**Summary of Results Analyzing k-nTS+ (k = 3) Performance on M3**

Variants on k-nTS+:

* Version in original paper draft
* With steps similar to deja-vu paper: seasonal decomposition, loess smoothing, scaling by the point one period before the forecast origin
* Same as above but with longer swapping windows
* With scaling by the point one period before the forecast origin and outlier replacement in protected versions
* With data supplemented using synthetic time series (selected the 15 series most similar to each original series on the features chosen by RReliefF and RFE)

*Privacy and Accuracy Results for k-nTS+ (k = 3) variants. “% Change MAE” is the average percent change in absolute error across all protected series. “Change MASE” is the difference (Protected MASE – Original MASE) where the denominator is the in-sample error of the naïve/seasonal naïve model on the original version of the series. “Ident. Prop.” is the average proportion of time series which were correctly identified when an adversary had access to 10 confidential data points from each original series and used these to predict the identities of each series. This is then averaged across 20 simulations and a weighted average (based on the number of series in each data set) is then calculated.*

|  |  |  |  |
| --- | --- | --- | --- |
| Version | % Change MAE | Change MASE | Ident. Prop. |
| Original k-nTS+ | 127% | 1.81 | 0.055 |
| k-nTS+ with Preprocessing | 24% | 0.25 | 0.482 |
| k-nTS+ with Preprocessing and longer windows | 28% | 0.31 | 0.384 |
| k-nTS+ with scaling and outlier replacement | 34% | 0.37 | 0.324 |
| k-nTS+ augmented with synthetic series (GRATIS) | 244% | 4.12 | 0.095 |

The protected series from the original version of k-nTS+ were identified a similar proportion of the time across all of the M3 data compared to our previous results on the just the M3 monthly micro data (5.5% compared to 3.3%). These privacy results are where we want them, but accuracy is unacceptable (127% increase in MAE on average, and an average increase in MASE of 1.81. On the other hand, if we take measures to improve the forecast accuracy by performing seasonal decomposition, smoothing, scaling the swapped values to the original series, and adding the seasonal component, our accuracy improves significantly (a 24% increase in MAE and 0.25 increase in MASE) but our privacy results are unacceptable: 48% of series are correctly identified on average. Increasing the length of the swapping windows reduces accuracy (28% increase in MAE and 0.31 increase in MASE) and improves privacy to 38% identified series (which is still unacceptable). If we do not perform the seasonal decomposition or smoothing, and instead only perform the scaling and replace any outliers, we still have about 32% of series identified on average, with accuracy results that are worse than with the full pre-processing steps. The results using synthetic time series are inconsistent, since privacy is slightly worse than under the original k-nTS+ (9.5% compared to 5.5% identified) but forecast accuracy is significantly worse (244% increase in MAE).

The original k-nTS+ likely provides the best privacy protection because exact confidential values are swapped across all series. So the adversary has difficulty distinguishing the identities of protected series since the known confidential values could match multiple protected series. The versions of k-nTS+ with preprocessing apply scaling to the swapped values to improve their similarity to the values of the original series, hence privacy is worse, not to mention the pre-processing step that preserved the seasonality of the original series.

I did another analysis operating under the hypothesis that data sets with series that have different feature values will have worse forecast accuracy under k-nTS+.

Our goal is to understand why our method performed so well on the M3 monthly micro data, but so poorly on the rest of the M3 data sets. My theory is that the similarity (based on features) and the spectral entropies (proxying for how well the original series can be forecast) will help determine when our method will perform well and when it won't. Our method is selecting a few features that are not only more likely to vary across time series with different forecast accuracies (RReliefF), but they are also almost as predictive as using all selected features to predict forecast accuracy across the original and baseline protected data sets (RFE). So, if we include these features in the swapping process, we should maintain their values (assuming there are similar series to swap with) and thereby maintain forecast accuracy. HOWEVER, we saw in the previous M3 monthly micro results that the spectral entropies of the protected series tended to be higher than those of the original series. This makes sense - even though we are preserving features that are important/predictive for forecast accuracy, we are introducing randomness into the time series, which reduces autocorrelation, which increases spectral entropy. Past research (Spiliotis et al. 2020) shows through multiple linear regression that, holding many other features constant (strength of trend, strength of seasonality, skewness, kurtosis, etc.) an increase in spectral entropy is associated with an increase in the MASE of the forecasted M4 time series at a statistically significant level. I assume this result would extend to other time series as well.

Based on these findings, I wanted to see whether data sets with time series that (1) are highly similar on the selected features, and (2) have high

spectral entropy will have the best results from k-nTS+. Data sets with series that are less similar on the important features may not be able to swap values without destroying forecast accuracy. And, even if a data set has time series with very similar features, the forecast accuracy for series with low spectral entropies may still be unacceptable since we are introducing randomness.

Kang et al. (2020) suggest using t-SNE to visualize the feature similarity of time series since linear dimension reduction techniques (PCA) can produce misleading results when many potentially non-linearly correlated features are present. Note that we compute feature similarity within the data subsets with series that have the same length.

***Fig 1:*** *plots of t-SNE dimension reduced feature vectors. The original feature vectors of each data subset contained only the features that were chosen using our feature selection method. Points are colored by the spectral entropy of the original series.*

A graph of different types of numbers

Description automatically generated with medium confidence

Figure 1 shows that there is a lot of variation in the feature similarity across the data subsets. Even within the monthly micro subsets, subsets 2 and 3 contain series that are much more similar to each other relative to subset 1. These subsets also have some of the highest spectral entropies (on average).

*Table 2: Data subsets with the ten highest average spectral entropies (calculated across the original time series).*

|  |  |  |
| --- | --- | --- |
| **Data Subset** | **Average Spectral Entropy** | **Percent Change MAE** |
| monthly-MICRO1 | 0.9623388 | -7.53 |
| monthly-MICRO2 | 0.8624881 | 19.72 |
| quarterly-FINANCE1 | 0.6836378 |  |
| monthly-INDUSTRY1 | 0.6148396 |  |
| monthly-MICRO3 | 0.6110978 | 10.36 |
| yearly-MICRO1 | 0.5943620 |  |
| monthly-INDUSTRY2 | 0.5933019 |  |
| monthly-INDUSTRY3 | 0.5850340 |  |
| yearly-DEMOGRAPHIC1 | 0.5657703 |  |
| yearly-DEMOGRAPHIC2 | 0.5383269 |  |

At this point, I wanted to see whether there was a measurable statistical relationship between the spectral entropy of a time series, the distance of a time series from its nearest neighbors based on features, and the forecast accuracy of the protected series. For this, I calculated the average distance between each time series and its three nearest neighbors using the t-SNE features and regressed the protected forecast accuracy (absolute error) on this distance, the spectral entropy, their interaction, and the original absolute error of the series. This allows us to interpret how changes in spectral entropy and feature similarity would affect the protected forecast accuracy holding the original accuracy constant. So that we capture variation in forecast accuracy due to these quantities and *not* differences in model performance, I ran one regression for each model. Results are in Table 3 below.

***Table 3***: *coefficients from regression of the absolute error from forecasting a protected time series on the absolute error from forecasting the original version of that time series, the spectral entropy of the original time series, the average distance of the original time series to its three closest neighbors based on the two t-SNE variables, and the interaction of spectral entropy and average neighbor distance.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Regression Coefficients (\*=10% significant, \*\*=5% significance, \*\*\*=1% significance) | | | | |  |
| *Model* | *Intercept* | *Original Absolute Error* | *Spectral Entropy* | *Average Neighbor Distance* | *Interaction(Spectral Entropy, Avg Neighbor Distance)* | *, adjusted* |
| SES | 659.4\*\*\* | 0.942\*\*\* | -224.7\*\* | 0.0015\*\*\* | -0.0008 | 0.254, 0.253 |
| DES | 610.1\*\*\* | 0.719\*\*\* | 73.38 | 0.0018\*\*\* | -0.0013\* | 0.159, 0.157 |
| TES | 529.7\*\*\* | 0.841\*\*\* | -184.3\* | 0.0019\*\*\* | -0.0018\*\*\* | 0.269, 0.267 |
| ARIMA | 698.4\*\*\* | 0.863\*\*\* | -107.9 | 0.0022\*\*\* | -0.0021\*\*\* | 0.209, 0.207 |
| VAR | 641.1\*\*\* | -0.00030 | 955.4\*\*\* | 0.0036\*\*\* | -0.0035\*\*\* | 0.032, 0.030 |
| LGBM | 771.4\*\*\* | 1.12\*\*\* | -0.052\*\*\* | 0.0013\*\*\* | -0.00082 | 0.387, 0.386 |

The average distance between a series and its three closest neighbors consistently has a positive, statistically significant effect. Holding the original forecast error ad spectral entropy constant, increasing the distance between a series and its neighbors will increase forecast error. The amount that the forecast error increases when neighbor distance increases is dependent on the spectral entropy of the original series (although this is not statistically significant for all models). For example, when forecasting with TES, the absolute error of forecast from the protected data increases by for every one-unit increase in the average distance between a time series and its neighbors. This means for series that have very high spectral entropies (e.g., Monthly Micro), having less similar time series is not as damaging to forecast accuracy.

See the plots for monthly-MACRO3 and monthly-MICRO1 in Figure 1. The series in both data sets are relatively dissimilar. However, the percentage increase in forecast accuracy is several orders of magnitude smaller for the monthly-MICRO1 series. **Table 4** contains the average neighbor distance and average percentage increase in forecast accuracy across the series in each of these data sets. Even though the average neighbor distance is larger for the monthly-MICRO1 series, the percentage increase in forecast error is much less than for the Monthly-MACRO3 series.

|  |  |  |
| --- | --- | --- |
| Data Subset | Average Neighbor Distance | Average % Increase in Absolute Error |
| Monthly-MACRO3 | 243826 | 3704% |
| Monthly-MICRO1 | 626065 | 133% |