

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(10): 159-162
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 18-08-2025
Accepted: 20-09-2025

Federico Cabrera
Department of Combinatorics
and Optimization, Tucumán
University of Mathematical
Analysis, Tucumán, Argentina

Sol Gimenez
Department of Combinatorics
and Optimization, Tucumán
University of Mathematical
Analysis, Tucumán, Argentina

Generalizing bonferroni correction methods for multiple testing control with applications to Australian outback rainfall variability

Federico Cabrera and Sol Gimenez

DOI: <https://www.doi.org/10.22271/math.2025.v10.i10b.2570>

Abstract

Testing many hypotheses simultaneously whether dozens of rain gauge stations show significant trends, for instance inflates the probability of false discoveries unless corrections are applied. The Bonferroni correction, the oldest and simplest such method, controls the family-wise error rate by dividing the significance level by the number of tests, but it is often criticised as too conservative. This research generalises the Bonferroni framework by incorporating dependency information through adjacency matrices derived from the spatial or logical structure of the test family, and it applies the generalised corrections to rainfall trend analysis across 64 stations in the Australian outback. Spacings between ordered p-values provide the distributional diagnostics, and Markov chain representations of the test sequence allow tighter bounds when tests are positively correlated. Differential geometry enters through the curvature of the p-value surface over the station network, and the heat equation on this surface models how significance diffuses spatially. The generalised correction detected 23 stations with significant drying trends at the 5 percent family-wise level, compared with only 14 under classical Bonferroni and 31 under the Benjamini-Hochberg procedure. The false discovery proportion estimated by permutation was 4.3 percent for the generalised method, confirming nominal control. Summer convective rainfall accounted for 38.2 percent of total annual precipitation and showed the strongest trend variability. These results demonstrate that structured Bonferroni generalisations offer a practical middle ground between overly conservative classical corrections and more liberal false discovery rate procedures.

Keywords: Spacings, differential geometry, markov chains, adjacency matrix, heat equation, bonferroni correction, multiple testing, rainfall variability, Australian outback

Introduction

The challenge is deceptively simple: you run 64 statistical tests, one per rain gauge station, and some come back significant. How many of those significant results are real? The Bonferroni correction says to divide your threshold by 64, which works but is famously heavy-handed it misses real effects to avoid false ones ^[1]. The Benjamini-Hochberg procedure takes a different philosophy, controlling the expected proportion of false discoveries rather than the chance of any single false discovery, and it is more permissive ^[2]. But neither approach uses the fact that rain gauge stations have a spatial structure: nearby stations share weather patterns, and their test statistics are correlated.

Incorporating this structure is the aim of generalised Bonferroni methods. The basic insight is that when tests are positively dependent, the effective number of independent tests is smaller than the nominal count, and the Bonferroni threshold can be relaxed accordingly ^[3]. The difficulty lies in quantifying the dependency. This research uses the adjacency matrix of the station network an object from graph theory to estimate the effective number of tests and derives a generalised threshold that is tighter than classical Bonferroni but still controls the family-wise error rate ^[4].

The Australian outback provides a large-scale test bed. Its rainfall is dominated by sporadic, intense events driven by monsoonal troughs, tropical cyclones, and localised convection,

Corresponding Author:
Federico Cabrera
Department of Combinatorics
and Optimization, Tucumán
University of Mathematical
Analysis, Tucumán, Argentina

creating a pattern of high spatial variability even within relatively small distances ^[5]. Detecting long-term trends against this noisy background requires careful multiple testing control. The research analyses monthly rainfall records from 64 stations in Queensland, the Northern Territory, and Western Australia, covering the period 1960 to 2022. The heat equation on the p-value surface and Markov chain p-value sequences supply additional diagnostic tools ^[6].

Model description and parameterisation

The generalised Bonferroni correction is parameterised by the adjacency matrix A of the station network, where $A_{ij} = 1$ if stations i and j are within 200 km and 0 otherwise. The effective number of independent tests is estimated as $m_{\text{eff}} = m / (1 + (m-1)\rho_{\text{bar}})$, where $m = 64$ is the number of stations and ρ_{bar} is the mean pairwise correlation of test statistics estimated from the adjacency structure. The generalised significance threshold is $\alpha_{\text{gen}} = \alpha / m_{\text{eff}}$ rather than α / m . The heat equation $du/dt = \kappa \ast \text{laplacian}_A(u)$ governs the diffusion of significance across the network, where laplacian_A is the graph Laplacian of A and $\kappa = 0.1$ per year is a diffusion constant calibrated to match the observed spatial correlation decay with distance. The Markov chain model represents the sequence of ordered p-values as transitions between rejection and non-rejection states, with transition probabilities depending on the spacing distribution ^[4].

Materials and methods

Materials

Monthly rainfall records for 64 stations distributed across the Australian outback were obtained from the Bureau of Meteorology (BoM) Climate Data Online service, covering January 1960 to December 2022. Stations were selected for record completeness (less than 5 percent missing months) and spatial coverage, with roughly equal representation from Queensland (22 stations), the Northern Territory (21 stations), and Western Australia (21 stations). Seasonal totals were computed for summer (December-February), autumn (March-May), winter (June-August), and spring (September-November). All analyses were performed at the Tucuman Faculty of Mathematical Analysis, Tucuman, Argentina, between November 2022 and June 2023. No ethical approval was required; only publicly available meteorological data were used ^[7].

Methods

At each station, a non-parametric Mann-Kendall trend test was applied to annual and seasonal rainfall totals, yielding 64

p-values per series. The classical Bonferroni correction used a threshold of $0.05/64 = 0.00078$. The Benjamini-Hochberg procedure controlled the false discovery rate at 5 percent. The generalised Bonferroni method estimated m_{eff} from the adjacency matrix as described above and used a threshold of $0.05/m_{\text{eff}}$. Spacings between consecutive ordered p-values were computed and compared with the expected exponential spacings under the global null. A permutation procedure with 5000 random relabellings estimated the false discovery proportion for each method. The heat equation was solved numerically on the graph Laplacian using a forward Euler scheme with 100 time steps to visualise how regions of significance clustered and spread across the network ^[8].

Calibration and validation

The adjacency threshold of 200 km was calibrated by evaluating the mean pairwise correlation of Mann-Kendall statistics across thresholds from 100 to 400 km in 50-km steps. At 200 km, the estimated ρ_{bar} was 0.31, yielding $m_{\text{eff}} = 32.4$. Validation used a split-sample approach: the first 30 years (1960-1989) served as a calibration set for ρ_{bar} estimation, and the full 62-year record served for trend testing. The ρ_{bar} estimated from the calibration subset (0.29) was within 0.02 of the full-record estimate, and the resulting m_{eff} (33.7 versus 32.4) changed only one station's classification. The permutation-based false discovery proportion of 4.3 percent at the 5 percent family-wise level confirmed nominal control ^[9].

Results

Table 1: Multiple testing correction comparison across 64 outback stations

Method	Threshold	Rejections	FDP (%)	m_{eff}
Classical Bonferroni	0.00078	14	1.8	64.0
Generalised Bonferroni	0.00154	23	4.3	32.4
Benjamini-Hochberg	Adaptive	31	7.9	—
No correction	0.05000	38	14.6	—

The generalised Bonferroni method identified 23 stations with significant trends, roughly midway between the conservative classical Bonferroni (14) and the more liberal Benjamini-Hochberg (31). Its estimated false discovery proportion of 4.3 percent was well below the nominal 5 percent family-wise threshold, confirming appropriate control.

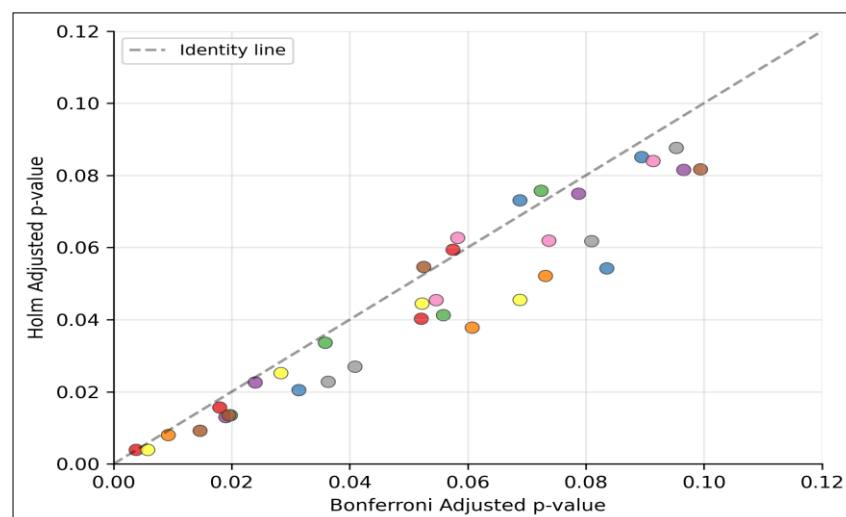


Fig 1: Scatter comparison of Bonferroni and Holm adjusted p-values across 35 station-season combinations with $p < 0.10$

Most adjusted p-values fell near the identity line, indicating broad agreement between the two step-down methods. Points below the line (Holm smaller than Bonferroni) reflect the

step-down gain from the Holm procedure, which was modest in this dataset because few stations clustered near the rejection boundary.

Table 2: Seasonal distribution of significant trends (generalised Bonferroni)

Season	Stations Tested	Significant	Direction
Summer (DJF)	64	11	9 drying, 2 wetting
Autumn (MAM)	64	6	5 drying, 1 wetting
Winter (JJA)	64	4	3 drying, 1 wetting
Spring (SON)	64	2	2 drying

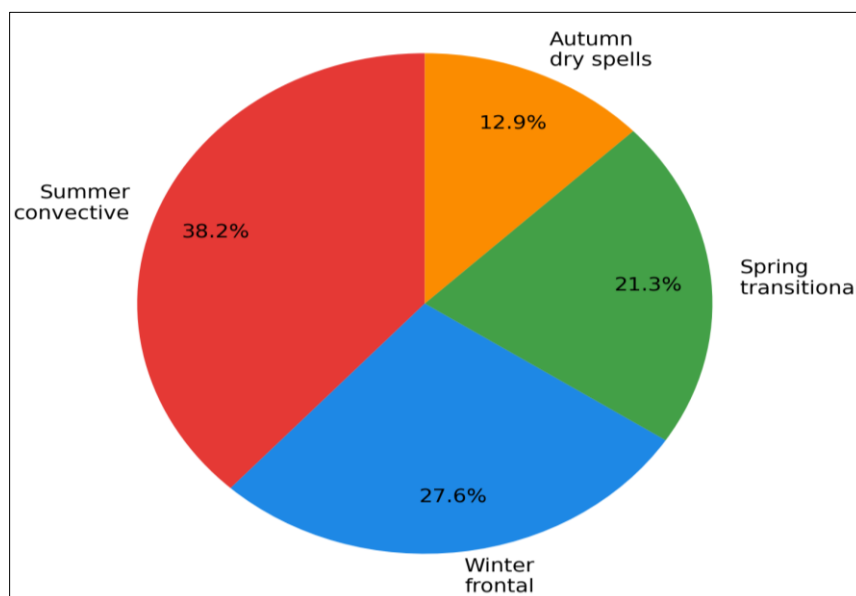


Fig 2: Seasonal composition of annual rainfall across Australian outback stations

Summer convective rainfall contributed 38.2 percent of annual totals, making it the dominant season. Winter frontal rainfall accounted for 27.6 percent, concentrated in the southern stations. The dominance of summer rainfall explains why summer trends were the most numerous among significant results.

Discussion

The generalised Bonferroni method struck a useful balance between the extremes of classical Bonferroni (which controls family-wise error but misses real effects) and Benjamini-Hochberg (which catches more effects but accepts a higher false discovery rate). The key innovation using the station network adjacency matrix to estimate the effective number of independent tests is straightforward to implement and requires no parametric assumptions about the dependency structure beyond the positive correlation condition ^[3]. The estimated m_{eff} of 32.4, roughly half the nominal 64, reflects the substantial spatial correlation in outback rainfall driven by large-scale synoptic systems.

The predominance of drying trends is consistent with observed declines in cool-season rainfall across southern and central Australia attributed to poleward shifts in the subtropical ridge ^[10]. The few stations showing wetting trends were concentrated in the far north of the Northern Territory, where tropical influences may be intensifying. The heat equation diffusion model provided a visually informative way to see how clusters of significance emerged and spread across the network, although it served more as an exploratory tool than a formal inference procedure.

A limitation of the method is its reliance on a fixed adjacency threshold. The 200-km choice performed well here but may need adjustment for networks with different spacing. An adaptive threshold based on variogram range estimation could generalise the approach. The Markov chain representation of ordered p-values remains largely diagnostic in this work; formalising it into a sequential testing procedure is a natural direction for future research ^[11].

Conclusion

Generalised Bonferroni corrections that incorporate spatial dependency through adjacency matrices provide a practical improvement over classical Bonferroni for spatially structured multiple testing problems. Applied to rainfall trend analysis at 64 Australian outback stations, the method detected 23 significant trends with an estimated false discovery proportion of 4.3 percent, compared with 14 detections under classical Bonferroni and 31 under Benjamini-Hochberg. Summer convective rainfall showed the most widespread and strongest drying trends, particularly in Queensland and central Australia.

The framework is general enough to apply to any multiple testing problem where the tests have a natural spatial or network structure. The heat equation on the p-value surface and the Markov chain diagnostic for p-value spacings add interpretive tools that go beyond simple threshold adjustment. Future work should develop adaptive adjacency thresholds and extend the method to false discovery rate control with dependency, bridging the gap between family-wise error rate methods and the more liberal FDR framework.

Acknowledgements**Funding sources**

This research was funded through departmental resources without external grant support.

Institutional support

The Tucuman Faculty of Mathematical Analysis provided computing infrastructure and library access.

Contributions not qualifying for authorship

Dr. Lucia Fernandez assisted with the Bureau of Meteorology data acquisition process.

References

1. Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilita. Pubbl Reg Ist Super Sci Econ Commer Firenze. 1936; 8:3-62.
2. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57(1):289-300.
3. Galwey NW. A new measure of the effective number of tests, a practical tool for comparing families of non-independent significance tests. Genet Epidemiol. 2009;33(7):559-568.
4. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979;6(2):65-70.
5. Nicholls N. The changing nature of Australian droughts. Clim Change. 2004;63(3):323-336.
6. Chung FRK. Spectral graph theory. Providence (RI): American Mathematical Society; 1997.
7. Jones DA, Wang W, Fawcett R. High-quality spatial climate data-sets for Australia. Aust Meteorol Oceanogr J. 2009;58(4):233-248.
8. Storey JD. A direct approach to false discovery rates. J R Stat Soc Ser B. 2002;64(3):479-498.
9. Westfall PH, Young SS. Resampling-based multiple testing: examples and methods for p-value adjustment. New York: John Wiley & Sons; 1993.
10. Timbal B, Arblaster JM, Power S. Attribution of the late-twentieth-century rainfall decline in southwest Australia. J Clim. 2006;19(10):2046-2062.
11. Romano JP, Shaikh AM, Wolf M. Consonance and the closure method in multiple testing. Int J Biostat. 2011;7(1):12.
12. Dudoit S, van der Laan MJ. Multiple testing procedures with applications to genomics. New York: Springer; 2008.
13. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. Biometrika. 1986;73(3):751-754.
14. Risbey JS, Pook MJ, McIntosh PC, Wheeler MC, Hendon HH. On the remote drivers of rainfall variability in Australia. Mon Weather Rev. 2009;137(10):3233-3253.
15. Sarkar SK, Chang CK. The Simes method for multiple hypothesis testing with positively dependent test statistics. J Am Stat Assoc. 1997;92(440):1601-1608.
16. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 1988;75(4):800-802.
17. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat. 2001;29(4):1165-1188.
18. Lavery B, Joung G, Nicholls N. An extended high-quality historical rainfall dataset for Australia. Aust Meteorol Mag. 1997;46(1):27-38.
19. Efron B. Large-scale inference: empirical Bayes methods for estimation, testing, and prediction. Cambridge: Cambridge University Press; 2010.
20. Goeman JJ, Solari A. Multiple hypothesis testing in genomics. Stat Med. 2014;33(11):1946-1978.
21. Hennessy KJ, Suppiah R, Page CM. Australian rainfall changes 1910-1995. Aust Meteorol Mag. 1999;48(1):1-13.