LETTER • OPEN ACCESS

Optimisation-based refinement of genesis indices for tropical cyclones

View the article online for updates and enhancements.

Environmental Research Communications

LETTER

Optimisation-based refinement of genesis indices for tropical cyclones

Guido Ascenso[1] , Leone Cavicchia[2] , Enrico Scoccimarro[2] and Andrea Castelletti[1]

1   Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, Italy
2   Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy

E-mail: guido.ascenso@polimi.it

Abstract

Tropical cyclone genesis indices are valuable tools for studying the relationship between large-scale environmental fields and the genesis of tropical cyclones, supporting the identification of future trends of cyclone genesis. However, their formulation is generally derived from simple statistical models (e.g., multiple linear regression) and are not optimised globally. In this paper, we present a simple framework for optimising genesis indexes given a user-specified trade-off between two performance metrics, which measure how well an index captures the spatial and interannual variability of tropical cyclone genesis. We apply the proposed framework to the popular Emanuel and Nolan Genesis Potential Index, yielding new, optimised formulas that correspond to different trade-offs between spatial and interannual variability. Result show that our refined indexes can improve the performance of the Emanuel and Nolan index up to 8% for spatial variability and 16%–22% for interannual variability; this improvement was found to be statistically significant ($p < 0.01$). Lastly, by analysing the formulas found, we give some insights into the role of the different inputs of the index in maximising one metric or the other.

1. Introduction

Tropical cyclones (TCs) are among the deadliest and costliest natural disasters [1]. Over the last decades, the accuracy of short-term forecasts of TC genesis has increased substantially [2]. Yet, predicting global TC activity on longer time scales, from seasonal to multi-decadal, remains challenging [3]. This is due to the lack of a complete understanding of the processes behind the genesis of tropical cyclones [4]. In this context, empirical methods exploring the relationship between tropical cyclones and the large-scale environment are a suitable alternative [5] to developing models for short- and long-term TC forecasting and tracking and understanding how climate change will affect TC frequency and intensity globally and locally.

Two approaches exist for studying the TC-climate relationship. The first involves the use of general circulation models (GCMs) [6–9]. Once regarded as unable to simulate TC-like vortices accurately [10], such models have been improved considerably over the years thanks to increasing computational power. They are now robust enough that realistic TCs can be found in their simulations [11–14]. Nevertheless, they are either computationally expensive to run [15] or have relatively low spatial and temporal resolution [13]. Therefore, obtaining robust predictions of future TC activity from GCMs is still challenging [16]. However, to overcome the low resolution of such models, downscaling strategies can be applied which have been shown to be useful tools for studying climate projections of TC genesis [17, 18]. The most confident GCM projections [19] estimate that a 2 °C increase in global temperatures will lead to higher TC inundation levels [20], due to a combination of sea level rise, projected slower TC translation speed [21, 22] and higher TC precipitation rates [23–25]. Furthermore, whereas the frequency of TCs is projected to decrease (though this estimate is less confident than the ones above [20]), their average (and highest) intensity is projected to increase, and, as intensity is often linked with damage, so is their cost [11, 26].

Alternatively, the connection between environmental drivers and TCG can be empirically modelled [10, 16]. Typically, this connection takes the form of indices (namely Genesis Indices (GIs)) that combine a set of large-scale dynamic and thermodynamic atmospheric and oceanic predictors into a single variable. The magnitude of the index indicates how favourable environmental conditions are to the formation of cyclones. Under the assumption that the input variables can be skilfully predicted, it is possible to estimate how cyclone generation frequency and distribution will change under future climate conditions or predict if an upcoming hurricane season is likely to be particularly severe. Using GIs has some key advantages over relying on GCM-generated TCs:

(i) Indexes are easy to compute—they typically consist of the product of the predictors scaled by empirically derived exponents and coefficients [27].

(ii) They can be applied to data at different temporal and spatial resolutions, from regional to global scale. They can even be applied to data produced by different models.

(iii)
Currently, climate models are more skilful at simulating large-scale environmental variables than TC-like vortices, meaning the performance of GIs is hindered mainly by the chosen predictors and the way they are combined together.

GIs have been applied in a variety of studies, such as linking changes in TC activity to patterns of climate variability such as the Madden-Julian Oscillation [28] and the El Niño/Southern Oscillation [27, 29], or assessing the role of specific large-scale fields in TCG [5]. Yet, GIs suffer from some limitations. Although they are generally robust at replicating TCG's spatial distribution and seasonal cycle [16, 30, 31], they struggle to capture its interannual variability. Yu *et al* [16] reported a correlation coefficient of -0.04 between a particular GI and historical data for the interannual variability of TCG over the western North Pacific. Refined index versions involving different combinations of variables have been shown to improve this specific aspect [32]. However, the optimality of the chosen parameterisation and its effect on the index skill have not yet been explored. Furthermore, most of the commonly used GIs cannot explain the decreasing trend of TCs occurrence detected in climate projections [33]. Finally, though GIs work fairly well on different reanalysis datasets [34], they are typically fitted to a specific dataset with a given bias and spatio-temporal resolution and are not universally optimised [27]. Consequently, GIs need to be scaled by a factor when applied to datasets other than the one they were originally fit, making it cumbersome to compare the performance of different GIs or their performance across different datasets [30].

In this paper, we introduce an optimisation-based framework to refine GIs for three main purposes:

(i) To improve the performance of GIs by finding the globally optimum combination of the chosen predictors.

(ii) To provide a systematic assessment of the sensitivity of the index performance in reproducing a set of statistical properties of the observed TCs (e.g. spatial pattern and interannual variability) on the numerical coefficients used to weight the large-scale variables in the index.

(iii)
To offer several versions of the optimised GI, each corresponding to a different trade-off (selected by the user) between two performance metrics.

The rest of the paper will describe how the proposed framework is structured and present different optimised versions of the Emanuel and Nolan Genesis Potential Index [35], widely regarded as one of the best GIs available [30].

2. Datasets and methodologies

2.1. The ENGPI
The Emanuel and Nolan Genesis Potential Index (hereafter ENGPI) is a refinement of Gray's seasonal genesis parameter parameter [36], the first GI ever developed. We chose it because it is arguably the most studied GI in the literature [5, 15, 31, 37, 38, 10, 27, 28, 30, 16, 32]. Yet, the same framework described here can be applied to any other GI with minimal changes.

The ENGPI's formula is:

$$ENGPI = (1 + 0.1 \ V_{shear,850-200})^{-2} \left(\frac{RH_{700}}{50}\right)^3 \left(\frac{MPI}{70}\right)^3 \ |10^5 \eta_{850}|^{3/2} \tag{1}$$

where $V_{shear,850-200}$ is the magnitude of the vertical wind shear between 850 and 200 hPa (in ms^{-1}); RH_{700} is the relative humidity at 700 hPa (in %); η_{850} is the absolute vorticity at 850 hPa (in s^{-1}); and MPI is the Maximum Potential Intensity (in ms^{-1}), which is the theoretical maximum intensity a TC is expected to reach. MPI was originally formulated by Emanuel [39] and later modified by Bister and Emanuel [40], taking the form of the following equation:

$$MPI^2 = \frac{C_K}{C_D} \frac{T_s - T_o}{T_o}(h_o^* - h^*) \tag{2}$$

where C_K and C_D are the surface enthalpy and momentum exchange coefficients; T_s is the sea surface temperature (in K); T_o is the outflow temperature (in K); and h_o^* and h^* are the saturation moist static energies of the sea surface and of the free atmosphere, respectively. The MPI was arguably the most crucial addition in the ENGI formula compared to Gray's index, which instead uses the near-surface ocean thermal energy (integrated over waters with temperature above 26.5°C) and the vertical gradient of the equivalent potential temperature between the surface and 500 hPa.

2.2. Datasets

We used the best-track data from the International Best track Archive for Climate Stewardship (IBTrACS) project (version v04r00, downloaded from:https://www.ncdc.noaa.gov/ibtracs/; [41]), at a temporal resolution of 3h, between 1980 and 2020 as a source for observed TCG. IBTrACS is a curated collection of TC data compiled by multiple international organisations. In this study, we define TCG as the first time step in IBTrACS when a TC reached an intensity (in terms of maximum sustained wind speed) of 35 kt. We excluded all the events classified in IBTrACS as either spurious or extra-tropical [42] and counted all remaining TCG events over a global 2.5° × 2.5° latitude/longitude grid.

We took the inputs to the ENGPI from two reanalysis datasets: ECMWF's ERA5 dataset (https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset; [43]) and NASA's MERRA2 dataset (https://disc.gsfc.nasa.gov/datasets?project=MERRA-2; [44]). Prior to their download, we interpolated both datasets to a 2.5° × 2.5° latitude/longitude grid over the 1980–2020 period, with a monthly temporal resolution.
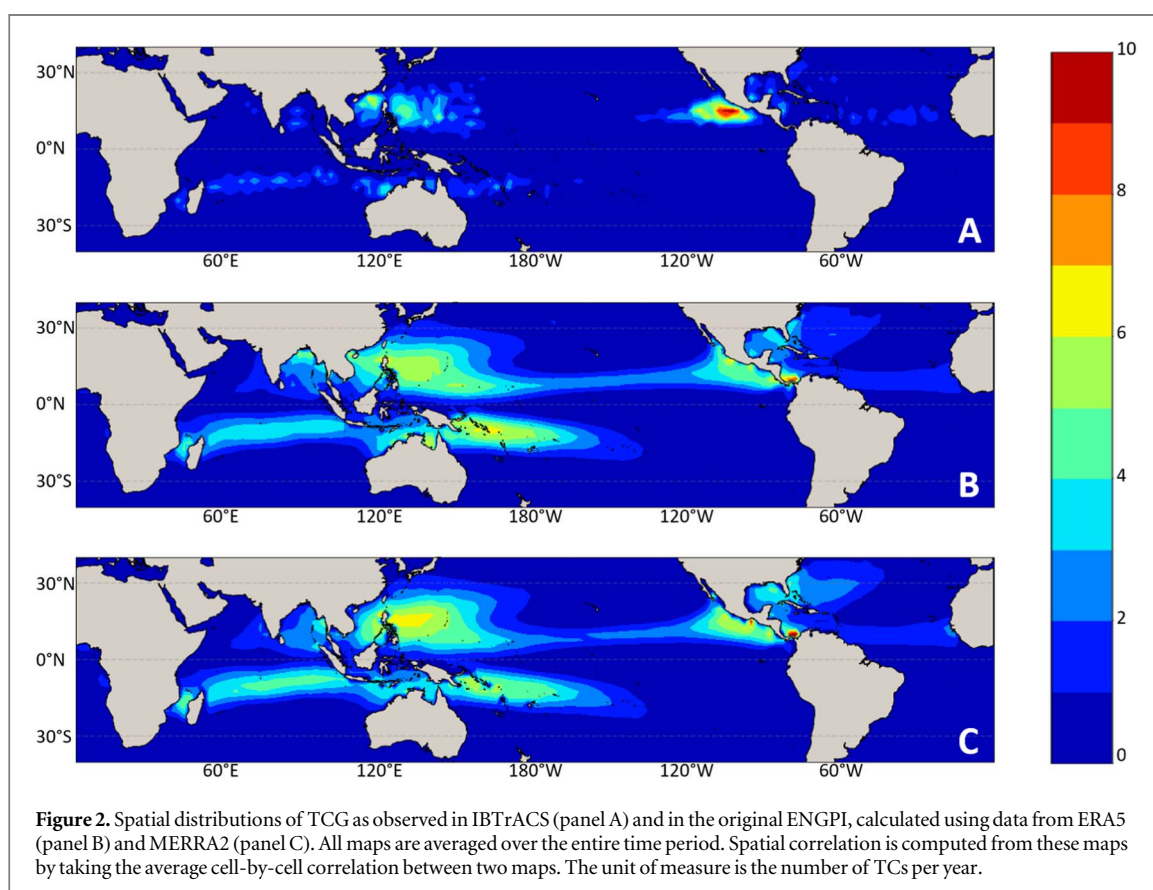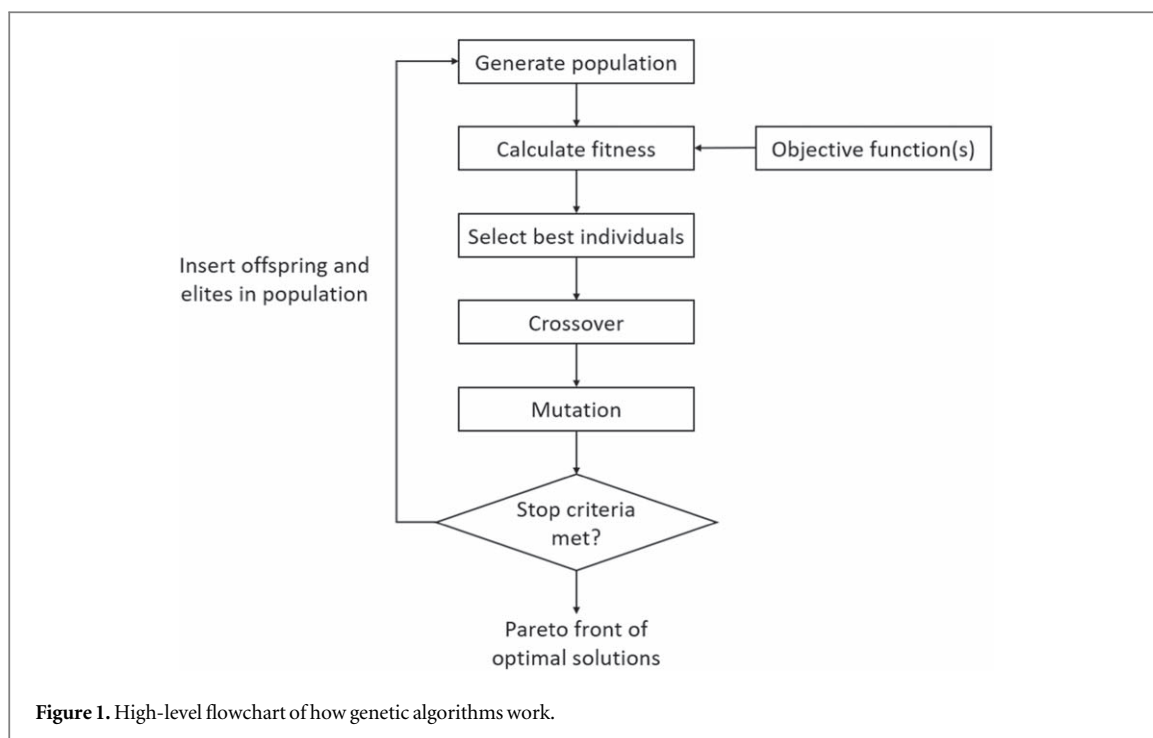
2.3. Optimisation framework

To find optimal configurations of the ENGPI, we used the NSGA-II multi-objective genetic algorithm [45]. Genetic algorithms have properties that make them especially suitable for this task:

- They efficiently explore an arbitrarily large space of solutions.

- Unlike gradient-based optimisation algorithms, they do not require derivatives to be computed and therefore can be applied to a wider range of functions.

- They support the definition of multiple optimisation objectives and they output a *Pareto front* of solutions, i.e., they generate a set of solutions that represent a spectrum of optimal trade-offs between maximizing one objective or the other(s).

- They are robust against local minima and maxima [46].

We chose the NSGA-II genetic algorithm because it is one of the most well known and studied multi-objective genetic algorithms and it has been found to perform well on a variety of tasks. Furthermore, it has been shown to outperform other genetic algorithms, particularly when few objectives are used [47–49]. An in-depth description of NSGA-II is beyond the scope of this paper[2], but a simple, high-level overview of how it works is given in figure 1. In brief, given a series of parameters to optimise, the algorithm generates an initial set (*population*) of random candidates, then selects the ones that perform best according to some specified objective function(s), and finally introduces randomness in the selected solutions by randomly mixing them together (*crossover*) and mutating some of their parameters; a new population is then created (retaining the offspring of the previous generation and the best-performing (*elite*) individuals of all previous generations), and the process continues until either a set number of *generations* has been created or the objectives have reached a desired value.

In our proposed framework, the NSGA-II algorithm searches the space of the parameters in the ENGPI formula and optimises them so that the resulting ENGPI maximises two objective functions:

---

[2] Readers interested in the details of its implementation should refer to its original paper: [45]

Figure 1. High-level flowchart of how genetic algorithms work.

Figure 2. Spatial distributions of TCG as observed in IBTrACS (panel A) and in the original ENGPI, calculated using data from ERA5 (panel B) and MERRA2 (panel C). All maps are averaged over the entire time period. Spatial correlation is computed from these maps by taking the average cell-by-cell correlation between two maps. The unit of measure is the number of TCs per year.

- *spatial correlation*, defined as the cell-by-cell correlation between the ENGPI and the observed TCG data from IBTrACS (averaged over the entire period) (e.g., figure 2).

- *temporal correlation*, defined as the correlation between the curves of yearly TCG of the ENGPI and the observed TCG data (e.g. figure 4).

We ran NSGA II for 10,000 iterations searching the exponent's space in the range of [-3, 3] and coefficient's in the range of [0.01, 100]; we chose the extremes of the ranges so that the search space would be as large as possible while preventing the computed indices from incurring into numerical over/underflow (i.e. due to too large/small exponents). The algorithm was also allowed to choose at which pressure level to select each variable. The pressure levels from which we allowed the algorithm to choose were 1000, 850, 700, 600, 500, 250, 200, 100, 50, 10, and 1 hPa; we selected these levels to give the algorithm the largest possible search space while avoiding taking all pressure levels present in ERA5 and MERRA2, due to computational constraints. Furthermore, following the improvement over the ENGPI proposed by Wang and Murakami [32], we added an additional parameter (a power of the exponential function) to the ENGPI formula, transforming it into:

$$ENGPI = (1 + 0.1\ V_{shear, 850-200})^{-2} \left(\frac{RH_{700}}{50}\right)^3 \left(\frac{MPI}{70}\right)^3 |10^5 \eta_{850}|^{3/2}\ e^x \qquad (3)$$

where x, the exponent of the exponential, is optimised along with all other exponents and coefficients. The purpose of this parameter is merely to scale the ENGPI and should not impact the index's physical properties.

We re-evaluated the optimal solutions on both datasets using three metrics: spatial correlation, temporal correlation, and seasonal correlation, defined as the correlation between the curves of monthly total TCG in ENGPI and IBTrACS (e.g., figure 3). Here, two of the objectives correspond to two of the metrics, but this is not necessarily always the case, as the optimisation and evaluation phases are disentangled.

Following Emanuel and Nolan [35] we use the seasonal correlation as an evaluation metric and therefore it was included here for comparison.

3. Results and discussion

3.1. Baseline
The evaluation metrics reported in table 1 constitute the baseline against which to compare the solutions found by the optimisation algorithm. The plots from which the correlations in table 1 were calculated are shown in figures 2, 3, and 4 for spatial, seasonal, and temporal correlation, respectively. As per Menkes *et al* [30], the values shown in figure 1 are scaled by a factor that brings the total average yearly number of TCs to equal that of data from IBTrACS.

As previously reported, these results shown how the original ENGPI captures well the spatial and seasonal distribution of TCG (and it does so equally well across different datasets), but fails to capture the interannual variability of TCG.

3.2. Optimised solutions
The optimisation runs for the ERA5 and MERRA2 datasets generated two Pareto fronts (figures 5 and 6, respectively). To illustrate the differences between the solutions found, we sampled three solutions from each front (see table 2 and figures 5 and 6): one for either extreme of the front, and a balanced trade-off between spatial and temporal correlation.

Let us first consider solutions (b) and (e), which are arguably the most well-balanced solutions found for ERA5 and MERRA2, respectively. For the dataset on which they were optimised, both solutions performed better than the baseline by about 8% for spatial correlation and 16%–22% for temporal correlation. The statistical significance ($p < 0.01$) of these results was confirmed using a Monte Carlo approach: 50,000 versions of the ENGPI were created with random exponents, and their performance (in terms of spatial and temporal correlation) was tested against the baseline ENGPI, creating a statistical distribution of the improvement of the randomised indices over the baseline ENGPI; Solutions (b) and (e) were found to be within the 99th percentile of this distribution. By comparing these solutions with neighbouring ones in the Pareto front, we found that the main similarity between them was that they all selected absolute vorticity at 600 hPa instead of at 850 hPa. Although most numerical models use vorticity at 850 hPa, some evidence of the physical importance of vorticity at 600 hPa exists. Specifically, Wang [50] and Murthy *et al* [51] show that in the vertical structure of TCs the peak of vorticity occurs at 600 hPa, not at 850 hPa. Therefore, the algorithm chooses to take vorticity at the height at which the maximum occurs as a way to assign more importance to this variable.

As for the temporal correlation improvements, in the majority of solutions found they are related mostly to lowered exponents for the thermodynamic variables (relative humidity and MPI). Indeed, by inspecting solutions (a) and (d), which are the ones with the highest temporal correlation for ERA5 and MERRA2, respectively, it appears that the algorithm is trying to remove these two variables from the equation by setting very low exponents for them and even taking relative humidity at 1 hPa. As at such a height, we expect there not to be any relative humidity, this indicates that, once the exponents are set to low enough values, this variable has little to no influence on the performance of the equation. This finding is in agreement with the work by Wang
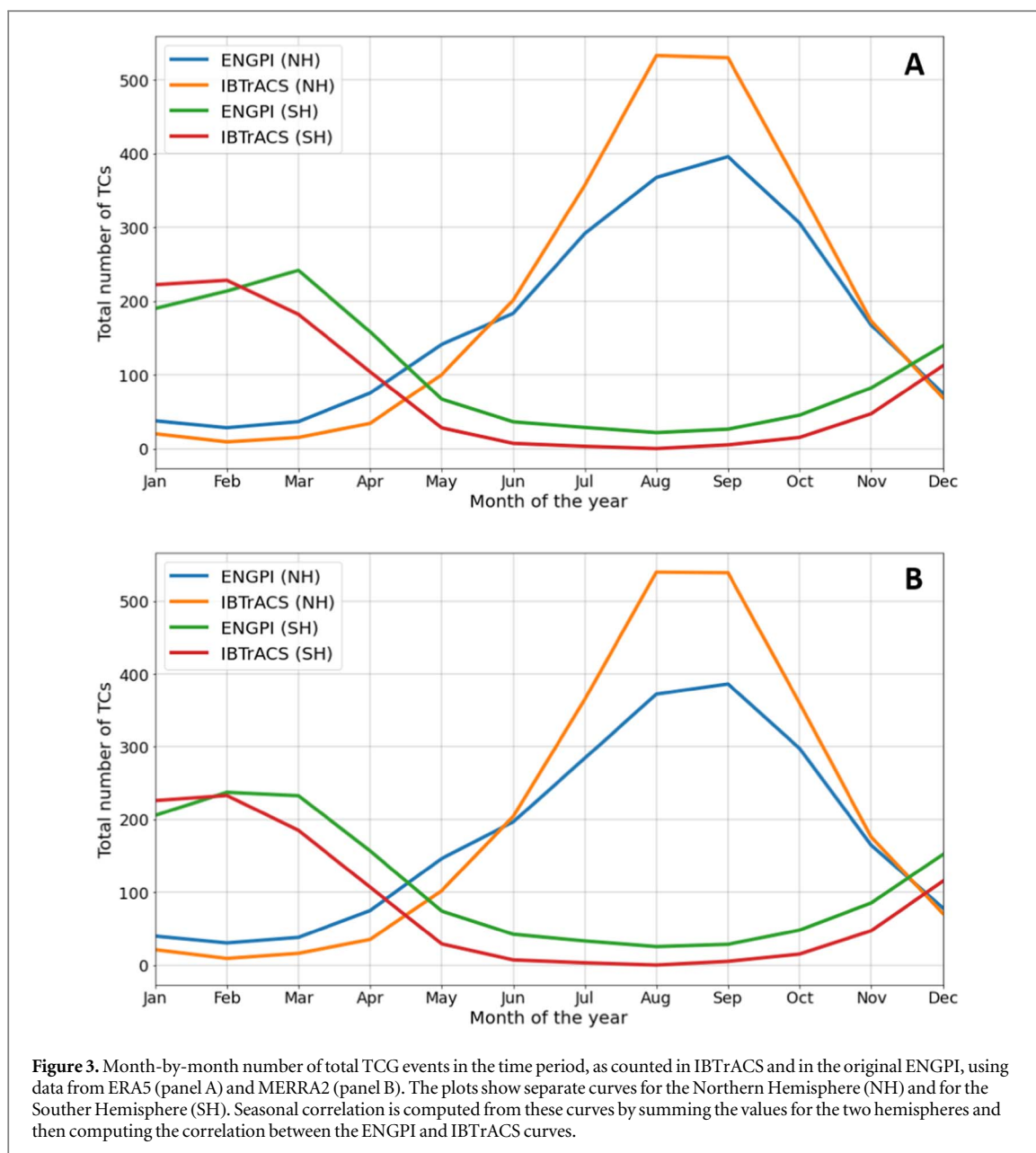
Figure 3. Month-by-month number of total TCG events in the time period, as counted in IBTrACS and in the original ENGPI, using data from ERA5 (panel A) and MERRA2 (panel B). The plots show separate curves for the Northern Hemisphere (NH) and for the Souther Hemisphere (SH). Seasonal correlation is computed from these curves by summing the values for the two hemispheres and then computing the correlation between the ENGPI and IBTrACS curves.

Table 1. Performance of the original ENGPI.

Evaluation metric	ERA5	MERRA2
$R_{spatial}$	0.616	0.635
$R_{seasonal}$	0.967	0.977
$R_{temporal}$	0.269	0.170

and Murakami [32], who use an input variable selection algorithm to find the best variables to construct a novel GI and found that no thermodynamic variables were selected. Furthermore, as indicated by Sharmila and Walsh. [52], relative humidity and MPI are not relevant predictors of the annual TCG variability in the western and eastern Pacific oceans. As these two basins contain more than 50% of all TCG events, it is plausible that a globally optimised index would attempt to ignore those two thermodynamic variables, as, on average, they help little to improve the temporal correlation as calculated here.

However, caution should be employed when interpreting solutions from either extreme of the Pareto fronts, as one objective may have been maximised at the expense of the physical meaning of the other. Indeed, the spatial distribution maps for solution (a) are considerably less realistic than the baseline (see figure 7), with a non-trivial likelihood of TCG being detected as far south as 40°S and as far north as in the Mediterranean Sea.
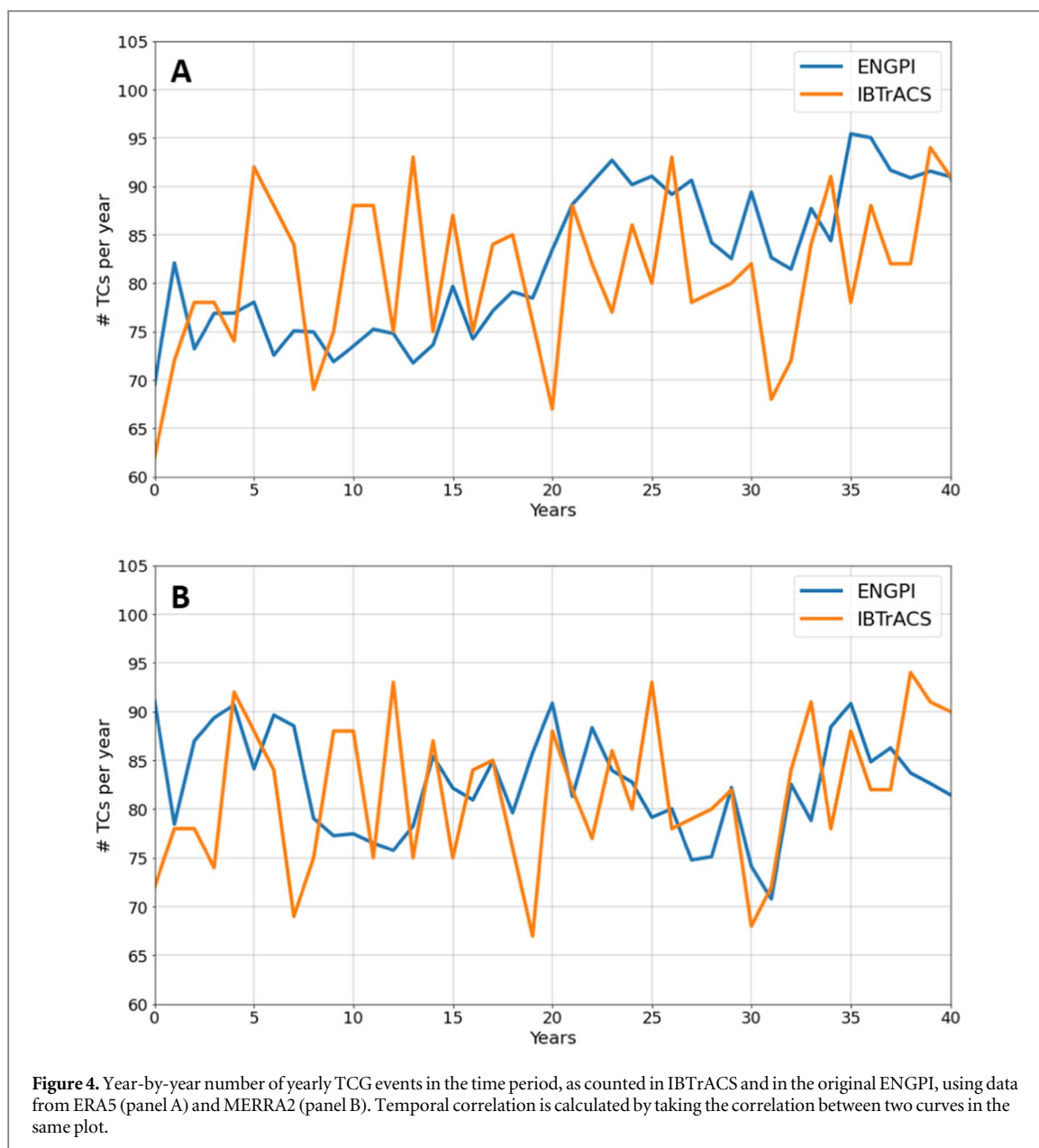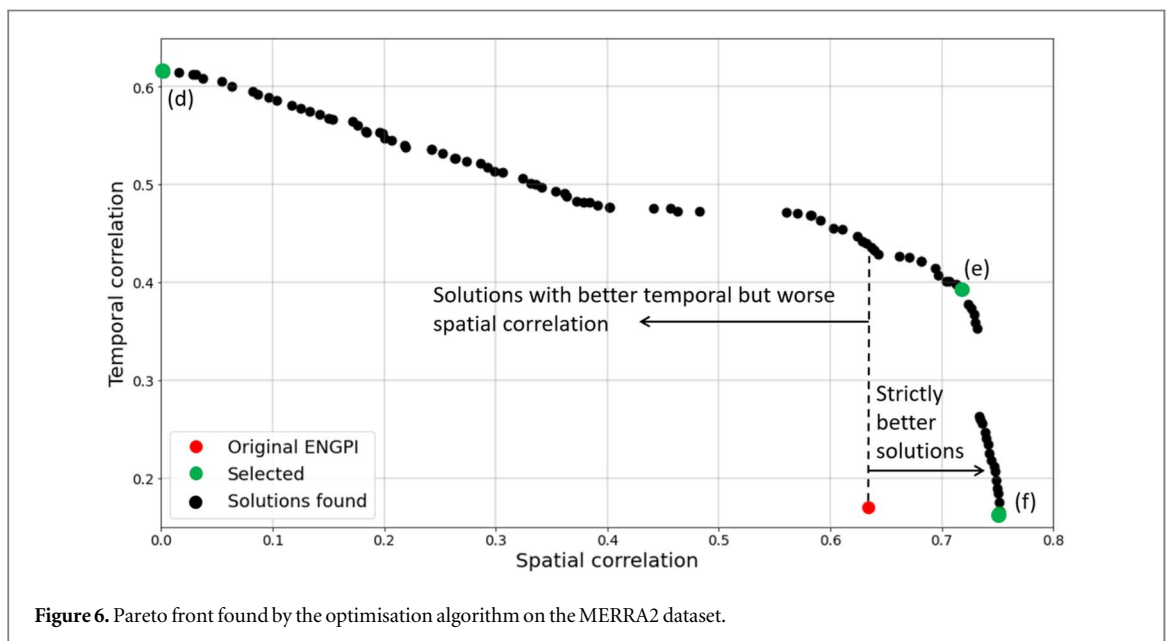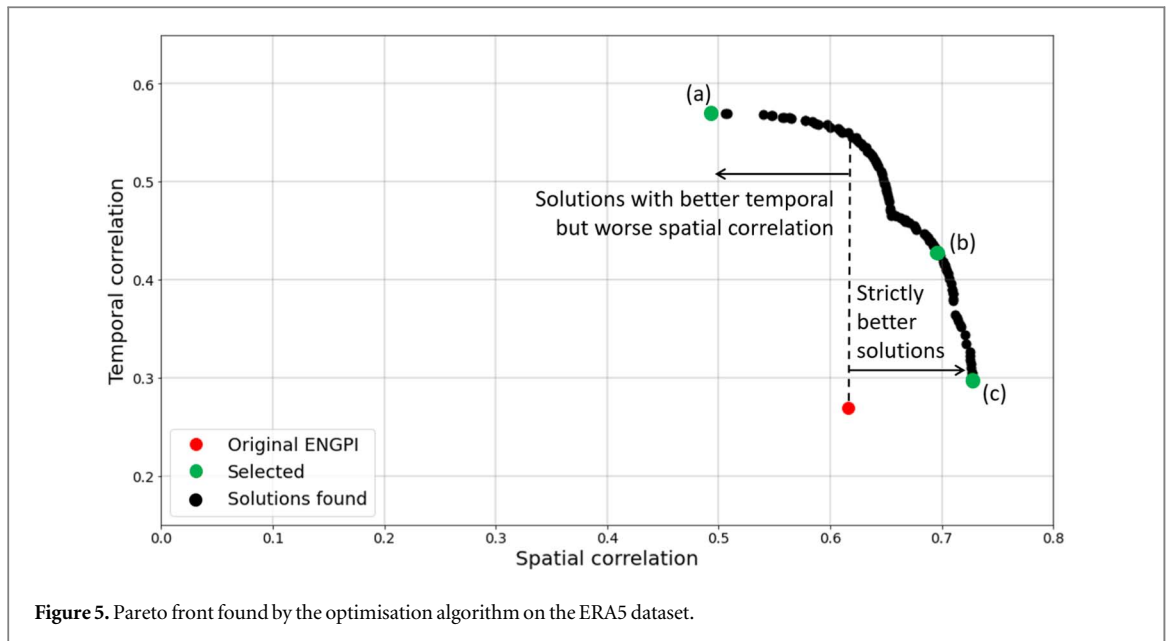
Figure 4. Year-by-year number of yearly TCG events in the time period, as counted in IBTrACS and in the original ENGPI, using data from ERA5 (panel A) and MERRA2 (panel B). Temporal correlation is calculated by taking the correlation between two curves in the same plot.

This indicates that the solutions at the tails of the Pareto fronts may have achieved high values in one objective function just by virtue of chance, but do not represent a physically realistic model of TCG. Furthermore, even though the temporal correlation in solution (a) is more than twice as higher than in the baseline, it still does not capture the true interannual variability of TCG (figure 8), potentially because the variables that truly explain the interannual variability of TCG are not present in the ENGPI. Furthermore, one would expect that a formulation of the ENGPI that captured the true underlying mechanisms of TCG would not need to optimise one objective at the expense of the other. Therefore, it is possible that the solutions in the middle of the Pareto front, where both objectives are optimised simultaneously, are closer to the true physical model of TCG.

Another observation we derive from table 2 is that solutions found by optimising on one dataset are almost equivalently valid when evaluated on the other. For example, figures 9 and 10 show how the spatial distribution maps and interannual variability curves of solution (b) (optimised on ERA5) are similar when evaluated on ERA5 and MERRA2. Also, the main difference between solutions found in the two datasets seems to be that solutions in MERRA2 tend to select variables at higher altitudes than solutions in ERA5. Two striking similarities are that exponents for vertical wind shear are almost always selected to be in the range $[-3, -2.4]$, and those for absolute vorticity in the vicinity of 2.

Finally, especially in the neighbourhood of solutions (b) and (e), there exist many formulations of the ENGPI that lead to almost equivalent performance (figures 5 and 6). The fact that no single solution sticks out as the clear global optimum for both objective functions indicates that the formula may be over-parameterised. Therefore, it may be beneficial to extend the proposed framework to GIs composed of different input variables

Figure 5. Pareto front found by the optimisation algorithm on the ERA5 dataset.

Figure 6. Pareto front found by the optimisation algorithm on the MERRA2 dataset.

from the one in the ENGPI, potentially even letting an input variable selection algorithm choose which variables to include.

4. Conclusions

In this paper, we presented a framework to optimise the parameters of tropical cyclone genesis indices with respect to different evaluation metrics. Rather than resulting in a single expression of the index, which is arbitrarily selected according to a chosen evaluation metric and whose optimisation is sensitive to the training data, this framework provides as an output a family of equally optimal solutions, together with an assessment of the skill of each of the family members on the different evaluation metrics. Therefore, the users of the index can select their preferred trade-off between the spatial and temporal correlation metrics.

As an application of this framework, the reported results show that the framework can substantially improve the performance of the Emanual and Nolan Genesis Potential Index (ENGPI) performance in terms of both spatial and temporal correlation. Selecting a solution with a balanced improvement of the two considered performance metrics, we show that one of the most notable improvements over the baseline ENGPI is due to the selection of absolute vorticity at 600 instead of 850 hPa, which leads to improvements in spatial correlation of
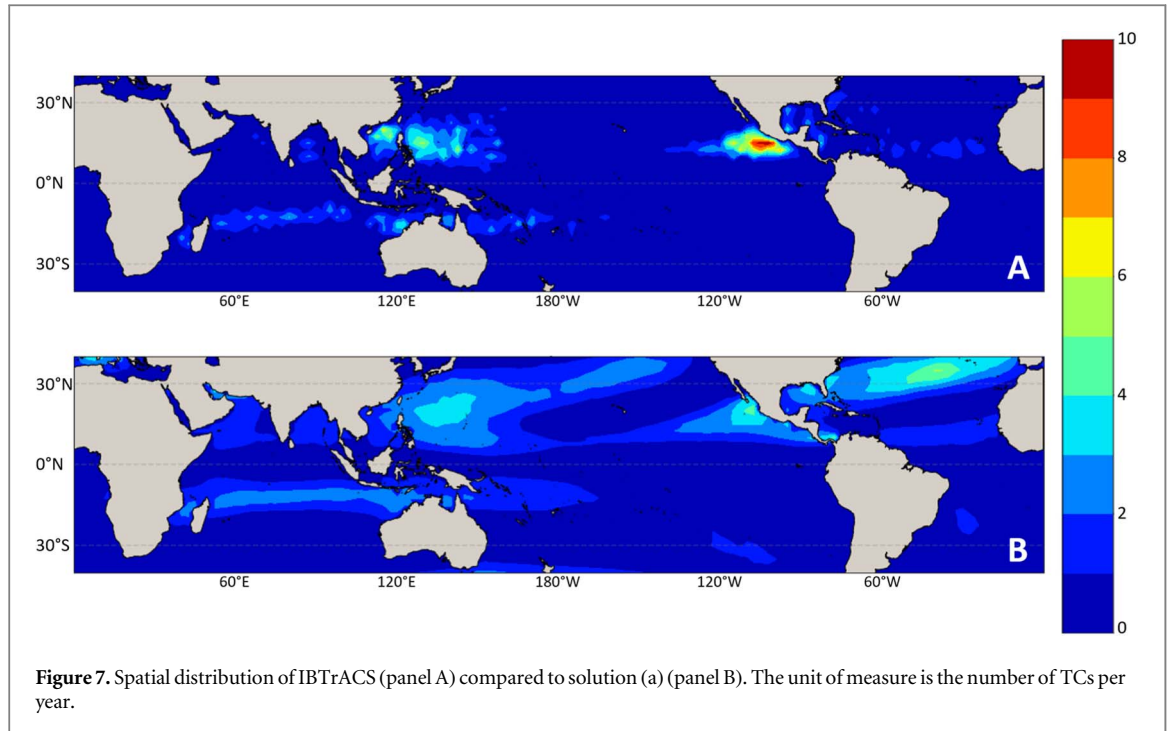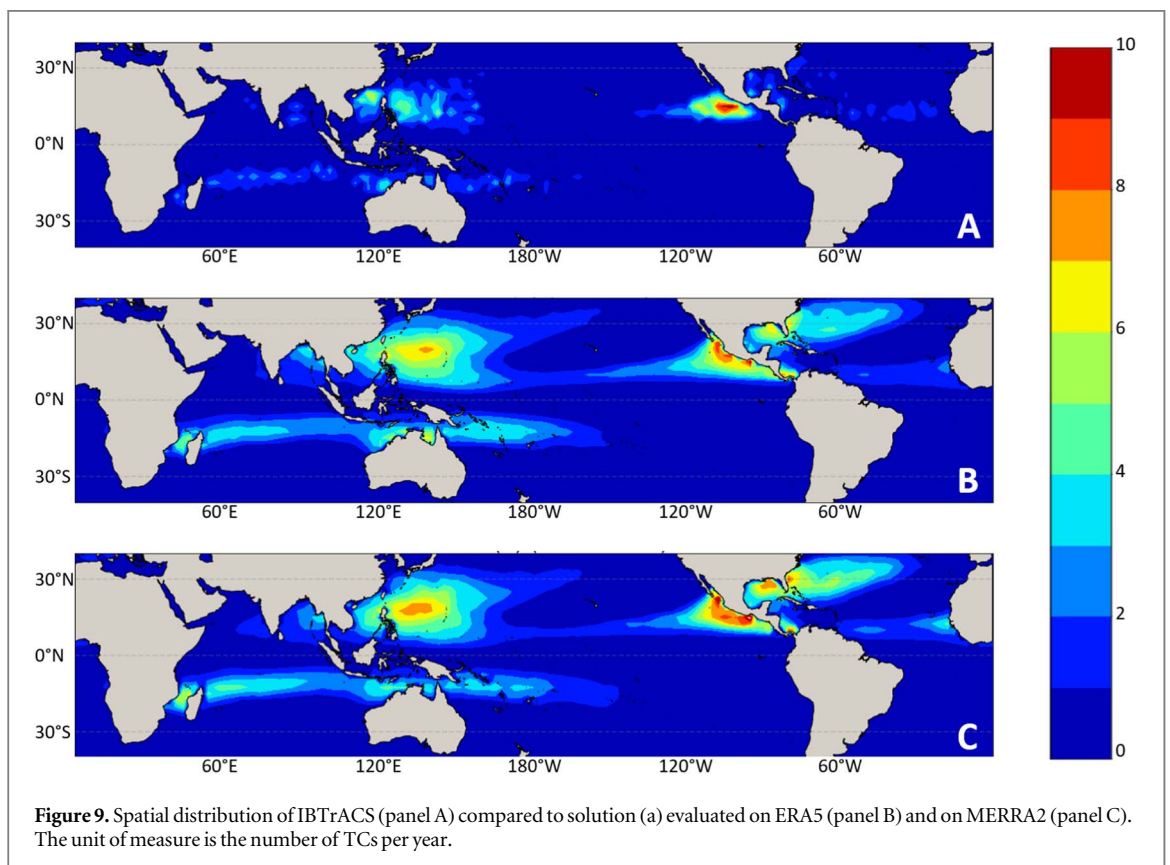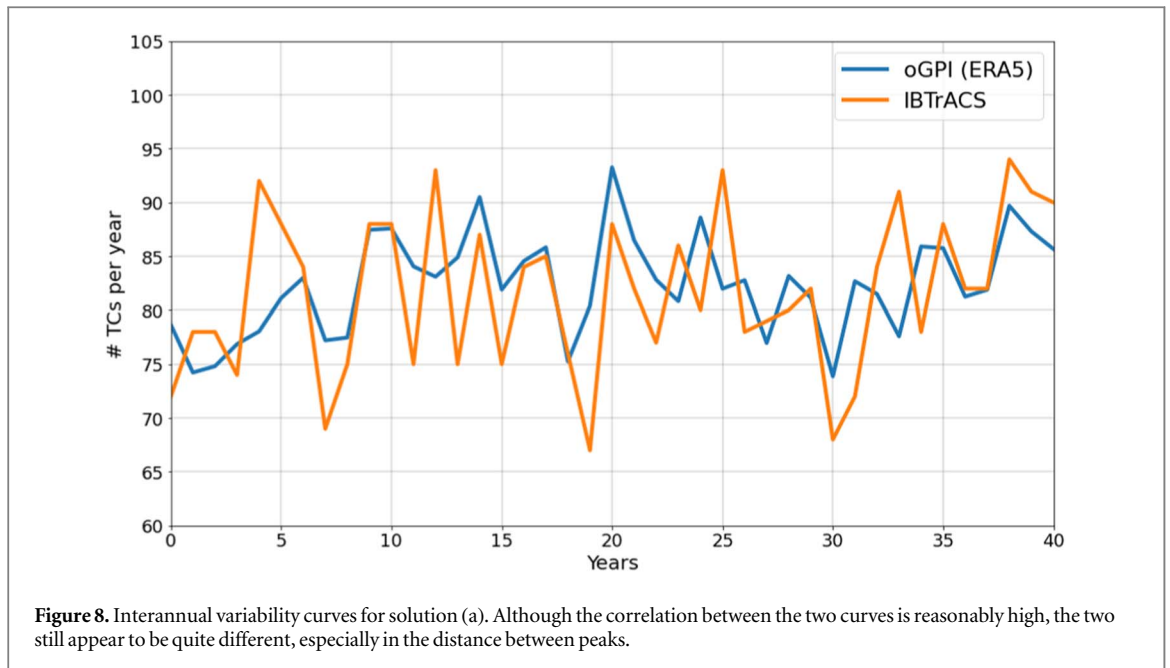
Figure 7. Spatial distribution of IBTrACS (panel A) compared to solution (a) (panel B). The unit of measure is the number of TCs per year.

Table 2. Performance of the solutions selected from figures 5 and 6. Each horizontal block of the table reports a solution, with the corresponding equation at the bottom of the block. The second column indicates on which dataset the solution was optimised, whereas columns four and five indicate on which dataset the reported evaluation metrics were calculated.

			Evaluated on	
Solution	Optimised on	Evaluation metric	ERA5	MERRA2
(a)	ERA5	$R_{spatial}$	0.493	0.500
		$R_{seasonal}$	0.978	0.978
		$R_{temporal}$	0.570	0.310
$oGPI(a) = (91.0 + 20.0~V_{S,850-250})^{-3.0} \frac{RH_{(1)}^{0.4}}{37.4} \frac{MPI^{0.3}}{54.0} \|\eta_{600}~10^5\|^{1.3} e^{2.5}$				
(b)	ERA5	$R_{spatial}$	0.696	0.700
		$R_{seasonal}$	0.990	0.985
		$R_{temporal}$	0.430	0.260
$oGPI(b) = (98.9 + 23.5~V_{S,850-250})^{-2.9} \frac{RH_{700}^{2.3}}{43.7} \frac{MPI^{1.8}}{68.4} \|\eta_{600}~10^5\|^{2.0} e^{3.0}$				
(c)	ERA5	$R_{spatial}$	0.727	0.741
		$R_{seasonal}$	0.988	0.985
		$R_{temporal}$	0.219	0.055
$oGPI(c) = (91.2 + 44.3~V_{S,600-250})^{-2.9} \frac{RH_{600}^{3.0}}{70.6} \frac{MPI^{2.6}}{71.6} \|\eta_{500}~10^5\|^{2.1} e^{2.9}$				
(d)	MERRA2	$R_{spatial}$	0.046	−0.005
		$R_{seasonal}$	0.483	0.05
		$R_{temporal}$	0.391	0.617
$oGPI(d) = (14.7 + 99.2~V_{S,250-50})^{0.6} \frac{RH_{1000}^{3.0}}{77.2} \frac{MPI^{0.6}}{99.3} \|\eta_{1000}~10^5\|^{2.1} e^{2.1}$				
(e)	MERRA2	$R_{spatial}$	0.688	0.718
		$R_{seasonal}$	0.961	0.966
		$R_{temporal}$	0.279	0.393
$oGPI(e) = (15.9 + 28.4~V_{S,500-200})^{-2.5} \frac{RH_{100}^{1.9}}{75.5} \frac{MPI^{0.6}}{88.6} \|\eta_{600}~10^5\|^{2.0} e^{1.9}$				
(f)	MERRA2	$R_{spatial}$	0.713	0.752
		$R_{seasonal}$	0.989	0.984
		$R_{temporal}$	0.168	0.165
$oGPI(f) = (20.3 + 18.3~V_{S,500-250})^{-2.4} \frac{RH_{600}^{3.0}}{75.8} \frac{MPI^{2.7}}{94.8} \|\eta_{600}~10^5\|^{2.1} e^{1.7}$				

about 8%. Also, the results suggest that the two thermodynamic variables (relative humidity and MPI) hinder the ability of the index to capture the interannual variability of TCG, and that lowering their weight in the equation by assigning small exponents to them may lead to values of TC interannual variability closer to the observations.

Figure 8. Interannual variability curves for solution (a). Although the correlation between the two curves is reasonably high, the two still appear to be quite different, especially in the distance between peaks.

Figure 9. Spatial distribution of IBTrACS (panel A) compared to solution (a) evaluated on ERA5 (panel B) and on MERRA2 (panel C). The unit of measure is the number of TCs per year.

The flexibility of the framework allows for tests similar to the one carried here to be repeated under different conditions: using different data, different input variables, different baseline indices, and different optimisation algorithms altogether. Future work should aim to extend the proposed framework to these different settings. Furthermore, future work should explore the option of selecting new inputs for the ENGPI by using an input variable selection algorithm.
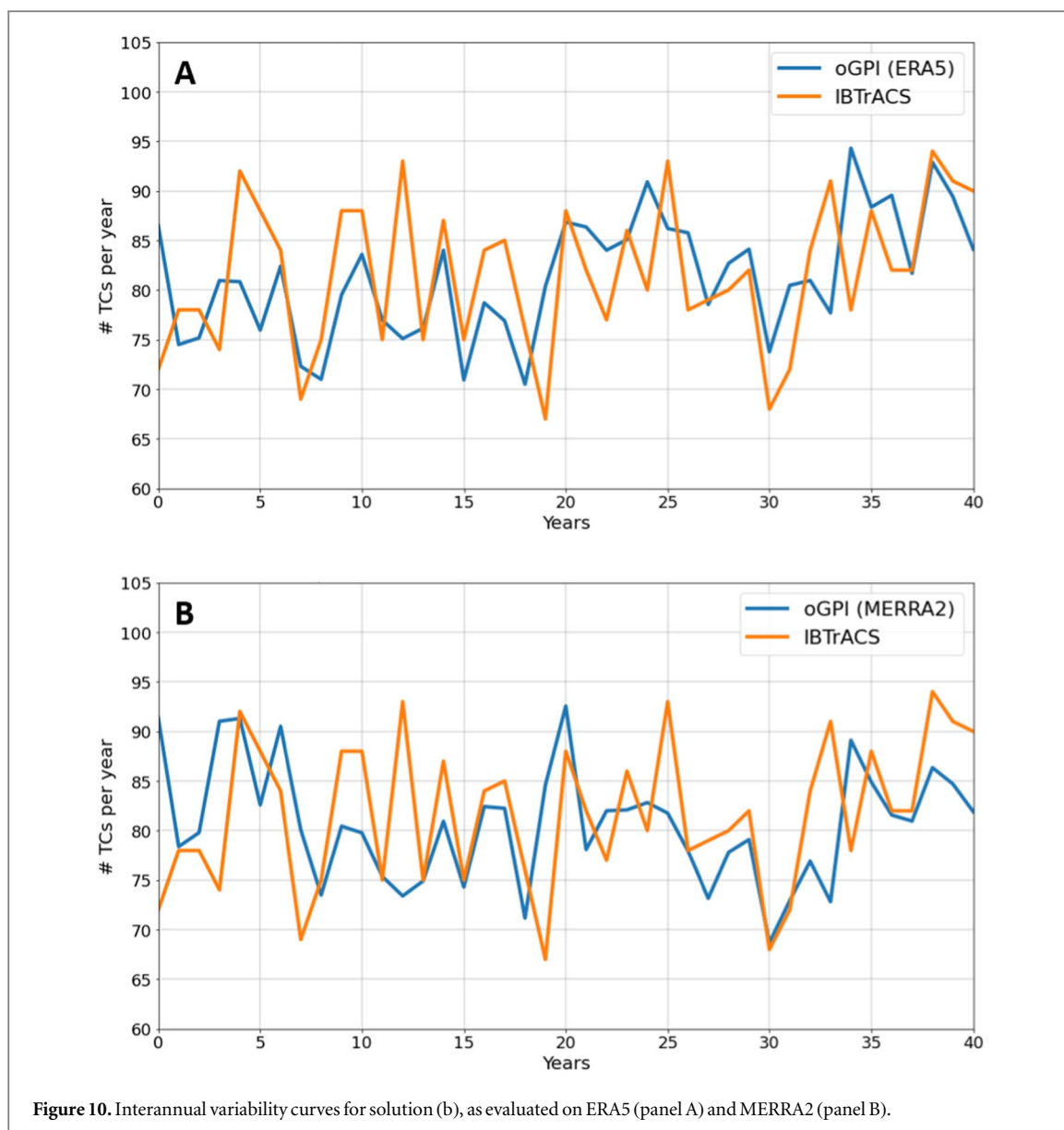
Figure 10. Interannual variability curves for solution (b), as evaluated on ERA5 (panel A) and MERRA2 (panel B).

Data availability statement

The data that support the findings of this study are openly available at the following URLs/DOIs:
doi: https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset
https://disc.gsfc.nasa.gov/datasets?project=MERRA-2.

ORCID iDs

Guido Ascenso ● https://orcid.org/0000-0002-9050-9892
Leone Cavicchia ● https://orcid.org/0000-0001-9857-7817
Enrico Scoccimarro ● https://orcid.org/0000-0001-7987-4744
Andrea Castelletti ● https://orcid.org/0000-0002-7923-1498

References

[1] Mendelsohn R, Emanuel K, Chonabayashi S and Bakkensen L 2012 The impact of climate change on global tropical cyclone damage *Nat. Clim. Change* **2** 205–9

[2] Robertson A W, Vitart F and Camargo S J 2020 Subseasonal to seasonal prediction of weather to climate with application to tropical cyclones *J. Geophysical Research: Atmospheres* **125** e2018JD029375

[3] Befort D J, Hodges K I and Weisheimer A 2022 Seasonal prediction of tropical cyclones over the North Atlantic and Western North Pacific *J. Clim.* **35** 1385–97

[4] Sobel A H *et al* 2021 Tropical cyclone frequency *Earth's Future* **9** e2021EF002275

[5] Tippett M K, Camargo S J and Sobel A H 2011 A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis *J. Clim.* **24** 2335–57

[6] Vecchi G A *et al* 2014 On the seasonal forecasting of regional tropical cyclone activity *J. Clim.* **27** 7994–8016

[7] Murakami H and Wang B 2010 Future change of North Atlantic tropical cyclone tracks: projection by a 20 km-mesh global atmospheric model *J. Clim.* **23** 2699–721

[8] Murakami H *et al* 2012 Future changes in tropical cyclone activity projected by the new high-resolution MRI-AGCM *J. Clim.* **25** 3237–60

[9] Zhao S and Suzuki K 2019 Differing impacts of black carbon and sulfate aerosols on global precipitation and the ITCZ location via atmosphere and ocean energy perturbations *J. Clim.* **32** 5567–82

[10] Camargo S J, Sobel A H, Barnston A G and Emanuel K A 2007 Tropical cyclone genesis potential index in climate models *Tellus A* **59** 428–43

[11] Walsh K J *et al* 2015 Hurricanes and climate: the US CLIVAR working group on hurricanes *Bull. Am. Meteorol. Soc.* **96** 997–1017

[12] Cavicchia L, Ascenso G, Scoccimarro E, Castelletti A, Giuliani M and Gualdi S 2022 Tropical cyclone genesis potential in CMIP6 climate models *Copernicus Meetings*

[13] Roberts M J *et al* 2020 Impact of model resolution on tropical cyclone simulation using the HighResMIP-PRIMAVERA multimodel ensemble *J. Clim.* **33** 2557–83

[14] Camargo S J and Wing A A 2016 Tropical cyclones in climate models *Wiley Interdiscip. Rev. Clim. Change* **7** 211–37

[15] Bruyère C L, Holland G J and Towler E 2012 Investigating the use of a genesis potential index for tropical cyclones in the North Atlantic basin *J. Clim.* **25** 8611–26

[16] Yu J Y, Hsiao L P and Chiu P G 2018 Evaluating the Emanuel-Nolan genesis potential index: contrast between North Atlantic and western North Pacific *Terrestrial, Atmospheric & Oceanic Sciences* **29** (2) 201–214

[17] Knutson T R, Sirutis J J, Garner S T, Held I M and Tuleya R E 2007 Simulation of the recent multidecadal increase of Atlantic hurricane activity using an 18 km-grid regional model *Bull. Am. Meteorol. Soc.* **88** 1549–65

[18] Scoccimarro E 2016 Modeling tropical cyclones in a changing climate *Oxford Research Encyclopedia of Natural Hazard Science* (https://doi.org/10.1093/acrefore/9780199389407.013.22)

[19] Roberts M J *et al* 2020 Projected future changes in tropical cyclones using the CMIP6 HighResMIP multimodel ensemble *Geophys. Res. Lett.* **47** e2020GL088662

[20] Knutson T *et al* 2020 Tropical cyclones and climate change assessment: part II: projected response to anthropogenic warming *Bull. Am. Meteorol. Soc.* **101** E303–22

[21] Kossin J P 2018 A global slowdown of tropical-cyclone translation speed *Nature* **558** 104–7

[22] Lanzante J R 2019 Uncertainties in tropical-cyclone translation speed *Nature* **570** E6–15

[23] Wright D B, Knutson T R and Smith J A 2015 Regional climate model projections of rainfall from US landfalling tropical cyclones *Clim. Dyn.* **45** 3365–79

[24] Knutson T R *et al* 2015 Global projections of intense tropical cyclone activity for the late twenty-first century from dynamical downscaling of CMIP5/RCP4. 5 scenarios *J. Clim.* **28** 7203–24

[25] Wang C C, Lin B X, Chen C T and Lo S H 2015 Quantifying the effects of long-term climate change on tropical cyclone rainfall using a cloud-resolving model: Examples of two landfall typhoons in Taiwan *J. Clim.* **28** 66–85

[26] Gettelman A, Bresch D N, Chen C C, Truesdale J E and Bacmeister J T 2018 Projections of future tropical cyclone damage with a high-resolution global climate model *Clim. Change* **146** 575–85

[27] Camargo S J, Emanuel K A and Sobel A H 2007 Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis *J. Clim.* **20** 4819–34

[28] Camargo S J, Wheeler M C and Sobel A H 2009 Diagnosis of the MJO modulation of tropical cyclogenesis using an empirical index *J. Atmos. Sci.* **66** 3061–74

[29] Tsuboi A and Takemi T 2014 The interannual relationship between MJO activity and tropical cyclone genesis in the Indian Ocean *Geoscience Letters* **1** 1–6

[30] Menkes C E *et al* 2012 Comparison of tropical cyclogenesis indices on seasonal to interannual timescales *Clim. Dyn.* **38** 301–21

[31] Shin J, Song C, Kim S and Park S 2019 A heuristic estimation of the genesis probability of tropical cyclones using genesis frequency and genesis potential index *Journal of the Korean Earth Science Society* **40** 561–71

[32] Wang B and Murakami H 2020 Dynamic genesis potential index for diagnosing present-day and future global tropical cyclone genesis *Environ. Res. Lett.* **15** 114008

[33] Camargo S J, Tippett M K, Sobel A H, Vecchi G A and Zhao M 2014 Testing the performance of tropical cyclone genesis indices in future climates using the HiRAM model *J. Clim.* **27** 9171–96

[34] Kim J H, Brown S J and McDonald R E 2011 Future changes in tropical cyclone genesis in fully dynamic ocean-and mixed layer ocean-coupled climate models: a low-resolution model study *Clim. Dyn.* **37** 737–58

[35] Emanuel K and Nolan D S 2004 Tropical cyclone activity and the global climate system *26th Conference on Hurricanes and Tropical Meteorolgy*

[36] Gray W M 1979 Hurricanes: their formation, structure and likely role in the tropical circulation *Meteorology over the Tropical Oceans* **155** 218

[37] Song Y, Wang L, Lei X and Wang X 2015 Tropical cyclone genesis potential index over the western North Pacific simulated by CMIP5 models *Adv. Atmos. Sci.* **32** 1539–50

[38] Zhang W, Gao J, Cheung K K, Chen J M and You L 2020 Intraseasonal variability of the genesis potential index and its relationship with tropical cyclogenesis over the western North Pacific *Int. J. Climatol.* **40** 22–35

[39] Emanuel K A 1995 Sensitivity of tropical cyclones to surface exchange coefficients and a revised steady-state model incorporating eye dynamics *J. Atmospheric Sciences* **52** 3969–76

[40] Bister M and Emanuel K A 1998 Dissipative heating and hurricane intensity *Meteorol. Atmos. Phys.* **65** 233–40

[41] Knapp K R, Kruk M C, Levinson D H, Diamond H J and Neumann C J 2010 The international best track archive for climate stewardship (IBTrACS) unifying tropical cyclone data *Bull. Am. Meteorol. Soc.* **91** 363–76

[42] Schreck C J III, Knapp K R and Kossin J P 2014 The impact of best track discrepancies on global tropical cyclone climatologies using IBTrACS *Mon. Weather Rev.* **142** 3881–99

[43] Hersbach H *et al* 2019 ERA5 monthly averaged data on single levels from 1959 to present *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)* **10** 252–66

[44] Reichle R H, Liu Q, Koster R D, Draper C S, Mahanama S P and Partyka G S 2017 Land surface precipitation in MERRA-2 *J. Clim.* **30** 1643–64

[45] Deb K, Pratap A, Agarwal S and Meyarivan T 2002 A fast and elitist multiobjective genetic algorithm: NSGA-II *IEEE Trans. Evol. Comput.* **6** 182–97

[46] Tang K S, Man K F, Kwong S and He Q 1996 Genetic algorithms and their applications *IEEE Signal Process Mag.* **13** 22–37

[47] Khare V, Yao X and Deb K 2003 Performance scaling of multi-objective evolutionary algorithms *International Conference on Evolutionary Multi-criterion Optimization, Springer* 376–90

[48] Purshouse R C and Fleming P J 2003 Evolutionary many-objective optimisation: an exploratory analysis *The 2003 Congress on Evolutionary Computation, 2003. CEC'03. IEEE* **vol. 3** 2066–73

[49] Purshouse R C and Fleming P J 2007 On the evolutionary optimization of many conflicting objectives *IEEE Trans. Evol. Comput.* **11** 770–84

[50] Wang Y 2008 Structure and formation of an annular hurricane simulated in a fully compressible, nonhydrostatic model—TCM4 *J. Atmos. Sci.* **65** 1505–27

[51] Murthy V S and Boos W R 2019 Understanding the vertical structure of potential vorticity in tropical depressions *Q. J. R. Meteorolog. Soc.* **145** 1968–91

[52] Sharmila S and Walsh K 2017 Impact of large-scale dynamic versus thermodynamic climate conditions on contrasting tropical cyclone genesis frequency *J. Clim.* **30** 8865–83