# Univariate Time Series Forecasting using ES-RNN and N-BEATS

**Yves Greatti**
New York University
yves.greatti@nyu.edu

**Zian Chen**
New York University
zian.chen@nyu.edu

**Sunjoo Park**
New York University
sunjoo.park@nyu.edu

## Abstract

In this work, we seek to replicate and improve the results reached by two neural networks: **ES-RNN** and **N-BEATS** on the M4 dataset competition. We also run different experimentations to compare the performances of these two deep learning techniques to a more classical statistical approach like Gaussian Process ($\mathcal{GP}$). We demonstrate that although Gaussian processes could be powerful for sampling tasks and simpler to configure, these neural networks outperform it for forecasting. Neural networks could have an overhead in term of the number of hyper-parameters to tune, but when using batching they scale up very easily and generalize well to a large number of time series (100K for M4). We are thus, pretty confident that, the two neural networks could forecast with the appropriate setting of hyper-parameters, other univariate time series beyond the M4 dataset.

## 1 Introduction

The dataset is provided by the M4 competition organized by the International Institute of Forecasters. These competitions also known as the "Makridakis competitions", have been happening since 1982, roughly every 10 years with an increasing number of time series to forecast starting from 1000 in 1982 to 100,000 in 2018, for the M4 competition. They attract people from academia as well practitioners, the last winner is Slawek Smyl, from Uber Technologies. The model used by Smyl is a hybrid approach combining Holts-Winter smoothing techniques with a recurrent neural network (RNN). Boris Oreshkin et al. want to challenge the conclusion that, mixed techniques are the future by proposing a pure DL model with interpretable outputs: **N-Beats**, which they claim outperforms **ES-RNN**, Smyl's model. N-Beats could be a very deep DL model and it is important for their authors that, in addition to hight accuracy, N-Beats has interpretable outputs which can match established statistical models such as ETS and ARIMA, robust, efficient, and automatic models.

## 2 Problem Description

---

- Given an history $[y_1, \cdots, y_T]$
- TASK: predict $[y_{T+1}, \cdots, y_{T+H}]$
- H: horizon, T: length of observations, m: periodicity of the data
- Standard scale-free metrics in the practice of forecasting:
  - sMAPE: symmetric Mean Absolute Percentage Error
  - MASE: Mean Absolute Scaled Error

---

For the purpose of this research, we use the predictive accuracy metrics sMAPE and MASE which are the standard metrics used across the forecasting community, disregarding OWA which is a metric defined by Makridakis and requires $\text{MASE}_{\text{Naive2}}$ as part of its computation, which is generated by a random walk model set up by the organizers of the competition.

$$\text{sMAPE} = \frac{2}{H}\sum_{i=1}^{H}\frac{|y_{T+i} - \hat{y}_{T+i}|}{|y_{T+i}| + |\hat{y}_{T+i}|}\,, \text{MASE} = \frac{1}{H}\sum_{i=1}^{H}\frac{|y_{T+i} - \hat{y}_{T+i}|}{\frac{1}{T+H-m}\sum_{j=m+1}^{T+H}|y_j - y_{j-m}|}$$

### 2.1 Data Description

M4 dataset extends the previous three competitions by increasing significantly the number of series. The time series are grouped into six categories and can be split into two groups: high-frequency data (weekly, daily, hourly) along with low-frequency data (yearly, quarterly and monthly). They were built from real, multiple, and diverse sources and are divided into six data types: demographic, finance, industry, macro, micro and other. The minimum of observations for the training set are 79 for hourly, 652 for daily, 67 for weekly, 24 for monthly, 8 for quarterly and 7 for yearly series. They have different length raging from 7 to about 1000 observations.

| **Frequency** | **Demographic** | **Finance** | **Industry** | **Macro** | **Micro** | **Other** | **Total** |
|---|---|---|---|---|---|---|---|
| Yearly | 1,088 | 6,519 | 3,716 | 3,903 | 6,538 | 1,236 | 23,000 |
| Quarterly | 1,858 | 5,305 | 4,637 | 5,315 | 6,020 | 865 | 24,000 |
| Monthly | 5,728 | 10,987 | 10,017 | 10,016 | 10,975 | 277 | 48,000 |
| Weekly | 24 | 164 | 6 | 41 | 112 | 12 | 359 |
| Daily | 10 | 1,559 | 422 | 127 | 1,476 | 633 | 4,227 |
| Hourly | 0 | 0 | 0 | 0 | 0 | 414 | 414 |
| Total | 8,708 | 24,534 | 18,798 | 19,402 | 25,1212 | 3,437 | 100,000 |

Table 1: M4 data by type and series frequency

| | **Mean** | **Std-Dev** | **Min** | **25%** | **50%** | **75%** | **Max** |
|---|---|---|---|---|---|---|---|
| Yearly | 25 | 24 | 7 | 14 | 23 | 34 | 829 |
| Quarterly | 84 | 51 | 8 | 54 | 80 | 107 | 858 |
| Monthly | 198 | 137 | 24 | 64 | 184 | 288 | 2776 |
| Weekly | 1009 | 707 | 67 | 366 | 921 | 1590 | 2584 |
| Daily | 2343 | 1756 | 79 | 309 | 2926 | 4183 | 9905 |
| Hourly | 805 | 127 | 652 | 652 | 912 | 912 | 912 |

Table 2: M4 series length statistics

## 3 ES-RNN

### 3.1 Model Description

ES-RNN algorithm consists in two major layers: a preprocessing layer which uses Holt-Winter smoothing technic to extract trend and seasonality parameters and a dilated RNN network. The Holt-Winter parameters are part of the back-propagation and are tuned as the network learns the characteristics of the time series. The ES-RNN model uses a modified version of the Holt-Winters formula, in which there is no local linear level coefficient, and level and seasonality terms are scaled instead of subtracted. The linear forecast is replaced by an RNN, which has for input the normalized and de-seasonalized prior observations.

$$l_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha) l_{t-1}$$
$$s_t = \beta \frac{y_t}{l_{t-1}} + (1-\beta) s_{t-m}$$
$$\hat{y}_{\textbf{win}} = \textbf{ES-RNN}(\frac{y_{ti}}{s_{ti} l_{ti}})$$
$$y_{\textbf{truth}} = (\frac{y_{to}}{s_{to} l_{to}})$$

where $l$ is a state variable, $s$ is a seasonality coefficient, $\alpha$ and $\beta$ are network parameters, $m$ is the periodicity of the data.

To model long-term dependencies in time series, ES-RNN uses dilated RNNs which allow to stack RNN cells by having skip connections between RNN cells . The dilation parameters are given in table Table **??**. DRNN cells can be vanilla RNN, LSTM, GRU, GRU cells provides the higher accuracy.

| **Time Series** | **Dilations** |
|---|---|
| Quarterly | (1, 2), (4, 8) |
| Monthly | (1, 3), (6, 12) |
| Daily | (1, 7), (14, 28) |
| Yearly | (1, 2), (2, 6) |
| Weekly | (1, 14), (14, 28) |
| Hourly | (1, 24), (24, 48) |

Table 3: ES-RNN dilation parameters

### 3.2 Data Preparation

The M4 time series have different length, to allow vectorization and to take advantage of the GPUs, they are chopped using a predetermined cut off value. To the exception of the daily time series, for which the cut-off value is 200, 72 is used as the default value (Table **??**).

| **Filtering rates (in %)** | | | | | |
|---|---|---|---|---|---|
| **Hourly** | **Daily** | **Weekly** | **Monthly** | **Quarterly** | **Yearly** |
| 6.93 | 6.93 | 18.11 | 26.83 | 45.42 | 60.61 |

Table 4: Percentage of time series eliminated by the cut-off value

The batch size is by default 1024, we observe that, it affects the overall performance of the network, depending the time series frequency, too small the validation loss may jump up and down, and in some cases the larger the batch size, the better.

For an hourly frequency, the effect of the batch size is given in table Table **??** .

| | **sMAPE per category** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Batch size** | **Demographic** | **Finance** | **Industry** | **Macro** | **Micro** | **Other** | **Total** |
| 512 | 6.32 | 3.34 | 3.95 | 2.52 | 2.43 | 3.08 | 3.02 |
| 1024 | 6.20 | 3.27 | 3.9 | 2.52 | 2.37 | 3.07 | 2.97 |
| 2048 | 6.26 | 3.32 | 3.92 | 2.49 | 2.42 | 3.1 | 3.01 |

Table 5: ES-RNN batch size statistics

Each time series is split into two windows: a backcast and forecast window. Observations in each window are normalized and de-seasonalized using the coefficients of the exponential smoothing. In addition, a one-hot representation of the time series category is added to the input window.

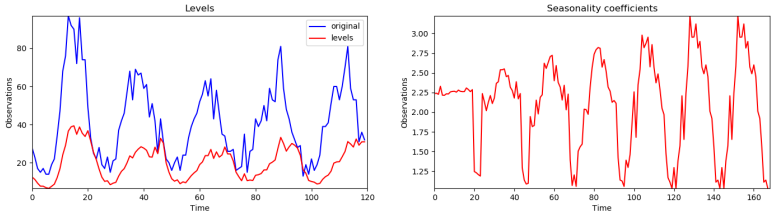


Figure 1: Levels are a smoothing version of the original time series and seasonality coefficients are between 0 and 1.

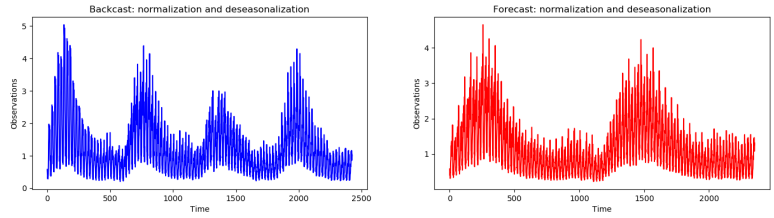


Figure 2: The input and ouput of the RNN are normalized and de-seasonalized, compared to N-BEATS which are not scaled.

Dropout or batch normalization do not provide any significant improvement, the smoothing and normalization of the data seem enough for a convergence to a local minimum. In both networks, using the pinball loss function brings definitely a boost in performance compared to L2 or L1 losses. The pinball loss function is popular in forecasting. Two different slopes are used depending whether the forecast $\hat{y}_t$ value is greater or smaller than the truth value $y_t$. Two models are then trained corresponding to the lower and upper bounds of the desired level of prediction interval:

$$L_\alpha = \begin{cases} \alpha(\hat{y}_t - y_t) & \text{if } \hat{y}_t \geq y_t \\ (1-\alpha)(y_t - \hat{y}_t) & \text{if } y_t > \hat{y}_t \end{cases}$$

Then $\mathrm{P}(\hat{y}_{t,lower} \leq y_t < \hat{y}_{t,upper}) = 1 - 2\alpha$.

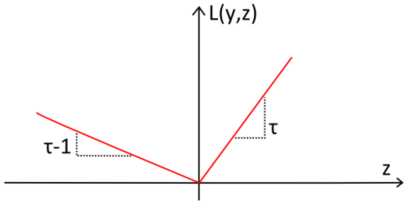


Figure 3: $\tau$ controls the desired imbalance in the quantile forecast.

## 4 General formatting instructions

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points. Times New Roman is the preferred typeface throughout, and will be selected for you by default. Paragraphs are separated by ½ line space (5.5 points), with no indentation.

The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow ¼ inch

space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the page.

For the final version, authors' names are set in boldface, and each name is centered above the corresponding address. The lead author's name is to be listed first (left-most), and the co-authors' names (if different address) are set to follow. If there is only one co-author, list both author and co-author side by side.

Please pay special attention to the instructions in Section 6 regarding figures, tables, acknowledgments, and references.

# 5 Headings: first level

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

## 5.1 Headings: second level

Second-level headings should be in 10-point type.

### 5.1.1 Headings: third level

Third-level headings should be in 10-point type.

**Paragraphs** There is also a `\paragraph` command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

# 6 Citations, figures, tables, references

These instructions apply to everyone.

## 6.1 Citations within the text

The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

The documentation for `natbib` may be found at

`http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

Hasselmo, et al. (1995) investigated…

If you wish to load the `natbib` package with options, you may add the following before loading the `neurips_2019` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{neurips_2019}
```

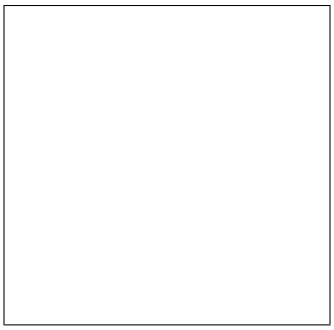

Figure 4: Sample figure caption.

As submission is double blind, refer to your own published work in the third person. That is, use "In the previous work of Jones et al. [4]," not "In our previous work [4]." If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form "A. Anonymous."

### 6.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

Note that footnotes are properly typeset *after* punctuation marks.[2]

### 6.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively.

You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.

### 6.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table 6.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

`https://www.ctan.org/pkg/booktabs`

This package was used to typeset Table 6.

## 7 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

[1] Sample of the first footnote.

[2] As in this example.

Table 6: Sample table title

| | Part | |
|---|---|---|
| Name | Description | Size ($\mu$m) |
| Dendrite | Input terminal | $\sim$100 |
| Axon | Output terminal | $\sim$10 |
| Soma | Cell body | up to $10^6$ |

# 8 Preparing PDF files

Please prepare submission files with paper size "US Letter," and not, for example, "A4."

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NeurIPS. Please see `http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf`
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

  ```
  \usepackage{amsfonts}
  ```

  followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for $\mathbb{R}$, $\mathbb{N}$ or $\mathbb{C}$. You can also use the following workaround for reals, natural and complex:

  ```
  \newcommand{\RR}{I\!\!R} %real numbers
  \newcommand{\Nat}{I\!\!N} %natural numbers
  \newcommand{\CC}{I\!\!\!\!C} %complex numbers
  ```

  Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

## 8.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (`http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf`)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

### Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

## References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Remember that you can use more than eight pages as long as the additional pages contain *only* cited references.**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.