k-nTS+ swapping method is explicitly explained.

2. Section 2.3:  The literature review of time series features for forecast accuracy could profit from including the relevant works such as:  
  
   Kang Y, Cao W, Petropoulos F, et al. Forecast with forecasts: Diversity matters[J]. European Journal of Operational Research, 2022, 301(1): 180-190.  
  
   Li L, Kang Y, Petropoulos F, et al. Feature-based intermittent demand forecast combinations: accuracy and inventory implications[J]. International Journal of Production Research, 2022: 1-16.  
  
   Montero-Manso P, Athanasopoulos G, Hyndman R J, et al. FFORMA: Feature-based forecast model averaging[J]. International Journal of Forecasting, 2020, 36(1): 86-92.

We greatly appreciate these suggestions and added these references to our literature review on page XXX. Additionally, in our results section on page XXX, we show that XXXX. Compared to Literature (XXX), our results suggest that….

3. Section 4.1: Why only use the monthly micro dataset from M3 competition? I recommend using all M3 competition data and discussing the performance of the proposed k-nTS+ method for the data with different frequencies. More recent M4 Competition data is also a better option.

Per the “Major Changes to the Manuscript” (Point 2), we now analyze the full M3 competition data set. The results are included on pages XX – XX. We also analyze a subset of the M5 competition to test the proposed method on hierarchical forecasts in a retail setting. The results are listed on pages XXXX to XXXX.

A screenshot of a computer

Description automatically generated

Compared to the prior version of the manuscript, we find that:

* Full M3 Data
  + Different frequencies
  + Other, include a table
* M5 Data
  + Retail data is more privacy sensitive (???). Prior research (Li et. al. 2022) discuss that retail data is not private.
  + The swapping mechanism performs poorly (???) when forecasts are aggregated up to the next category. This is because our nearest time series method can swap values from *different* trees in the hierarchy based on similarity. On page XXXX, we list this as a limitation of our proposed method and suggest future research on incorporating pre-defined hierarchical structures into the data protection process itself.

4. Section 4.2: The authors should clarify the details of the feature selection, e.g., why select such features for k-nTS and add new features for k-nTS+. The process seems subjective. The authors should give all the alternative features and explain the reason.

We apologize for the lack of clarity on the previous manuscript. To clarify this point, we try to avoid selection bias by including all features which could help to improve accuracy in the swapping process. On page XXX, we now differentiate between our manual feature selection based on the literature (sources here XX) and our proposed feature selection method.

The proposed feature selection method uses all time series features and selects those features which balance the trade-off between forecast accuracy and data privacy well. In our application, we select over XXX features that (1) occur commonly in the literature (Sources XX,XX), and (2) can be conveniently computed using the *tsfeatures* package in R. After the process is completed, the data protection methodology selects the best 5 or 6 features based on the out of bag error for each forecasting model.

The results show that regardless of forecasting model chosen, the data protection methodology retains the same 4 out of 6 features. To further clarify this, we now include Table XXXX on page XXX which displays the average ranking of time series features selected for k-nTS+. We hope this helps clarify the details of the feature selection process.

5. A brief discussion of the computational cost is useful for other researchers.

Thank you for brining this omission to our attention. We added Table XXX on page XXX to display the computation cost as the number of time series increases.

* Using simple model, slow part is feature selection
* Using complex model, slow part is generating baseline forecasts

6. Is there any reason for the error measures used in sections 4.4 and 4.5 to differ? If MSE does not provide significantly different results than MAE, personally I would prefer a consistent measure to be used for both sections.

We now use MAE to assess both the accuracy of forecasts in Section 4.4 and the out of bag errors for the random forest predictions of the forecast MAE in Section 4.5.

7. Figure 6: Each diagram should be numbered differently, such as A.1, A.2, A.3, B.1, B.2, B.3.

Thank you, we have adjusted Figure XXX (previously Figure 6) accordingly.

8. Figure 7: The ordinate title should not have "average"? It is unclear what "time series features for each privacy method" means exactly.

Thank you we have corrected the titles in Figure XX (previously Figure 7).

**Reviewer 2 Comments:**

The authors propose a method for preserving data privacy in time series data through a swapping technique. This approach focuses on maintaining forecast accuracy by swapping the data values only if the essential features of the time series, such as mean and autocorrelation function (ACF), are likely to remain unchanged. The proposed method assumes a centralized approach, where a single data owner possesses the time series data. In this scenario, a forecaster selects a forecasting model F, and the data owner performs data swapping to prevent a decline in F accuracy.

Thank you very much for your helpful comments. We hope that the newly revised manuscript addressed your main concerns.   
  
While the idea sounds interesting, the paper requires further clarification and enhancements to address the following points:  
  
- Applications: The authors should provide further clarification on the potential applications of their proposed method.

Thank you. Per the “Major Changes to the Manuscript” (Point 1), we discuss potential applications of our proposed method swapping method, including imputing missing values, nowcasting, and anonymizing sensitive time series, on page XXX.

* Citations for using features for these applications

- The authors should explain why the data owner cannot provide the forecaster with the original or degraded model weights. This comparison would help illustrate the advantages and limitations of the proposed approach.

Thank you for the suggestion. To address the question on whether the data owner can provide the forecasters with the original model weights, we save the parameters of the VAR models trained on the M3 data and generated a simulated time series corresponding to each original series.

We show that the main advantage of our methodology is that through sharing protected time series, forecasters gain access to the entire distribution of time series values, rather than just model weights. The protected data provides much better coverage on the distribution of the original time series features compared to time series simulated from the VAR model weights (see Figure below). The simulated series provided good protection against re-identification (on average only 5.8% of the simulated series were identified correctly) but at a high cost to the utility of the data. In addition, our method does not have to assume a single forecasting model, rather, we create a protected data set that works well for several forecasting models of varying complexities.

**A graph of a diagram

Description automatically generated with medium confidence**

Need to compare original weights vs. degraded weights vs. original data vs. protected data. Can we use the VAR model as an illustration as we suggested and simulate time series?? Best approach here… XXXX

Additionally, discussing whether this method could be extended or applied to cases where data are owned by multiple data owners (decentralized scenario) would be beneficial.

Thank you for the comment. We discuss this scenario on page XX in the conclusion.

- A crucial aspect missing is the impact of the swapping process on forecasting accuracy in a real-world setting. For instance, if a forecaster aims to perform a one-timestep ahead forecast using a VAR (Vector Autoregression) model with a lag of 1, he would require the value of X[T] to predict X[T+1]. Therefore, it is essential to evaluate whether the swapping process can change the last point of the time series without significantly affecting the forecasting accuracy.

To clarify from the previous version of the paper, our existing swapping process changed all past time series values including value X[T]. T values for each time series were swapped at time T when the data owner decided to protect all its past data. On average, this resulted in a 14% decrease in forecast accuracy. To address your comment, we expect a change in only the last point (X[T]) to be much less (Can we include how much here or is it too much work?), but this is not suggested because all of the time series data must be protected.

To investigate the impact of swapping on a VAR model with a lag of 1, we illustrate how the swapping process changed model weights (coefficients) and forecast accuracy on page XXXX. Interestingly, differential privacy on X[T] caused the model weights to change drastically which caused a 1000% degradation in forecast accuracy. Mathematical results are listed on page XXX.

Figure 6 in the paper shows that the last point in the window is very close to the original one. Does this mean that a curious forecaster could reconstruct the data by running the model for some time?

Thank you for this comment. Reconstruction attacks are a valid concern with non-differentially private protection mechanisms. In our paper, we measured the reidentification risk of an intruder which is only one form of a privacy attack. We simulated an intruder matching 10 time series values in a row to the protected dataset and found that our method performs quite well (X% reidentification risk). In our Conclusion section on page XXX, we now note the limitation that we only considered one privacy attack and it’s possible for an intruder to attack the data in another way.

However, with the theoretically private method of differential privacy, at acceptable levels of , we showed that the forecasts are unusable. When increasing , the forecast accuracy improved while maintaining some bound on overall privacy risk. However, our empirical application shows that results in significantly higher identification disclosure risk (50% for epsilon of 10) than our proposed method k-nTS+. Essentially, the differentially private data is either not useful or not protective and there is no acceptable trade-off between data privacy and forecast accuracy.

Overall, the message of our paper is that the usefulness (forecast accuracy) of anonymized data should be a forethought, rather than an afterthought (Blanco Justicia et. al. 2022)

Detailed comments:  
- Acronyms meaning is missing. Some examples: SES, DES, LGBM, OOB, MSE, MAE, etc.

- Equations should not be figures, e.g., (1) and (5).

- Figures 2 and 8 should be of better quality.

- Figure 4: I suggest a unique plot with different color/shape lines representing the methods.

- Notation needs to be introduced appropriately in many equations - see, e.g., (9).

Thanks for these minor comments. We have corrected all of them in the paper.

- I would include the detailed proposal version in the main text.

We have included the detailed version of the algorithm in the main text on page XX.

- Text needs revision. Typo example: "matices", sometimes ':' is used instead of '='.

Thank you. We have carefully edited the entire paper. We hope that the revised manuscript addresses your concerns and we sincerely appreciate your enhancements to our paper.