

How much weather effects is on NZ tracks visitations

an AI approach to divide and conquer

Dong Wang, Jeff Dalley, Elaine Wright

2020-09-18

Abstract

This article reports how weather effects were singled out and quantified on visitations for New Zealand's walks, employing an AI method. This includes:

- 1: a research statement
- 2: a gentle introduction to Neural Networks models and model specifications
- 3: sensitivity analyses: teasing apart weather effects from both holiday and weekend effects
- 4: case studies: weather impact on sites with close geo-locations, an indirect way to validate complex models
- 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	2
2	Methodology	3
2.1	ANNs Model Specifications	3
2.2	Peak-season models to establish daily relationship between VC and its influentials	4
2.3	ANNs configuration	4
2.4	Initial models testing	4
2.5	Sites selection to further build a picture of weather impact across New Zealand	4
2.6	Source of weather data and simulation software	4
2.7	Simulations for weather sensitivity analysis after ANNs modelling	4
3	Case studies: weather impact on sites with close geo-locations	8
4	Results and Discussions	11
4.1	Classification of weather impact across NZ	11
5	Acknowledgement	11
	Reference	11

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 this pioneer study, an AI approaches has been adopted that can tease apart weather impacts from weekend and holiday effects on visitor numbers on any tracks.

The output (Fig.1 and Fig.2) of this study can help infrastructure planners, pricing teams, managers and directors to make sound management decisions.

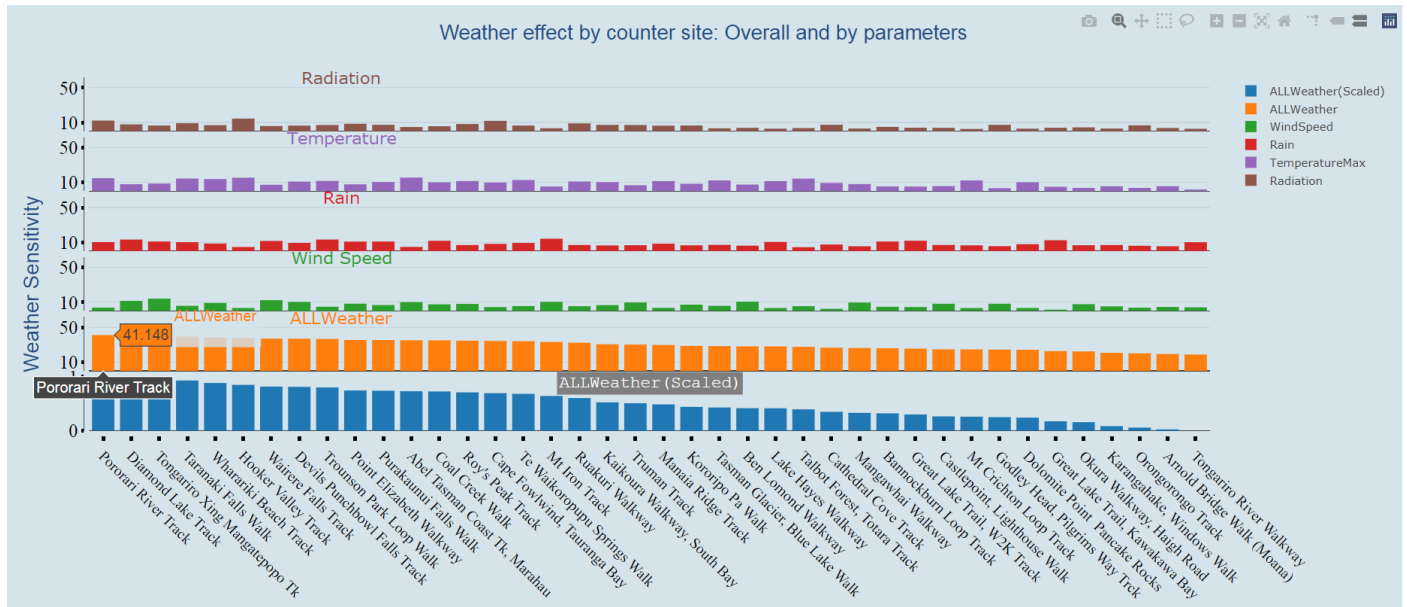


Figure 1: Weather effect on its original and standardised ([0,1]) scale

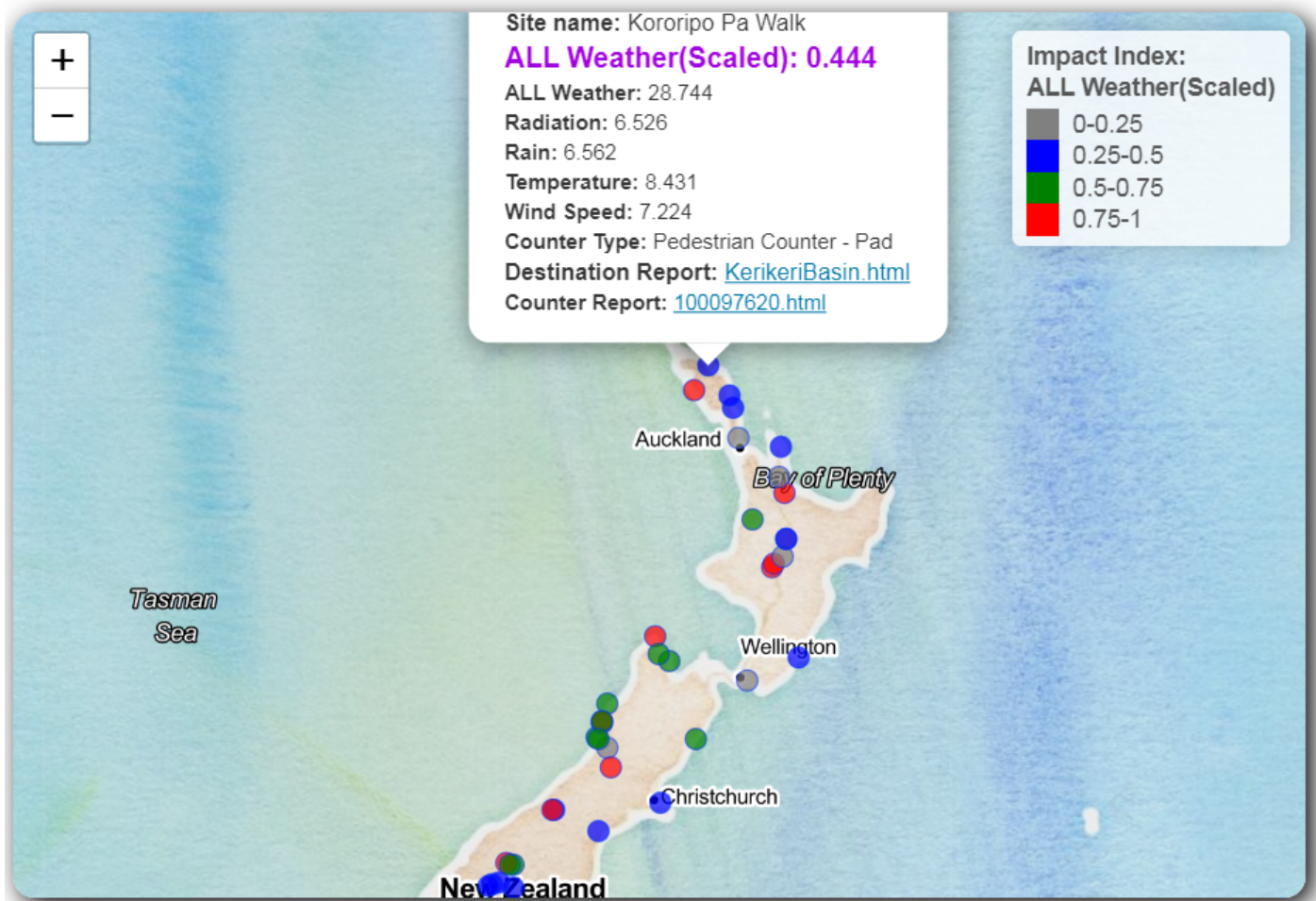


Figure 2: Weather effect on its original and standardised $([0,1])$ scale

2 Methodology

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. Leverageing ANNs such advantages, in this study, we are interested to use it to 1) predict visitor counts (VC), 2) quantify and explain what predictor(s) are most influential among a range of confounding predictors/factors on a specific track. Based on this, We will derive a generic weather index for each track that single out the effect from the weather predictors.

2.1 ANNs Model 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)), daily holiday indicator (flagged as 0 or 1), weekend indicator (flagged as 0 or 1) and a various number of lagged weather parameters. The output is daily 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 (2006b) 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).

2.2 Peak-season models to establish daily relationship between VC and its influentials

Daily data of both VC and weather parameters from months only in peak season: Oct., Nov., Dec., Jan., Feb., Mar. and Apr. were used in modelling. Data in peak season months from all available years were pooled and used. All available years here refer to years available in VC data. This is suppose to reduce the ‘noise’ when modelling: only months in peak seasons, visitors are abundance.

2.3 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).

2.4 Initial models testing

Initially, we tested configured ANNs on its ability to rank relative importance of inputs predictors on two selected New Zealand Walks: Tongariro Alpine Crossing and Dolomite Point. Encouraging results were obtained: 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.

2.5 Sites selection to further build a picture of weather impact across New Zealand

DOC owns 700+ tracks/walks across New Zealand. VC data on each walks has variant qualities, which pose a risk that could compromise ANNs modelling. Based on known ANNs performance and quality of VC data on two walks in Section 2.4, three criteria were considered when candidates walks were picked for ANNs modelling:

1. low % of hourly VC data missing ($\leq 10\%$)
2. high % of active hours: hours with VC > 0 ($\geq 40\%$)
3. ideally high number of years with data availability (≥ 2)

This results in 40+ candidates sites picked for modelling.

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.

2.7 Simulations for weather sensitivity analysis after ANNs modelling

ANNs are not statistical models. Once generated, ANNs are not statistically identifiable (deterministic) at all. For a given dataset and configuration, there can be unlimited numbers of ANNs with different weights that could generate very different predictions. Usually, the implicit relationship built by ANNs between the input and outputs are difficult to be interpreted directly, let alone ranking predictors relative importance or assessing models’s sensitivity. As such, a (always) common criticism is ANNs are notoriously known as ‘black box models’ that offer minimal insight into relationships between inputs and outputs variables. Wang and Lu (2006a) and among a few of others provided a rebuttal to this concern by describing methods to extract information about such variable relationships from ANNs. Here, based on solutions especially in Wang and Lu (2006a), we provided another version of methods to explore the relationship between the outcome variable, VC, and predictor(s) of interest.

Our methods, in a nutshell, is to explore the relationship of VC and a predictor of interest, say, WS, while holding other covariates at constant values (from 5th, 25th, 50th, 75th to 95th quartile), after ANNs being built.

As an example for illustration, let’s say if we want to know the effect of WS on VC in Tongariro Alpine Crossing on Holiday and Weekend scenario.

Steps

1. calculate ith (e.g., 5th) quartile for other weather covariates: T, SR and Rain on the daily data for Holiday and Weekend
2. plug ith quartile for other weather covariates data into ANNs

3. while holding 5th quartile for other weather covariates as constants, vary WS values from 0 to its known daily maximum, record the VC outputs from ANNs
4. run a linear regression model between VC outputs and WS, record and graph such linear regression model: slope, intercept, p-value and R-squared
5. the slope: -51.7 (and the intercept: 734) are WS effect on VC in this scenario, shown in Fig.3a

In this way, for a particular weather parameter, (WS here), we examined its effect on 20 possible scenarios: 5 (quartiles of other weather covariates) \times 2 (weekend or not) \times 2 (holiday or not) in Fig.3. Slopes and intercepts were then standardized which allow each slope in regression lines (i.e., weather effects) are comparable. Therefore, WS effect WS on VC in Tongariro Alpine Crossing on Holiday and Weekend scenario is -0.78, which was the standardized slope.

The overall effect for the particular weather parameter, (WS here) was a sum of the standardized slopes (after taking absolute values) of all 20 scenarios. For Tongariro Alpine Crossing, its effect was 14.157 as shown in Fig.1, 3 and 4.

Following the same steps and procedures above, we obtained other 3 weather parameters effects on Tongariro Alpine Crossing as shown in Fig.1 and 4. This allows us to quantify effects for each weather parameter separately and rank the relative importance of each weather parameters. The overall weather effects on any particular walks then is a simple sum of the effects from each separate weather parametrs. For Tongariro Alpine Crossing, it is 40.321.

We then processed each weather effects and its overalls for other eligible candidates sites selected as per Section 2.5 and shown in Fig.1.

The overalls of weather effects then were scaled in the range of [0,1], which allows us to rank the weather impact site-wise, as shown in Fig.1 and Fig.2

Overall WindSpeed Effect @ Tongariro Xing Mangatepopo Tk: 14.157

Effect (E): the mean change of daily visitor activity in the response given a one unit change in the WindSpeed
 Circles are simulations using Neural Networks, where Rain, Temperature/Max, Radiation, are fixed at a % Quantile
 Linear regression lines are fitted to circles. R-squared value means how good circles are being fitted by regression line: p-value from F test to test if full regression is significant

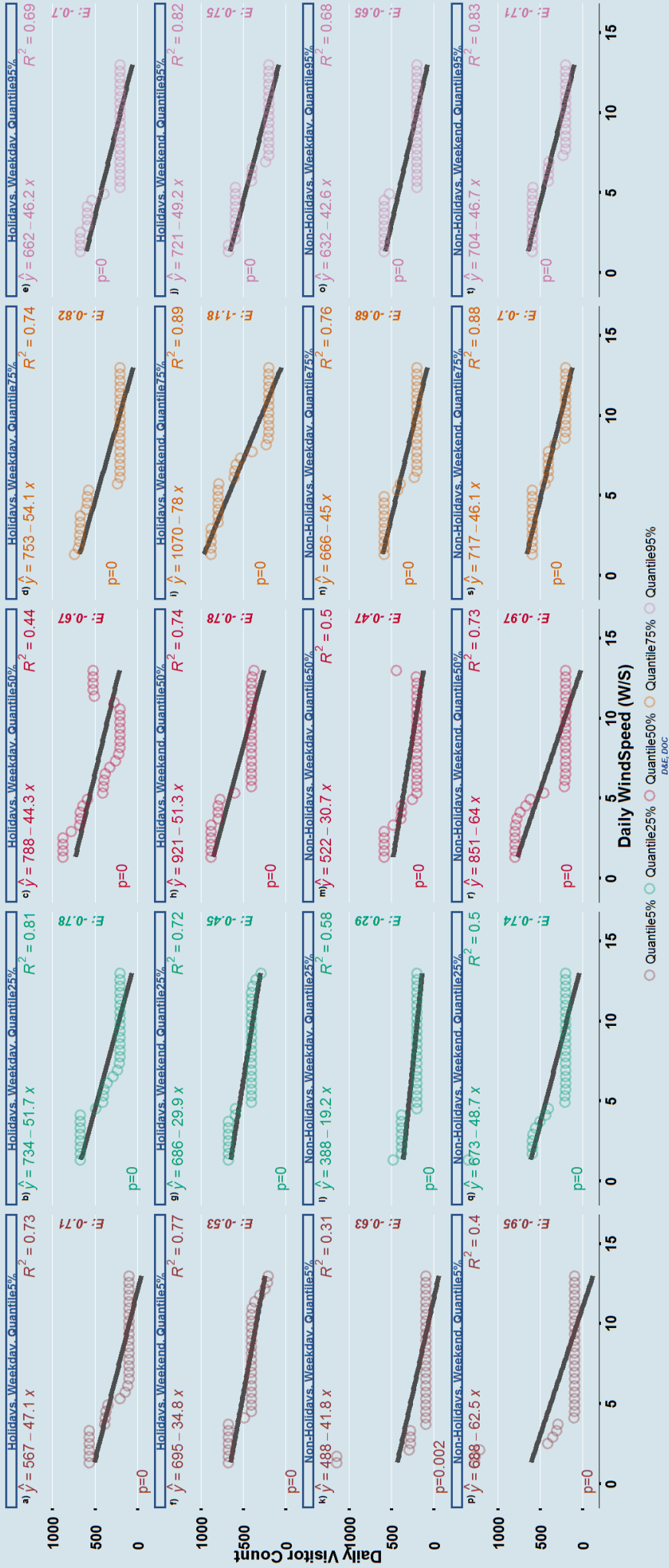
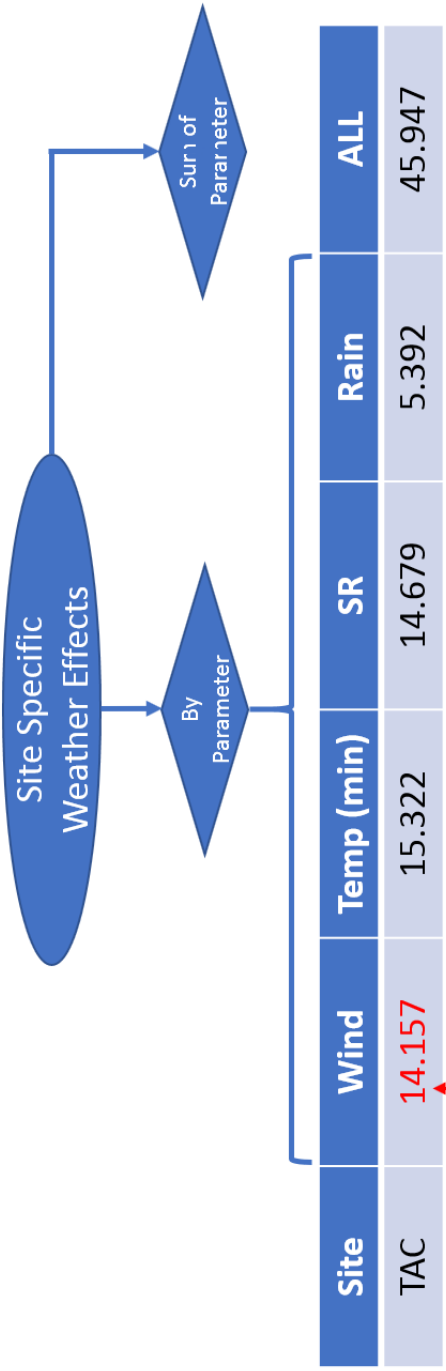


Figure 3: Wind Effects on Tongariro Xing Mangatepopo Tk (Tongariro Alpine Crossing) under 20 scenarios and its overall effects

Flowchart: to illustrate how a specific weather effect being calculated, Tongariro Alpine Crossing (TAC)



$$TAC_{wind} = |-0.71| + |-0.78| + \dots = 14.157$$

Sum of each effects (absolute values)

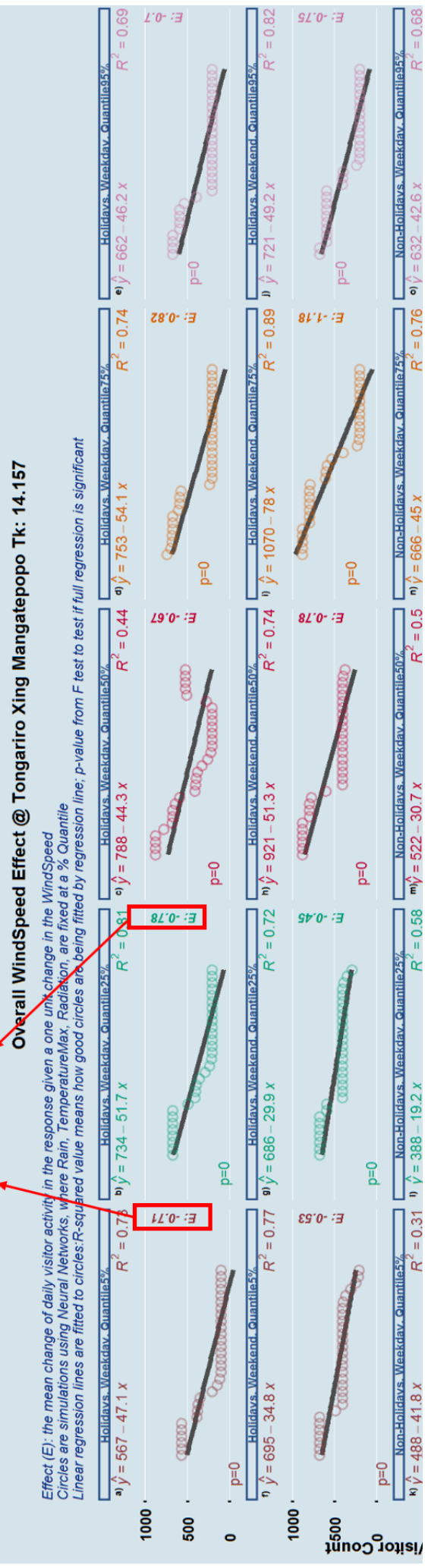


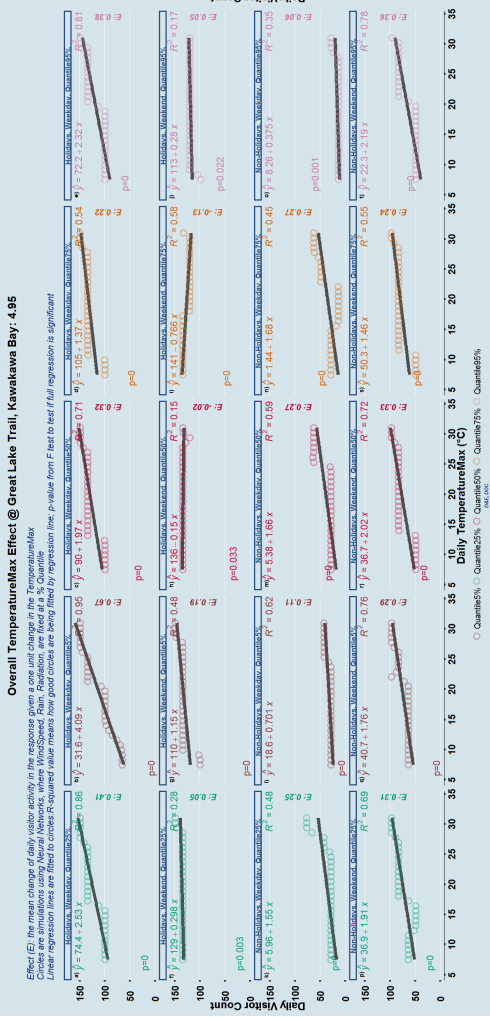
Figure 4: How overall weather Effects (before scaling) being calculated under 20 scenarios on Tongariro Xing Mangatepopo Tk (Tongariro Alpine Crossing)

3 Case studies: weather impact on sites with close geo-locations

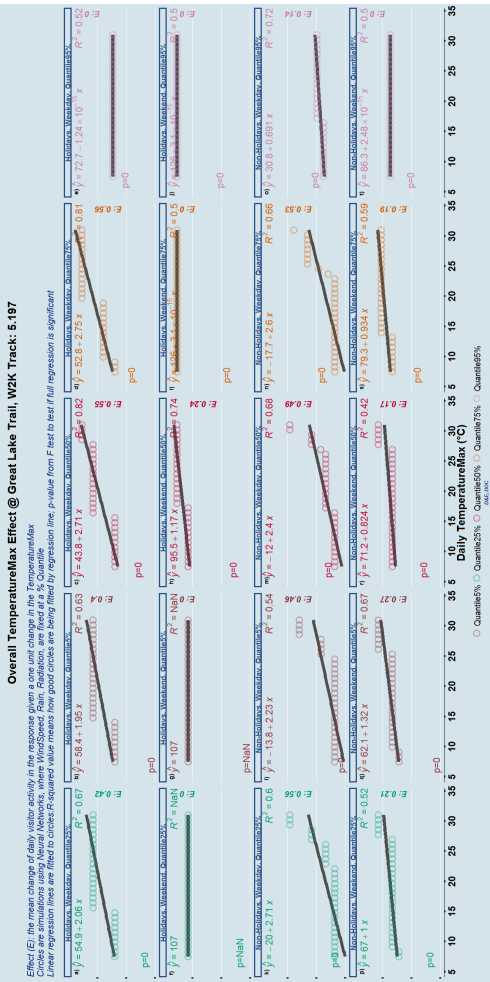
Track sites with close geo-locations are supposed to be have similar weather impact (**especially when visitation levels are similar**). It is interesting to test how AI models interpret this and validate this.

in [Fig.5](#) and [6](#) we singled out such a paired sites: Great Lake Trail, Kawakawa Bay and Great Lake Trail, W2K Track, and show you how similar weather impact could be on each's visitations.

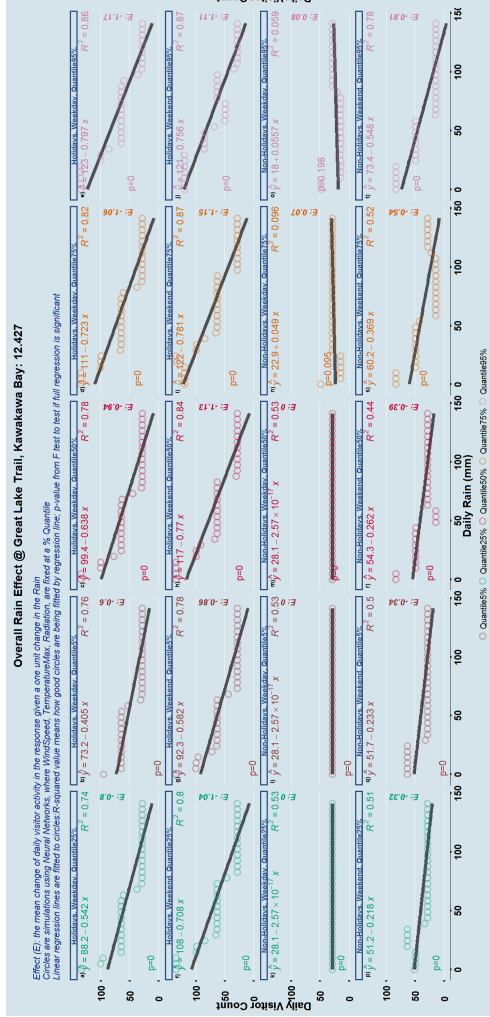
Temperature: 4.95



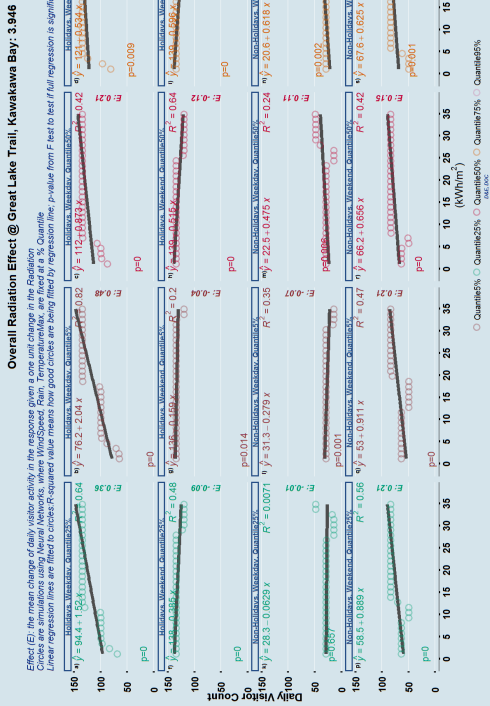
Temperature: 5.197



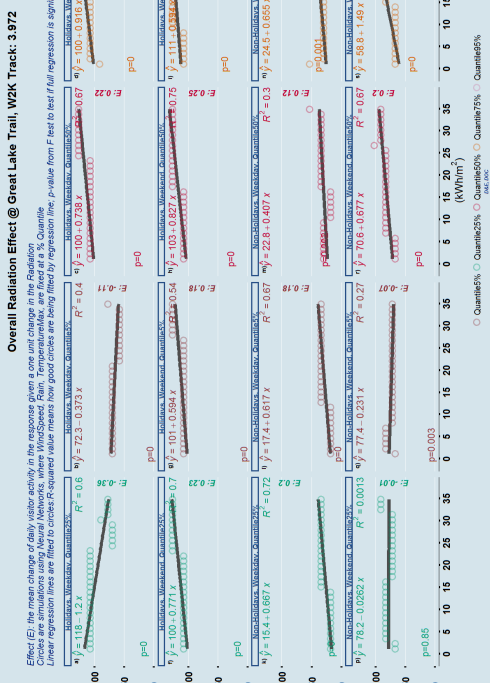
Rain: 12.427



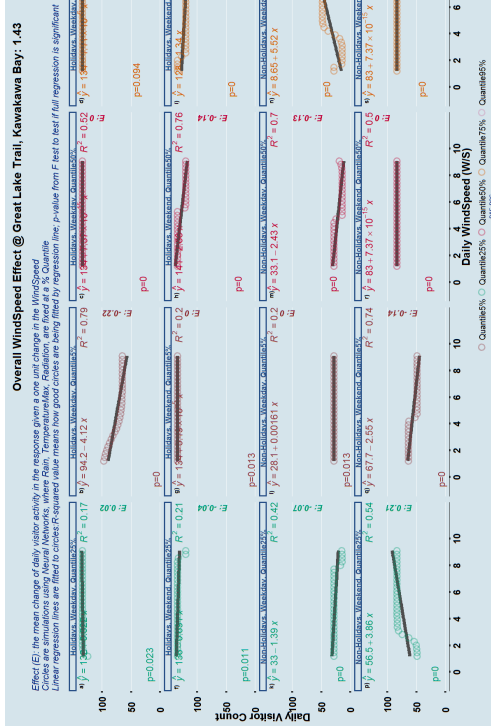
Radiation: 3.946



Radiation: 3.972



WindSpeed: 1.43



WindSpeed: 4.549

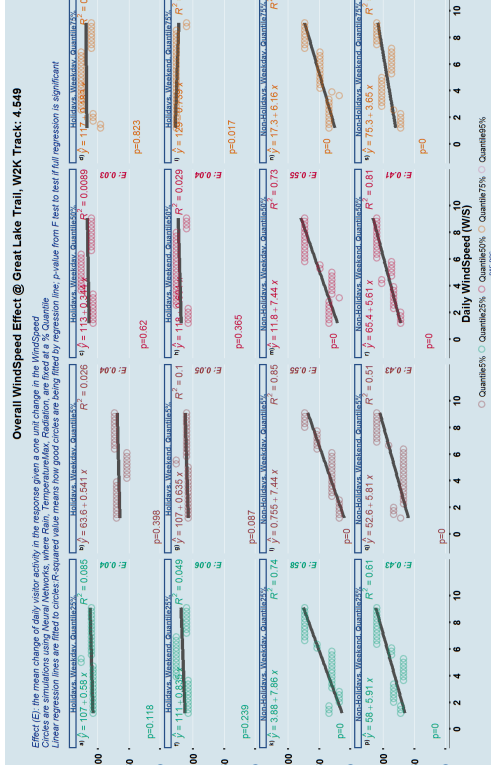


Figure 6: Radiation and WindSpeed effects for Great Lake Trail (Left), Kawakawa Bay and Great Lake Trail, W2K Track (Right)

4 Results and Discussions

4.1 Classification of weather impact across NZ

Four levels of weather impact were considered to classify weather impact on visitations across NZ

1. Level 1: weak. All weather Index is in the range of 0 to 0.25
2. Level 2: lower medium. All weather Index is in the range of 0.25 to 0.5
3. Level 3: higher medium. All weather Index is in the range of 0.5 to 0.75
4. Level 4: strong. All weather Index is in the range of 0.75 to 1

5 Acknowledgement

This work benefits from numerous discussions with DOC's other statisticians: Ian Westbrooke, Helene Thygesen and NIWA climate scientist Andrew Tait. Zac Taylor has been taken great caution in QA piles of counter data. Also, other NIWA's staffs are appreciated: Andrew Tait, Petra Pearce, Gregor Macara who kindly provided VCSN weather data.

Reference

- 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](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>.
- Stefaniak, Bogusław, Witold Cholewiński, and Anna Tarkowska. 2006. "Algorithms of Artificial Neural Networks - Practical Application in Medical Science." *Polski Merkurusz Lekarski : Organ Polskiego Towarzystwa Lekarskiego* 19 (January): 819–22.
- 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. 2006a. "Interval Estimation of Urban Ozone Level and Selection of Influential Factors by Employing Automatic Relevance Determination Model." *Chemosphere* 62 (April): 1600–1611. <https://doi.org/10.1016/j.chemosphere.2005.06.047>.
- . 2006b. "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>.