Department of Conservation
Te Papa Atawhai

Adaptive weather adjusted visitations trend reveals a more realistic popularity of New Zealand tracks

an AI approach

Dong Wang, Jeff Dalley, Elaine Wright

2021-06-28

Abstract

This article reports how weather effects were reduced or even removed on visitations for New Zealand's walks, employing an AI method combined with time series theory. This includes:

1: a research statement
2: a gentle introduction to Neural Networks models and time series decomposition method
3: analyses: modelling each temporal component separately using Neural Networks
4: case studies: Tongariro Xing Mangatepopo, 100055887
5: Recommendations and future works

Keywords: visitor counts, sensor data, weather effects, neural networks models, time series models

Contents

1 Introduction and preview of the tools

NZ is the home of hundreds of tracks geographically distributed across New Zealand. It is known that those track locations are impacted by heterogeneous weather patterns, with various degrees, which unavoidably influence Visitors' decisions to visit. What complicated visitor' decisions also include weekends or/and public holiday effects, i.e., visitors more likely to go during (long) weekends or/and holidays.

In our last pioneer study, an AI approaches has been adopted that have teased apart weather effects from weekend and holiday effects on visitor numbers and relative impacts from each impact on any tracks has been quantified and ranked.

In this study however, we focus on building a novel AI-powered tool to quantify weather adjusted interannual visitation trend and compare with its unadjusted counterpart in New Zealand tracks.

1.1 Why weather adjusted visitation trend is important for DOC

D & E team in DOC publish visitation series weekly based on the observed 'raw' visitor counts from counters. However, such raw visitor counts does not necessarily and always reflect the true demands for tracks from the visitors and does not show the underlying trend. This is simply because many factors like weather events especially extreme ones or/and unusual warm or cold seasons, can often obscure such true demands and the underlying trend. For examples, if we drill down to daily level, Fig.1 shows the site Tongariro Xing Mangatepopo Tk, renowned to its high wind speed, the visitor counts almost dropped to none on 2012-03-20, while on the same day and month but different years, like year 2010, when wind speed is in the 'normal' range (2-5 m/s), visitor counts can be over 1000. Therefore, it is hard to make management and investment decisions if one only based on the 'raw' visitor counts, if one looks at visitor counts for year on year change. The question is what the true demands and the underlying trend would be had unusual weather events and seasons were not there? The answer is to compute the weather adjusted demands or visitation trend.

date	count	RAIN	TMIN	RAD	RH	WIND
3/20/2009	220	0	4.8	19.5	78	4.2
3/20/2010	1153	0	7.4	14.9	91	4.2
3/20/2011	413	0	1.5	18.8	85	2.1
3/20/2012	5	0	7.0	14.0	83	13.0
3/20/2013	915	0	3.8	18.7	75	4.1
3/20/2014	211	0	1.5	19.3	87	3.3
3/20/2015	364	0	-5.5	21.7	66	2.9
3/20/2016	445	0	5.4	14.2	82	3.2
3/20/2017	614	0	3.1	16.1	86	3.9
3/20/2018	1063	0	3.3	16.2	70	2.6
3/20/2019	988	0	7.1	17.7	93	2.2

Note:

Here is a general comments of the table.

1 count: observed visit count;

2 RAIN (mm): rain amount;

3 TMIN (°C): minimal temerature;

4 RAD (W/m²): solar radiation;

5 RH (%rh): relative humidity;

6 WIND (m/s): wind speed;

Figure 1: Daily visitor counts and weather parameters on March 20 since 2009 for Tongariro Xing Mangatepopo Tk, 100055887

2 Methodology

Broadly speaking, traditional statistical science have three approaches to estimate the trend component of a time series.

The simplistic approach falls into linear regression realm where ordinary least-squares method and its variants are used to estimate a slope and intercept that describe the trend component. Such approach and its variants are relatively easy to use and ubiquitous in the literature. However, most of series in the real life can violate one or more (strong) assumptions that underpin linear regression (Montgomery, Peck, and Vining 2006), for example, by examining our visitation series, we often see noise component is not bell-shaped, assuming series is represented by additive trend and noise components with a time dependent scaling function.

There are also plethora methods in nonlinear approach, e.g., based on ARIMA (Storch and Cambridge 1999). Also, some researchers consider trend component are basically in a certain nonlinear form and use various numeric optimization algorithms, (e.g., weighted least-squares) to estimate the trend (Michalewicz and Fogel 2004).

Nonparametric approach is a drastically different one where the trend is estimated by a form of kennels function where researches have to carefully justify the selection of the associated kennel parameters, e.g., bandwidth in Epanechnikov kernel, and apply it to smooth the series (Gasser and Müller 2006).

Each apporach has its merit, however, the problem is that most of methodologies derived from those approaches are only suitable for estimating trend for the univariate time series, which applicable only if it have its own relatively isolated and independent data generating process. It therefore would be unrealistic to directly apply those methodologies to our visitation series, and ignore all 'co-evolved' time series from weather parameters and other latent variables, which have impact on it.

2.1 Our Approach

Overall, our approach is to integrate Artificial neural networks (ANNs) with classic time series decomposition theory, to compute weather adjusted trend for visitation series.

2.1.1 ANNs

We are living in the world that increasingly rely on products and services featured some Artificial Intelligence (AI). AI, especially those based on Artificial neural networks (ANNs), are rapidly becoming essential and dominant for analysis of complex data and for decision support.

ANNs are highly parameterized, non-linear models with sets of interconnected processing units called neurons that can be used to approximate the relationship between input and output signals of a complex system (Stefaniak, Cholewiński, and Tarkowska 2006). Typically, ANNs are applied to predict the response of one or more variables given one to many explanatory variables, where smooth functions are fitted to dataset while residual error are minimized through iterative training (Hornik 1991)

Compared to conventional statistical models where traditional statisticians used (e.g., generalized linear/additive regression), ANNs have been proved to have a more powerful (probably unmatched) predicting capability.

2.1.2 Time series decomposition

It is possible that a time series y_t can be decomposed in three different temporal components in an additive form, Equation (1): a trend component T_t , a seasonal component S_t and a reminder pattern R_t, all at time t (Hamilton 1994; West 1997). Here, we decomposed on the daily data for VC and all weather parameters in consideration for modelling as per Section 2.2.

$$y_t = T_t + S_t + R_t \tag{1}$$

T_t is the low frequency variation in data together with nonstationary, long-term change in level. S_t is the variation in the data at or near the seasonal frequency. R_t is the remaining variation in data beyond that in T_t and S_t.

2.2 ANNs Model and its Specifications

The input is a set of predictors: four weather parameters (daily wind speed (WS), daily minimal temperature (T), daily rain amount (Rain), daily solar radiation (SR)) and a various number of lagged weather parameters. The output is daily visitor count (VC).

The primary reason to consider lagged structure in model specifications is weather parameters are usually autocorrelated. Both over and under -specification of lags were known to have impact on the models response, i.e., VC. In this study, we followed Wang and Lu (2006) and Lu and Wang (2014) to identify the optimal lags. For a more systematic, generic and rigorous treatment on the effects of past history of predictors, we recommend the work from Jørgensen (2004).

2.3 Full-season models to establish daily relationship between VC and its influentials

Contrast with our previous weather sensitivity study, where only peak season data were used, in this study, daily data of both VC and weather parameters from full seasons were used in modelling, as we aims at comparing year-on-year change of both adjusted and unadjusted visitation series.

2.4 ANNs configuration

Number of layers of ANNs can affect models performance. We tested a range of hidden layers and found single hidden layer mostly sufficed our purpose. Increased layers, in some cases, did increase 'prediction' accuracy, but such gains are little and negligible. Root mean squared errors (RMSE) were used as the metric in model selections. For a through consideration of ANNs and its applications, we recommend the works from Subana Shanmuganathan (2016).

2.5 ANNs on three temporal components and their residuals as the adjusted weather trend

We tested the configured ANNs on its ability to establish the relationship between VC and those six weather parameters on one popular New Zealand walk: Tongariro Alpine Crossing, as promising results were obtained in our last weather sensitivity study (on peak season data): Tongariro Alpine Crossing are almost twice as much as affected by weather (summed by all 4 weather parameters) compared to Dolomite Point, which likely reflected the fact and experience as per the subject matter expert from tourism industry.

Here, we firstly tested the same configured and specifications ANNs on the full season data on y_t, then progressively on T_t, S_t, R_t, which were obtained from Section 2.1.2. Lastly, we extracted residuals for ANNs on each temporal components which then were summed up as the adjusted daily weather trend, denoted as $trend\_adj\_nonLinear$. We compared year-on-year change of both $trend\_adj\_nonLinear$ and unadjusted visitation trend, which is T_t but denoted as $trend\_unadj\_nonLinear$ for clarity.

2.6 Source of weather data and simulation software

All daily weather data were estimated on a regular (~5km) grid covering the whole of New Zealand, i.e., VCSN data simulated by NIWA. The estimates are produced every day, based on the spatial interpolation of actual data observations made at climate stations located around the country. A thin-plate smoothing spline model is used for the spatial interpolations. This model incorporates two location variables (latitude and longitude) and a third "pattern" variable (Tait, Sturman, and Clark 2012). The software used for the interpolations is ANUSPLIN (McKenney et al. 2011). We used the VCSN data in the grid that is the closest to the walk(s) in modelling.

3 Results

3.1 Temporal components from decomposition and modelled separately

Each series from VC and weather parameters were decomposed as per Equation (1). Here, only a portion of results from wind speed series were shown as Fig.2. By doing so:

1. the long term (LT) trends shown in red, here is T_t for wind speed as an example, were revealed as the unajusted trend, denoted also as $trend\_unadj\_nonLinear$. The same were applicable to VC series where we got $trend\_unadj\_nonLinear$ for VC.
2. irregularities or R_t as extreme high wind speed as an example here, could have less impact on corresponding visitor counts' R_t, as each temporal component were modelled separately by ANNs.
3. cyclic signal are S_t component and were also modelled separately by ANNs.

3.2 Summed residuals from ANNs for each temporal components as the weather adjusted trend

ANNs models residuals were extracted and summed from trend component T_t, seasonal component S_t and a reminder pattern R_t, which is the weather adjusted trend. Fig.3 and 4 compared weather adjusted with adjusted trend on the daily and yearly level.

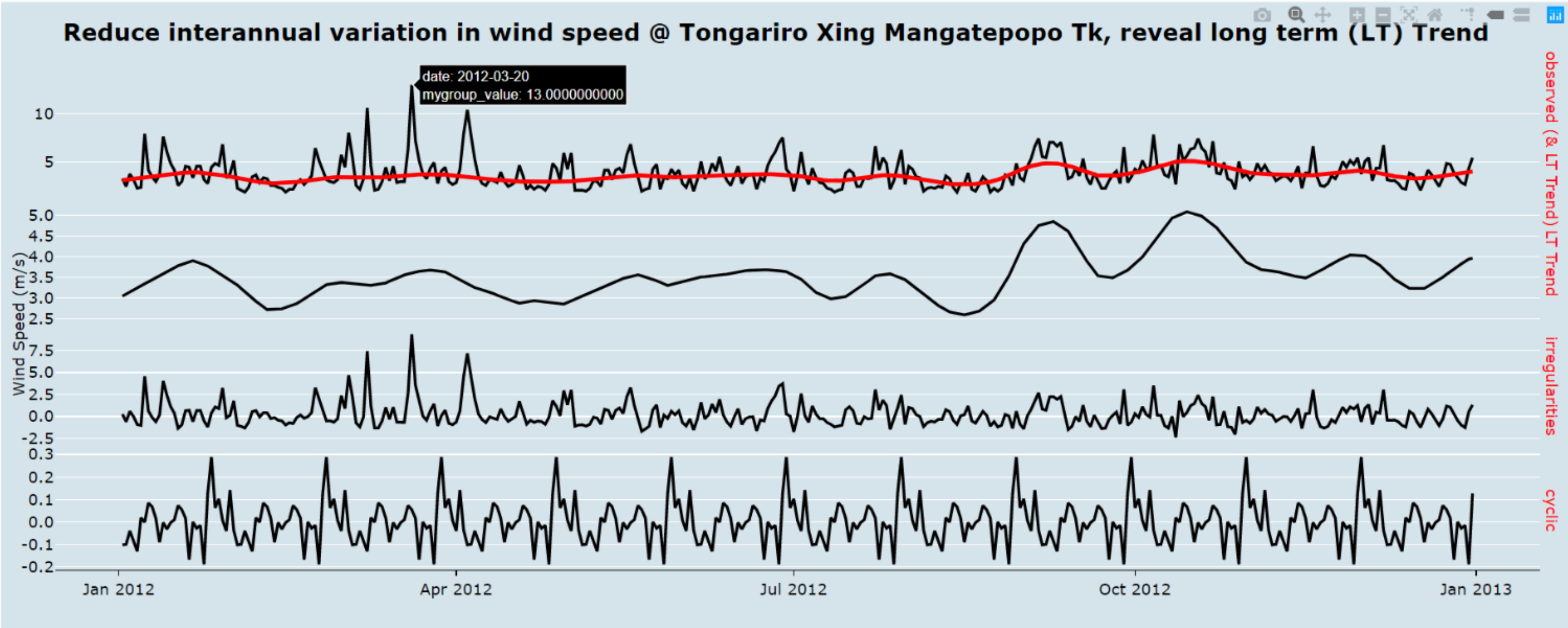

Figure 2: Wind speed on Tongariro Xing Mangatepopo Tk (Tongariro Alpine Crossing) decomposed

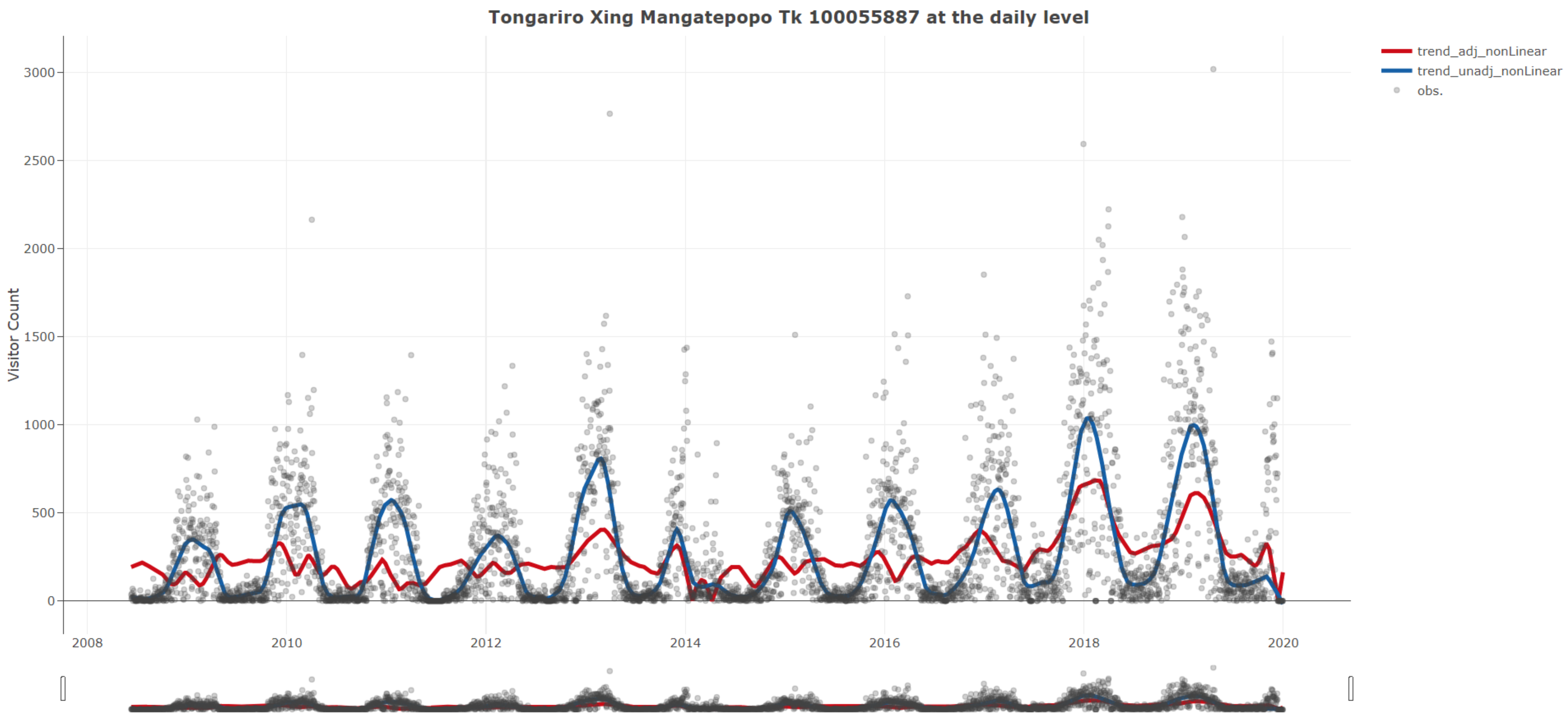

Figure 3: Comparision of weather adjusted and unadjusted VC on Tongariro Xing Mangatepopo Tk (Tongariro Alpine Crossing) at the daily level

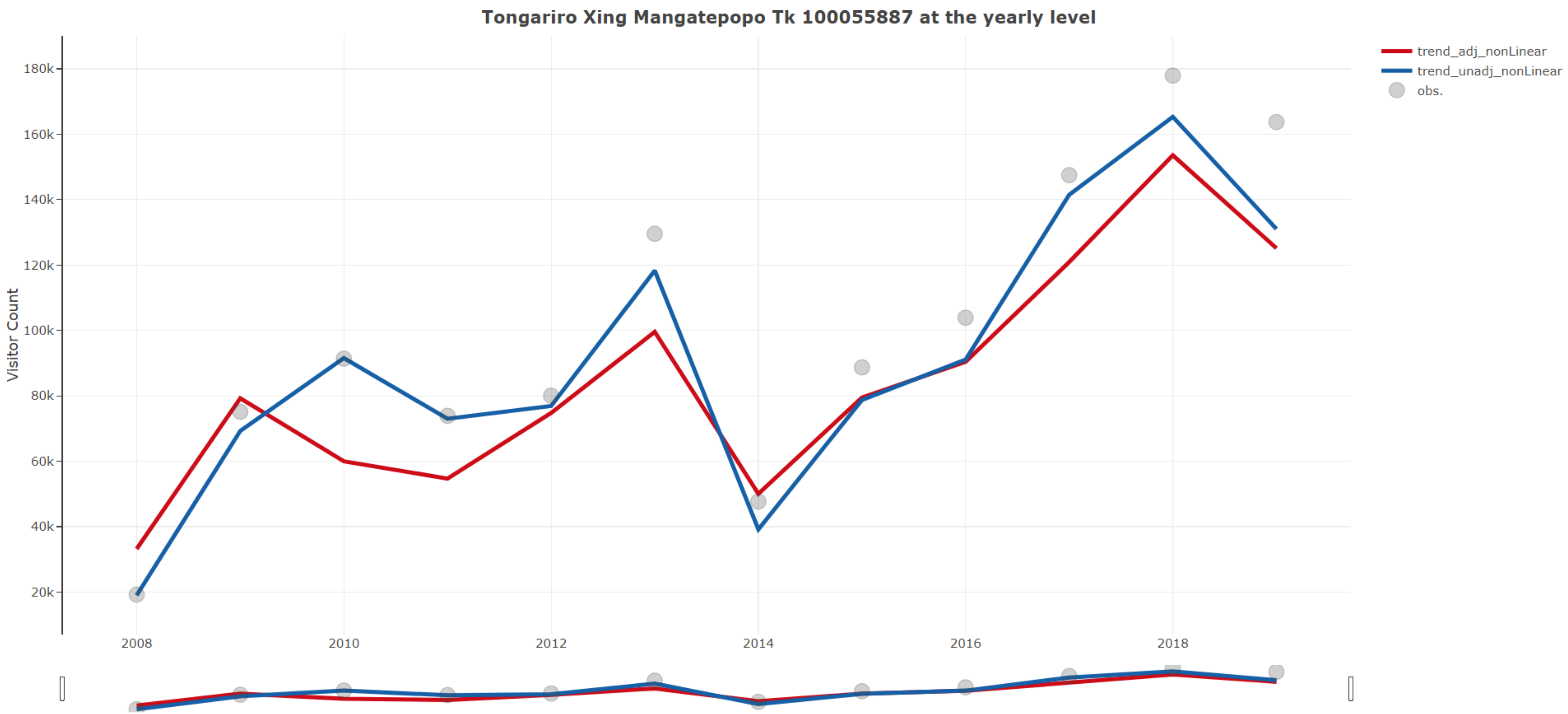

Figure 4: Comparision of weather adjusted and unadjusted VC on Tongariro Xing Mangatepopo Tk (Tongariro Alpine Crossing) at the yearly level

4 Discussions

In Fig.3, in general, *trend_adj_nonLinear* has smaller peak in summer but smaller depth of trough in winter in terms of on visitation counts than *trend_unadj_nonLinear*. This suggests weather adjustment worked: had the extreme weather events or unusual warm or cold seasons' effect reduced or even removed, we can get the true demands of visitation which should be 'smoother' than visitation that were not weather adjusted. In addition, both *trend_adj_nonLinear* and *trend_unadj_nonLinear* curves still reflected the seasonal pattern of original visitation counts.

What's more interesting is in Fig.4, as we know globally as well as New Zealand, there is a general trend of global warming, where weather indicators like TMIN over the studied years are on the rise. This means there were more and more better days in favorite of tramping and walking, which is reflected and aggregated in the *obs*. However, *obs* is not necessarily true demands of visitation or may not a genuine indicator of the popularity of a walking site: had the global warming were not happening, the site may not be as popular as what it suggested by *obs*, but the more realistic popularity may be milder shown by the curve of *trend_adj_nonLinear*.

5 Recommendations and future works

It will be interesting to see the current approach being applied to other New Zealand walks. This study also took a none-or-all approach to address weather adjustment. It is worth to consider a more granular adjustment where only the most influential weather parameters on a site is to be adjusted.

Reference

Gasser, Theo, and Hans-Georg Müller. 2006. "Kernel Estimation of Regression Functions." In *Lecture Notes in Mathematics*, 757:23–68. https://doi.org/10.1007/BFb0098489.

Hamilton, James D. 1994. *Time Series Analysis*. 1st ed. Hardcover; Princeton University Press. http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20/&path=ASIN/0691042896.

Hornik, Kurt. 1991. "Hornik, k.: Approximation Capabilities of Multilayer Feedforward Network. Neural Networks, 251-257." *Neural Networks* 4 (January). https://doi.org/10.1016/0893-6080(91)90009-T.

Jørgensen, S. E. 2004. "Model Selection and Multimodel Inference: A Practical Information—Theoretic Approach, Second Edition, Kenneth p. Brunham, David r. Anderson, Springer-Verlag, Heidelberg, 2002, 490 Pages, Hardbound, 31 Illustrations." *Ecological Modelling* 172 (February): 96–97. https://doi.org/10.1016/j.ecolmodel.2003.11.004.

Lu, Wei-Zhen, and Dong Wang. 2014. "Learning Machines: Rationale and Application in Ground-Level Ozone Prediction." *Applied Soft Computing* 24 (November): 135–41. https://doi.org/10.1016/j.asoc.2014.07.008.

McKenney, Daniel, M. F. Hutchinson, Pia Papadopol, Kevin Lawrence, John Pedlar, Kathy Campbell, Ewa Milewska, Ron Hopkinson, David Price, and Tim Owen. 2011. "Customized Spatial Climate Models for North America." *Bulletin of the American Meteorological Society* 92 (December): 1611–22. https://doi.org/10.1175/2011BAMS3132.1.

Michalewicz, Zbigniew, and David Fogel. 2004. *How to Solve It: Modern Heuristics*. https://doi.org/10.1007/978-3-662-07807-5.

Montgomery, Douglas C., Elizabeth A. Peck, and Geoffrey G. Vining. 2006. *Introduction to Linear Regression Analysis (4th Ed.)*. Wiley & Sons.

Stefaniak, Bogusław, Witold Cholewiński, and Anna Tarkowska. 2006. "Algorithms of Artificial Neural Networks - Practical Application in Medical Science." *Polski Merkuriusz Lekarski : Organ Polskiego Towarzystwa Lekarskiego* 19 (January): 819–22.

Storch, Hans Von, and Francis W. Zwiers Cambridge. 1999. "Statistical Analysis in Climate Research."

Subana Shanmuganathan, Sandhya Samarasinghe. 2016. *Artificial Neural Network Modelling*. 1st ed. Boca Raton, Florida: Springer International Publishing. https://www.springer.com/gp/book/9783319284934.

Tait, A., J. Sturman, and M. Clark. 2012. "An Assessment of the Accuracy of Interpolated Daily Rainfall for New Zealand." *Journal of Hydrology New Zealand* 51 (January): 25–44.

Wang, Dong, and Wei-Zhen Lu. 2006. "Ground-Level Ozone Prediction Using Multilayer Perceptron Trained with an Innovative Hybrid Approach." *Ecological Modelling* 198 (October): 332–40. https://doi.org/10.1016/j.ecolmodel.2006.05.031.

West, M. 1997. "Time Series Decomposition." *Biometrika* 84: 489–94. http://www.stat.duke.edu/%20mw/MWextrapubs/West1997c.pdf.