DEVELOPING SOUTH AFRICAN TOURISM CLIMATE INDEX

S Krishnannair *, I Nandutu† and M Ali ‡

Industry Representative

Jennifer Fitchett[1]

Study Group Participants

A. Mavimbele, M. Ali, D. Patel, I. Nandutu, M. Trivedi, S. Bhatt, S. S. S. Jamaludin, S. Mokhanda and S. Krishnannair

Abstract

Travel and Tourism as a sector has been resilient despite the diverse climatic conditions that affect tourists. Climatic conditions prevailing in the tourist destinations determine the number of tourists visiting those destinations. In 1985, Mieczkowski developed the Tourism Climate Index (TCI) to estimate the impact of climatic conditions on tourism. A mathematical model for Tourism Climate Index South African (TCI_{SA}) is proposed. Five climatic variables such as maximum temperature, minimum temperature, precipitation, cloud cover and wind speed are used as the input to the index to estimate the values of TCI_{SA}. The model for TCI_{SA} is validated using the climate data for the Durban region. Results show Durban as one of the popular tourist destinations due to its climate suitability for various tourist activities. This is comparable with the data from TripAdvisor. This study recommends the validation of the mathematical model by using more data points and climatic variables.

*Department of Mathematical Sciences, University of Zululand, Natal, South Africa . *email: KrishnanairS@unizulu.ac.za*

†Department of Mathematics, Rhodes University, South Africa. *email:*

‡School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa. *email: montaz.ali@wits.ac.za.*

[1]School of Geography, Archaeology and Environment Studies. *email:jennifer.fitchett@wits.ac.za*

1 Introduction

Climate plays a key role in the sustainability of the tourism sector of a country [1] [2]. Studies have shown that the climate change is considered one of the most affecting factors in decision making for the selection of tourist destinations during holidays and other leisure activities [3]. Among several metrics and indexes that have been developed to study the influence of climate change in the field of tourism is the Tourist Climate Index (TCI). Mieczkowsk (1985) introduced the concept of TCI as a measure to evaluate the influence of various climatic indices on the quality of experience for the average tourist. Generally, a suitable climate is defined by temperatures within human comfort levels, no rainfall, low wind speed low humidity, little to no cloud cover and maximised sunshine hours. This is determined by the TCI formula[4]

$$TCI = 2(4CD + CA + 2R + 2S + W)$$

where CD = daytime thermal comfort, CA = average thermal comfort, R = total monthly rainfall, S = total monthly sunshine hours and W = monthly average wind speed.

The tourism sector in South Africa (SA) attracted over 14.3 million people over the last decade [5]. It contributes 3% of the gross domestic product of South Africa [5] [6]. Tourism also plays an important role in reducing unemployment rates as more than half a million of the population in SA work in the tourism sector [7] [8] [5]. The South African tourism sector is negatively affected by the abrupt and unpredictable climate change due to the low adaptive nature of indoor and outdoor tourist attractions in SA. Therefore, the awareness of climate indices and their influence on tourism are essential knowledge for tourist activities in SA. The tourist destination cannot be easily identified by observing behaviour of the individual climate indices alone or separately in tourism. This is because the climate conditions of a region depend on a group of indices such as temperature, wind and rain which are highly correlated to each other. In this study a data driven approach based on a mathematical model is used to analyse the climate patterns and respective changes in the temporal distribution of the climate indices for tourism in SA. The proposed model is used to estimate Tourism Climate Index South African (TCI_{SA}) to specify the favourable or preferred month for tourism in different regions in SA.

2 Problem statement

In 1985 Mieczkowski developed the TCI, a method to quantify, classify and compare the climatic suitability of various destinations for tourism and to determine changes in climatic suitability through time. This index was based on expert opinion regarding the climatic factors which are of importance to tourists and the relative weighting thereof. In the past five years a proliferation of new indices have emerged such as the Climate Index for Tourism, the Beach Tourism Index [9], the Holiday

Climate Index (Urban), Holiday Climate Index (Beach) [10] and the Camping Climate Index [11]. Each of these is argued to be more suitable than the TCI as they are based on the experience of tourists, rather than expert opinion. They all rely on only the experiences of tourists in the Northern Hemisphere ignoring Africa. This problem involves developing a (TCI_{SA}) based on the experiences of tourists and behaviour of climatic variables/indices in different regions in South Africa. The index will need to be appropriate to the tourist expectation of southern Africa (sun, sea and nature), be operational using the regularly collected meteorological variables in southern Africa and produce output measures of climatic suitability that are consistent with the experiences of tourists.

3 Methods and results

The TCI_{SA} in this study uses five climatic variables which are related to the three components essential to tourism called thermal comfort, aesthetic and physical component. The five climatic variables used for the TCI_{SA} input are maximum temperature (X_1), minimum temperature (X_2), precipitation (X_3), cloud cover (X_4) and wind speed (X_5). The TCI_{SA} score is calculated based on the following formula:

$$TCI_{SA} = X_1 + X_2 + X_3 + X_4 + X_5 \tag{3.1}$$

where $X_1, \ldots, X_5$ are the input features of the climate elements. The estimation of TCI using a mathematical model involves the following steps.

3.1 Pre-processing data and visualisation

We collected data from Climate Explorer [12]. Variables that were considered are maximum temperature, minimum temperature, precipitation, cloud cover and wind speed from Durban. The data was split into a training set and a test set. We also visualised 2017 data of tourist opinions from Trip Advisor. Trip Advisor data helped us determine the Towns to consider when collecting climate data from climate explorer. Figure 1 shows people who consider weather asan important factor before travelling. Cape Town and Durban were the cities for which a larger number of tourists selected their destination based on the reviews.

To exhaustively understand the Tourism experience in South Africa we visualised Trip Advisor reviews where 464 climate reviews have been recorded. Referring to Figure 2, almost 40 per cent of tourists have commented on the cold weather, followed by 30 per cent who are very sensitive to the hot weather. The tourists also mentioned rainy days and sunny days, which are 10-11 per cent. Tourists mentioned cloud, humidity, mist and wind less frequently.

Figure 3 shows the number of climate references made by tourists to specific destinations. Tourists in Cape Town have the highest number, while the lowest is in Port Nolloth. Almost all 19 destinations in South Africa seemed to have received comments on 'cold' weather, as shown in Figure 4. The greatest number of comments

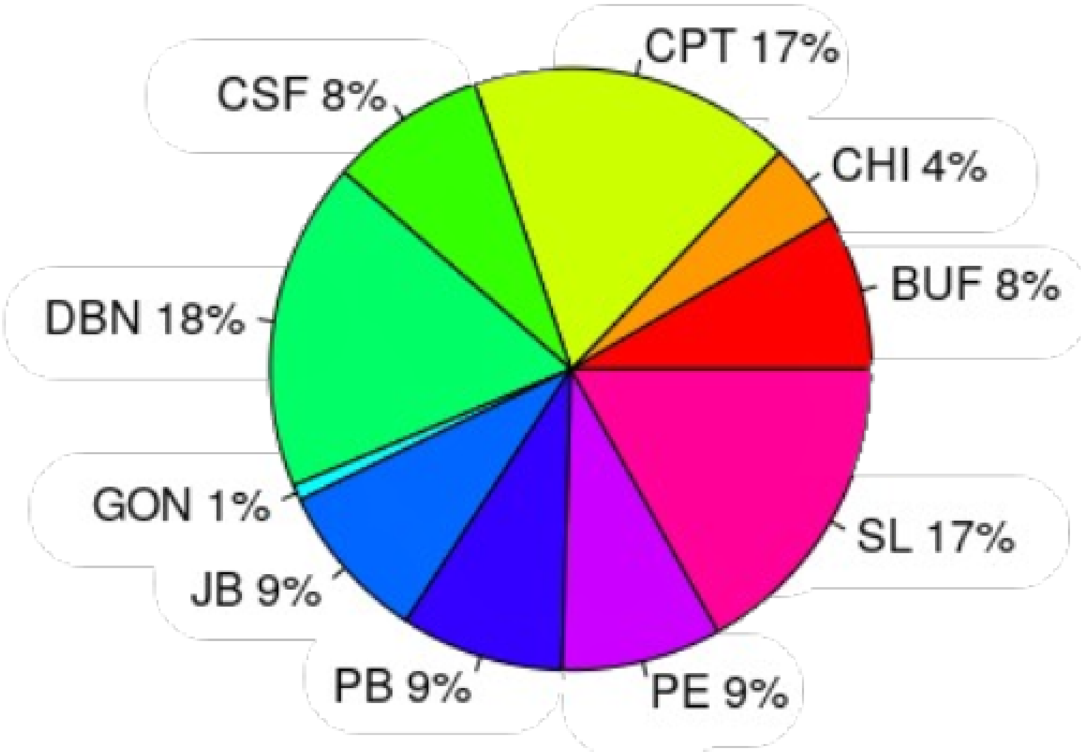


Figure 1: Proportions of Destination Weather Consideration before Travel

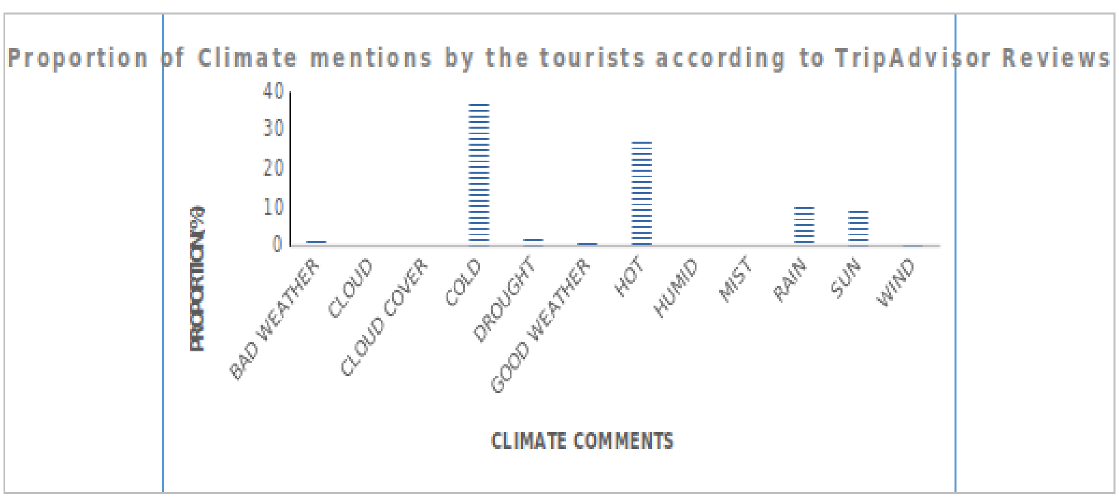


Figure 2: Proportion of Climate mentions by the tourists for all 19 destinations in South Africa.

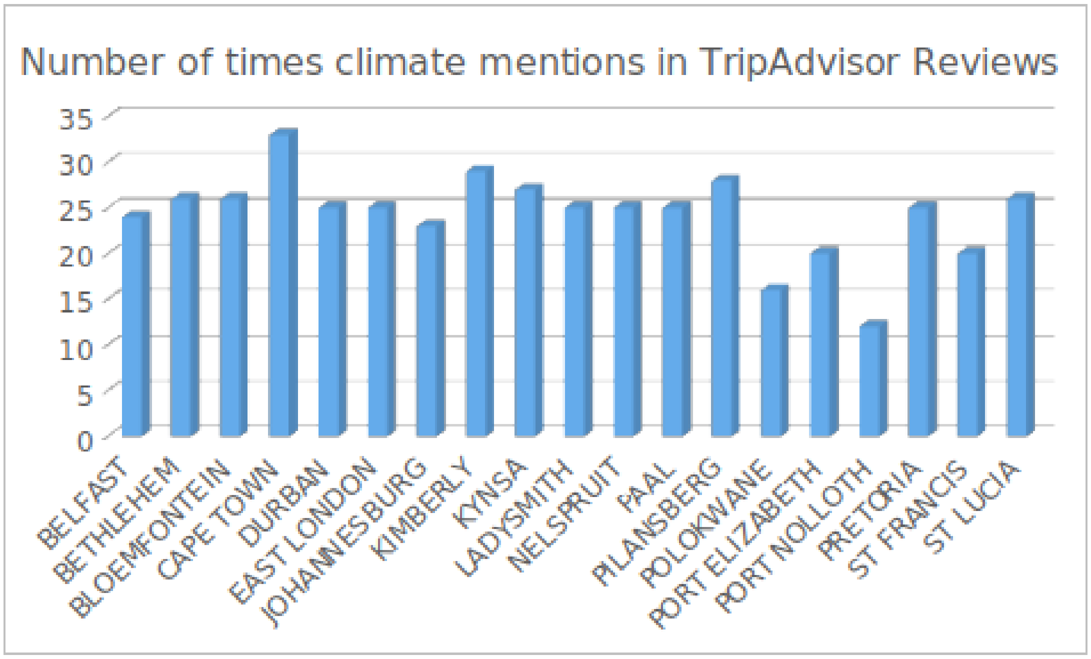


Figure 3: Climate mentioned for all 19 destinations in South Africa.

were made by the tourists who visited Bethlehem about the cold weather. On the other hand, comments on 'hot' weather were recorded in Kimberly.

Figure 5 shows the rating scale for the questionnaire data. Scale 1 is for the highest rating while 5 is for the lowest rating. All destinations record the highest rating for sunshine followed by temperature. Rating scale 5 is given for rainfall. These show the tourists prefer a sunny day compared to a rainy day to do their activities at the destinations. The sensitivity of tourists to the weather of each destination may give some indication of the climate suitability of the destination. Here, we conduct a statistical analysis using the contingency table to determine the link between the climate reviews and the destinations of the trip.

Ho: The climate reviews made by the tourists is not associated with the destination

H1: The climate reviews made by the tourists is associated with the destination

For the test hypothesis, a chi-square test was performed. A p-value=0.000, which is less than 5% of the significance level, is observed. We therefore conclude that the climate reviews made by the tourists are linked to the destination. In other words, it means that tourists have become aware of the climate.

3.2 Mathematical model - Regression Analysis

Now we develop a mathematical model to estimate the TCI. We consider data for Durban from 2005 to 2019 (downloaded from [12]) of the following variables:

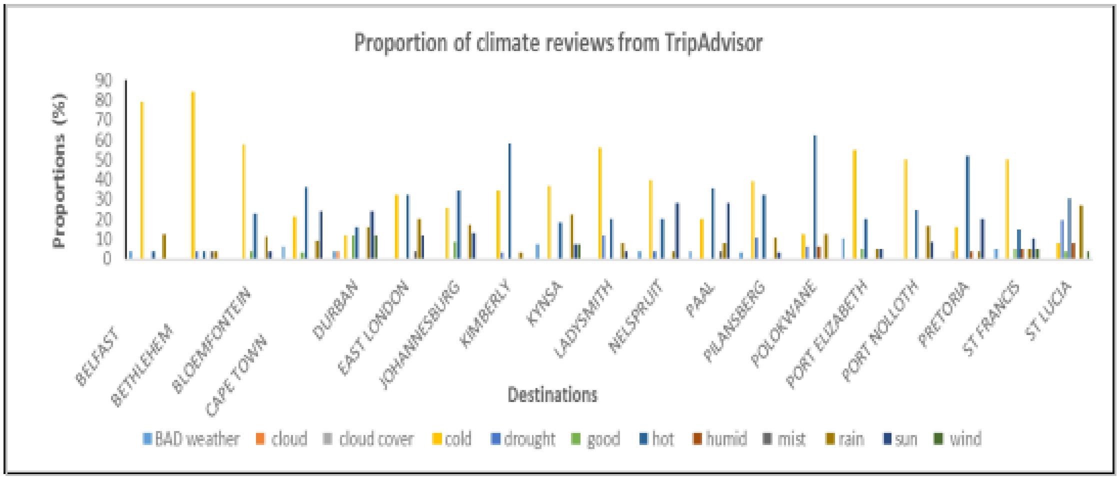


Figure 4: Proportion of climate reviews for all 19 destinations in South Africa.

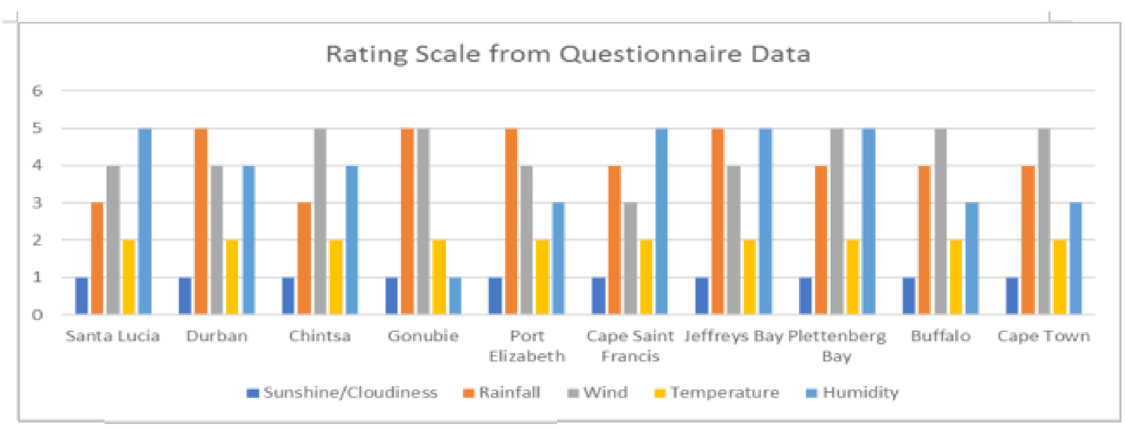
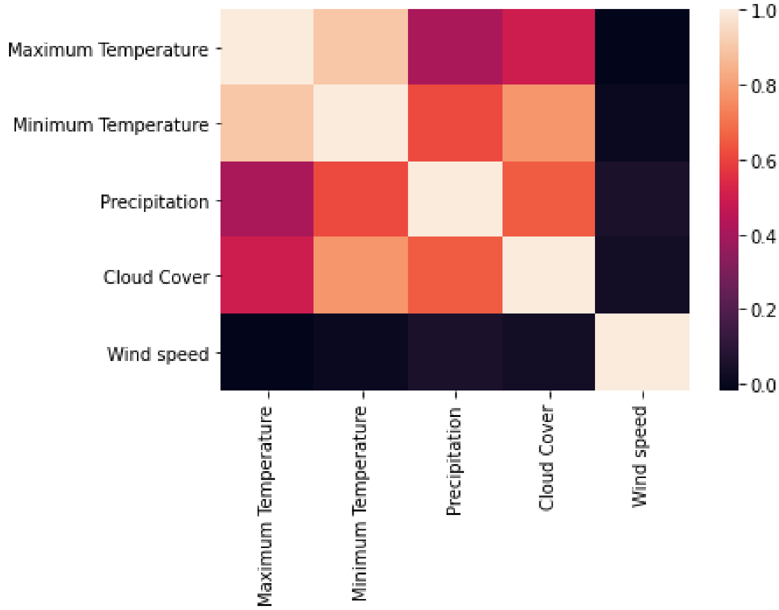


Figure 5: Rating Scale for Questionnaire Data.



h

Figure 6: Correlations between Climate Elements

1. Maximum temperature (x_1)
2. Minimum temperature (x_2)
3. Precipitation (x_3)
4. Cloud cover (x_4)
5. Wind speed (x_5)

There are 25 correlation pairs of the considered variables. The correlation matrix shown below and in Figure 6 indicates that there is no correlation of variable x_5 with the other variables.

$$C = \begin{pmatrix} 1 & 0.9042 & 0.4088 & 0.4988 & -0.0168 \\ 0.9042 & 1 & 0.6182 & 0.7803 & 0.0117 \\ 0.4088 & 0.6182 & 1 & 0.6543 & 0.0551 \\ 0.4988 & 0.7803 & 0.6543 & 1 & 0.0310 \\ -0.0168 & 0.0117 & 0.0551 & 0.0310 & 1 \end{pmatrix}$$

Hence correlations of x_5 with all the other variables can be ignored in the model. Moreover all the diagonal elements are correlations of a variable x_i with itself for $i = 1$ to 5. Hence the diagonal correlations along the diagonal can be ignored. The correlation matrix is a symmetric matrix. Therefore only the upper triangular correlations are considered as parameters of the model. Hence the matrix for the non-linear model is of the form

$$M = \begin{pmatrix} 0 & a_{12} & a_{13} & a_{14} & 0 \\ 0 & 0 & a_{23} & a_{24} & 0 \\ 0 & 0 & 0 & a_{34} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

From this matrix, consider the model

$$f_N(x) = a_0 + a^T x + x^T M x \tag{3.2}$$

where $a_0 \in \mathbb{R}$,
$a^T = (a_1, a_2, a_3, a_4, a_5)$,
$x^T = (x_1, x_2, x_3, x_4, x_5)$.

Expanding the right hand side of (3.2) using matrix multiplication, we obtain

$$f_N(x) = a_0 + \sum_{j=1}^{5} a_j x_j + \frac{1}{2} \sum_{i=1}^{3} \sum_{j=i+1}^{4} a_{ij} x_i x_j.$$

In expanded form

$$f_N(x) = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{14}x_1x_4$$
$$+ a_{23}x_2x_3 + a_{24}x_2x_4 + a_{34}x_3x_4 \ .$$

Hence there are 12 parameters a_0, a_j, a_{ij} for the non-linear model considered. To include the non-linearity in the model, consider the logistic function.

$$F(x) = \frac{e^x}{1 + e^x}.$$

Now substituting equation (3.2) in the logistic function, gives

$$F\left(f_N\left(x\right)\right) = \frac{e^{f_N(x)}}{1 + e^{f_N(x)}}.$$

Let

$$\bar{a}^T = (a_{12}, a_{13}, a_{12}, a_{23}, a_{24}, a_{34} \ .)$$

Now consider a single given data, say

$$x^{j^T} = (x_1, x_2, x_3, x_4, x_5) \ .$$

This implies from equation (3.2),

$$f_N(a_0, a, \bar{a}) = a_0 + a^T x^j + \frac{1}{2} x^{jT} M x^j \ .$$

From equation (3.2),

$$F\left(f_N\left(a_0, a, \bar{a}\right)\right) = \frac{e^{f_N(a_0,a,\bar{a})}}{1 + e^{f_N(a_0,a,\bar{a})}} \; .$$

For given data x^j, we can write

$$F\left(f_N\left(a_0, a, \bar{a}, x^j\right)\right) = \frac{e^{f_N\left(a_0,a,\bar{a},x^j\right)}}{1 + e^{f_N(a_0,a,\bar{a},x^j}}$$

and therefore

$$f(a_0, a, \bar{a}, x^j) = F(f_N(a_0, a, \bar{a}, x^j)) \; .$$

Consider the maximum likelihood

$$\max \prod_{j=i}^{N} F(f_N(a_0, a, \bar{a}, x^j)). \tag{3.3}$$

From the above maximum likelihood function, we obtain

$$\max_{a_0,a,\bar{a}} \ln \prod_{j=i}^{N} F(f_N(a_0, a, \bar{a}, x^j))$$

$$\Longrightarrow \max_{a_0,a,\bar{a}} \sum_{j=i}^{N} \ln F(f_N(a_0, a, \bar{a}, x^j))$$

$$\Longrightarrow \min_{a_0,a,\bar{a}} - \sum_{j=i}^{N} \ln F(f_N(a_0, a, \bar{a}, x^j))$$

$$\Longrightarrow \min_{a_0,a,\bar{a}} \sum_{j=i}^{N} \ln \frac{1}{F(f_N(a_0, a, \bar{a}, x^j)}$$

$$\Longrightarrow \min_{a_0,a,\bar{a}} \sum_{j=i}^{N} \ln \left[1 + e^{-f_N\left(a_0,a,\bar{a},x^j\right)}\right].$$

Let us denote

$$L(a_0, a, \bar{a}) \;= \sum_{j=1}^{N} \ln \left[1 + e^{-f_N\left(a_0,a,\bar{a},x^j\right)}\right].$$

The aim is to minimize the objective function L by adjusting the 12 model parameters represented by $a_0, a, \bar{a}$. For this optimization process, it is required to determine the gradient of the objective function L with respect to the 12 model parameters

$a_0, a, \bar{a}$. Now the gradient of L with respect to $(a_0, a, \bar{a})$ is

$$\begin{aligned}
\nabla L(a_0, a, \bar{a}) &= \sum_{j=1}^{N} \nabla \ln\left[1 + e^{-f_N\left(a_0, a, \bar{a}, x^j\right)}\right] \\
&= \sum_{j=1}^{N} \frac{1}{1 + e^{-f_N}} \nabla(1 + e^{-f_N}) \\
&= \sum_{j=1}^{N} \frac{e^- f_N}{1 + e^{-f_N}} \nabla\left(-f_N\left(a_0, a, \bar{a}/x^j\right)\right) \\
&= \sum_{j=1}^{N} \frac{-e^{-f_N}}{1 + e^{-f_N}} \nabla(f_N(a_0, a, \bar{a}))\ .
\end{aligned}$$

Hence finally we obtain the gradient of the objective function L with respect to the 12 model parameters $a_0, a, \bar{a}$:

$$\nabla L(a_0, a, \bar{a}) = \sum_{j=1}^{N} \frac{-e^{-f_N}}{1 + e^{-f_N}} \begin{pmatrix} 1 \\ x_1^j \\ x_2^j \\ x_3^j \\ x_4^j \\ x_5^j \\ x_1^j x_2^j \\ x_1^j x_3^j \\ x_1^j x_4^j \\ x_2^j x_3^j \\ x_2^j x_4^j \\ x_3^j x_4^j . \end{pmatrix}$$

Here, it is clear that $\nabla L(a_0, a, \bar{a}) \in \mathbb{R}^{12}$. The steps of the optimization procedure to minimize the objective function $L(a_0, a, \bar{a})$ by adjusting the 12 model parameters $a_0, a, \bar{a}$ is summarised below.

Optimization Procedure

Let $A = (a_0, a, \bar{a})$

1. Initialize $A^k, k = 0$ at random.
   Calculate the gradient of $A^k, k = 0$
   i.e. $\nabla L(A^0) \in \mathbb{R}^{12}$ or $\nabla L(A^k), k = 0$

2. Set $\beta = 1$

3. Find $A^{k+1} = A^k - \beta \nabla L(A^k)$
4. Compare if $L(A^{k+1}) < L(A^k)$
   Set $k = k + 1$ and go up to step (v)
   else $\beta = \frac{1}{2}\beta$ and go to step (iii)
5. Calculate $\nabla L(A^{k+1})$,
   Stop if $\|\nabla L(A^{k+1})\| < 10^{-3}$
   Else go to step (ii) with A^{k+1}

After this procedure, we obtain 12 optimized model parameters $A^* = (a_0^*, a^*, \bar{a}^*)$. On inserting the optimized model parameter values into equation (3.2), we define the TCI Predictor as

$$f(x) = \frac{e^{f_N(x)}}{1 + e^{f_N(x)}} \ .$$

Considering training data

$$x^i = (x_{i1}, x_{i2}, x_{i3}, x_{i4} x_{i5})^T,$$

we obtain a TCI value

$$TCI = f(x^i) \times 100.$$

Note that the proposed TCI predictor will lie in the range $[0, 100]$.

4 Discussion of results

1. Database for Durban city was downloaded from [12].
2. The data is for the years 1905-2019 and 5 parameters: maximum temperature, minimum temperature, precipitation, cloud cover and wind speed.
3. The correlation matrix between these parameters was constructed.
4. The non-linear model was developed based on the 12 parameters of the correlation matrix: 5 individual parameters, 6 pairs of parameters and 1 constant parameter.
5. The l ogistic function was used to determine the objective function.
6. The o ptimization procedure was proposed and a python code was attempted for the optimization procedure.
7. The optimization procedure was performed for five data points and one iteration.

For five data points for Durban and one iteration of the optimization procedure, the TCI index for 10 data points was estimated and is shown in Figure 7.
The limitation of the python code are:

```
For data point 1 TCI =  98.7203673814043
For data point 2 TCI =  98.5223043507944
For data point 3 TCI =  96.8039115112577
For data point 4 TCI =  95.8143348032390
For data point 5 TCI =  92.6261944137983
For data point 6 TCI =  85.3474842735185
For data point 7 TCI =  87.1983903080991
For data point 8 TCI =  86.3638204714567
For data point 9 TCI =  93.0047489922248
For data point 10 TCI =  95.1612308312618
```

Figure 7: Screenshot of python code, TCI_{SA} Values for the Logistic Function.

1. Only the data for Durban was considered.
2. Humidity data was unavailable.
3. For simplicity, only 5 data points are taken (against 1380).
4. For simplicity, the algorithm is run for only 1 iteration.

As a part of future work, the code can be updated to overcome above limitations.

5 Conclusion and recommendations

In conclusion, this paper aimed at developing a TCI_{SA} that contributes to determining the climate suitability for Tourists and increase on the Tourists tourism experience in South Africa. A mathematical model is proposed to estimate the TCI_{SA} index for South Africa. Five climate elements are used as an input for the model. The proposed model is validated using the climate data for Durban. Based on the results from this work, we recommend the following.

- Development of a Holiday Climate Index, Beach Climate Index and Camping Climate Index for South Africa is recommended.
- Customized TCI_{SA} index to integrate tourist opinion as part of the thermal comfort factors is recommended.
- The effect of other climatic indices such as Humidity , Sunshine and Extreme weather patterns in tourism can be of interest to specify the tourists popular destinations in SA.
- Use of the nonlinear statistical model and ML techniques is recommended to handle the multicollinearity of climatic indices in the estimation of TCI
- Because of competitive pressure within the tourism industry, tourism marketers need to innovate continuously to enjoy competitive advantage and to deliver customer satisfaction more effectively and efficiently. It will translate into long term development and growth of tourist destinations and help in

achieving strong market position. Hence, future research can be undertaken to develop an Index for Destination Attractiveness (IDA) where along with Climate, other important factors and corresponding variables can be studied.

Acknowledgement

Thank you to Jennifer Fitchett for the expert guidance.

References

[1] Becken S. Harmonising climate change adaptation and mitigation: The case of tourist resorts in Fiji, Global Environmental Change, **15**, 2005, 381-393.

[2] Gossling S, Scott D, Hall C M, Ceron J-P and Dubois, G. Consumer behaviour and demand response of tourists to climate change, Annals of Tourism Research, **39**, 2012, 36-58.

[3] Dabbas A A L, Gal Z and Attila B. Neural network estimation of tourism climatic index (TCI) based on temperature-humidity index (THI)-Jordan region using sensed datasets, Carpathian Journal of Electronic and Comoputer Engineering, **11**, 2018, 50-55.

[4] Mieczkowski Z. The tourism climatic index: a method of evaluating world climates for tourism. Canadian Geographer/Le Gographe Canadien, **29**, 1985, 220-233.

[5] Fitchett J M, Robinson, D and Hoogendoorn, G. Climate suitability for tourism in South Africa, Journal of Sustainable Tourism, **25**, 2017, 851-867.

[6] StatSA 2016. An economic look at the tourism industry. http://www.statssa.gov.za/?pD4362.

[7] Briedenhann J and Wickens E. Rural tourism – meeting the challenges of the new South Africa, International Journal of Tourism Research, **6**, 2004, 189-203.

[8] Rogerson C M. Tourism-led local economic development: The South African experience, Urban Forum, **13**, 2002, 95-119.

[9] Morgan R, Gatell E, Junyent R, Micallef A, Ãzha E and Williams A T. An improved user-based beach climate index. Journal of Coastal Conservation, **6**, 2000, 41-50.

[10] Scott D, Rutty M, Amelung B and Tang M. An inter-comparison of the holiday climate index (HCI) and the tourism climate index (TCI) in Europe, Atmosphere, **7**, 2016 80. 17 pages.

[11] de Freitas C R. Tourism climatology: evaluating environmental information for decision making and business planning in the recreation and tourism sector. International Journal of Biometeorology, **48**, 2003, 45-54.

[12] World Heterological Organisation. Climate Explorer https://climexp.knmi.nl/start.cgi