AMSI SUMMER RESEARCH SCHOLARSHIPS 2023–24

SET YOUR SIGHTS ON RESEARCH THIS SUMMER

Community Level Changes in First Flowering Day with Regard to Climate Extremes: Australian Millennium Drought

Dillon Batdorf

Supervised by Prof. Irene Hudson

RMIT University

AMSI AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Abstract

Climate change's impact on phenological patterns is a critical area of ecological research, especially concerning the timing of biological events such as the first flowering day (FFD) of plant species. This study investigates the FFD of 81 Australian flower species over 31 years to understand the effects of climate factors with a particular focus on the Millennium Drought period. Utilising time series clustering and Rank Biased Overlap (RBO) metric, the FFD data was analysed to identify changes in phenological patterns and stability of flowering sequences over time.

The time series clustering performed reveals distinct clusters, suggesting variability in FFD patterns that correspond to different climatic periods, including the significant Millennium Drought. The RBO analysis demonstrated moderate to high similarity in the flowering order across years, depending on whether the periods being compared were drought or non-drought periods.

The results implicate climate change in the alteration of FFD and suggest potential ecological consequences, such as challenges for agricultural management. This report also identifies limitations in the historical scope of the data, and advocates for the incorporation of a broader range of flower species and ecosystems, along with more recent data to allow more impactful future research, such as the production of predictive models. Overall, the findings made contribute to understanding how climate change affects plant phenology and underscore the need for adaptive strategies to combat climate change.

Introduction and Background

Phenology, as defined by Keatley and Hudson (2010), encompasses the study of life cycle phases in flora and fauna, including events such as the blossoming of plants, crop maturation, insect emergence and bird migration. This discipline examines the interplay between these phenological phenomena and climatic variables, investigating how the timing of these recurrent events is influenced by climatic changes (Hudson & Keatley, 2010; Keatley et al., 2013; Schwartz 2003). Temporal changes in phenology, primarily driven by global warming, are broadly acknowledged as the most significant indicator of worldwide environmental transformation (Hassan et al. 2023). Such phenological markers, notably the timing of the flowering and animal migration, exhibit significant sensitivity and vulnerability to climatic variations, including temperature, light intensity, precipitation, and humidity levels. Often, research in this field focuses on specific indicators, like the first flowering day (FFD), rather than capturing the full spectrum of phenological occurrences (CaraDonna et al. 2014).

Phenological milestones play a critical role in understanding ecological dynamics and biodiversity, prompting numerous studies on the subject. For instance, research highlighting a 250-year analysis of the initial bloom dates of 405 plant species in the UK revealed that, over the past 25 years, the onset of the first blooms advanced by 2.2-12.7 days compared to any similar period since 1760, correlating closely with average temperature increases. A temperature rise of one degree Celsius is associated with an earlier first flowering date by approximately five days (Amano 2010). Additionally, the average blooming time of three shrub species, as determined by the Spring Phenology Index (SPI) from 1963-2009, showed that the most recent 30-year SPI advanced by 2.1-6.3 days compared to any previous 30-year interval before 1970, according to an SPI time series reconstructed for 1834-2009 in eastern China. This SPI was found to correlate significantly with February to April temperatures, with a one-degree Celsius increase temperature advancing the SPI by 3.1 days (Ge 2014). Observation records from 1981 to 2017 for three species of lilac and honeysuckle indicated that the predictiveness of the spring index for vegetative and reproductive phenology varies across species and latitudes (Gerst 2020).

The FFD metric represents the day count January 1st until the observed first bloom of a species within a year. It is a critical observation in phenology, signifying a major phase in a plant's lifecycle that modulates its interaction with the surrounding environment (Ehrlén 2015; Keatley, Hudson, & Leemaqz 2018). Given susceptibility of phenological events to climatic alterations, shifts in the FFD are anticipated (Hudson, 2018a). A study by Bock et al (2014) found that the first flowering date herbaceous plants is postponed with increasing altitude, indicating temperature and photoperiod dependencies. Observations across five plant communities at varying altitudes revealed that higher temperatures and reduced rainfall are linked to earlier flowering periods (Prevéy 2020).

While European research has demonstrated how climate change impacts plant flowering times and FFD (Hudson, 2018a, 2018b), data on alterations in flowering sequences in Australia remains scarce, underscoring the need for further investigation in the Australasian context (Hudson, 2011; Hudson, 2010; Keatley et al. 2013). Understanding the sensitivity of a plant's flowering commencement and cessation to historical drought conditions can facilitate predictions on future plant responses to potential droughts. This report places particular emphasis on the variations in FFD during and following the drought years of 1996/7 to 2009/10 compared to non-drought years, along with other observable shifts in patterns. The significance of studying FFD and flowering order (FO) lies in its broad ecological implications, encompassing increased competition among plants for soil nutrients, adverse effects on bee populations and pollination processes, consequences for the food availability of foragers, and repercussions for the agricultural sector. Such changes in FFD or FO can have profound impacts on the intricate balance on ecosystems, affecting both flora and fauna and, by extension, the services they provide to human societies.

Within the Australian context, significant concerns have been raised regarding climate change, a situation underscored by a dearth of research focused on FFD despite its importance. Cole (2017) and Marchin (2015) have demonstrated that climate change markedly influences phenological events, affirming its prominence as a pressing issue in Australia. However, to date, only a limited number of studies, including those by Keatley, Hudson & Leemaqz (2018) and Hudson et al. (in preparation), have specifically analysed FFD. Australia's vulnerability to extreme weather, amplified by global climate changes, has led several severe drought periods, notably the Millennium Drought from 1997 to 2009. This drought had a widespread and prolonged impact on most of southern Australia, particularly affecting the densely populated southeast and southwest regions. Initially concentrated in Victoria and Tasmania, by 2001, the drought had extended across southern Australia (Van Dijk et al. 2013).

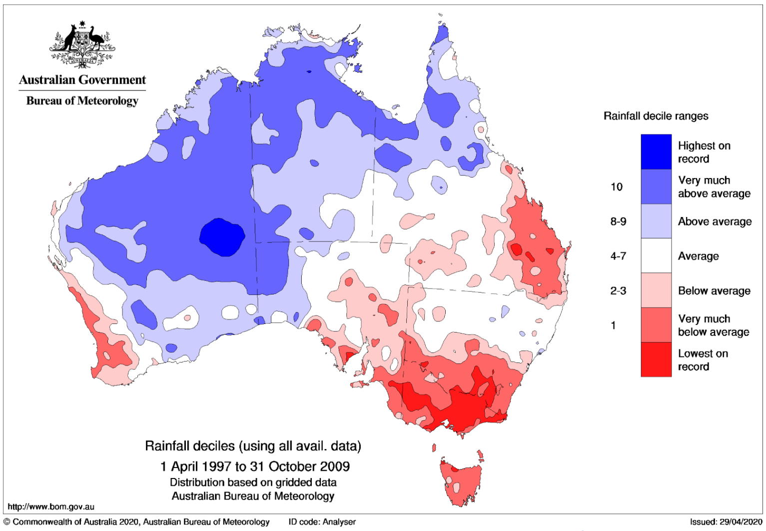

Figure 1: Millenium Drought (1997 to 2009) rainfall deciles.

The goal of this report is to delve into the effects of climate change, with a particular focus on the Millennium drought period by analysing the FFD of a selection of Australian flower species. This will be accomplished by using techniques such as time series clustering and Rank Biased Overlap (RBO). The Methodology section will provide the theory and details of how this was accomplished along with further detail on the dataset and software tools used to implement the analysis techniques. The findings made from the analysis of the data will be presented in the Results section. Following this, the implications of the results are explored in the Discussion section. Finally, a Conclusion section will summarise the key findings made with potential future research avenues outlined.

Statement of Authorship

The workload was divided as follows:

- Dillon Batdorf analysed this data using Python, and produced all results and interpretations written in this report.
- Professor Irene Hudson (Mathematical Sciences, STEM College, RMIT) supervised the project, providing project direction, assisted with interpreting results, and proofread the report.
- Dr. Marie Keatley (Applied Chemistry and Environmental Science, STEM College, RMIT), provided the data used in this project.

Methodology

This section of the report delineates the approach employed to examine the effects of climate change on the first flowering day (FFD) of Australian flower species, focusing on the Millennium Drought period. Utilising methods such as time series clustering and Rank Biased Overlap (RBO), this study aims to identify patterns in flowering times and assess shifts in species' flowering order relative to climatic changes. This section will detail the theoretical basis, implementation, and rationale behind these analytical techniques, and outline the software tools used. By integrating these methodologies, the research endeavours to offer a comprehensive analysis of how climate variations influence plant phenology, thereby contributing to our understanding of ecological responses to environmental stressors.

Data

The dataset used is a combination of two separate datasets that have been used in numerous studies focused on determining flowering changes in Australia (Keatley & Hudson 2018). The dataset contains yearly data from 1984 to 2014. One of the main parts of the data is the *First Flowering Day* (FFD) of 81 Australian native flower species. The FFD is the number of days since January 1st of each year when the flower species have first bloomed. The other part of the data relates to climate data related to the region of this datasets flower species. The climate measures in this dataset include maximum temperature, minimum temperature, rainfall, radiation, vapor pressure and radiation. All the variables present in the dataset are annual measures ranging from 1984 to 2014.

Analysis Techniques

Time Series Clustering

Clustering represents a technique in data mining that groups together similar datapoints into related or uniform categories, without prior understanding of the category definitions (Rai & Singh 2010). Clusters are created by assembling items based on their high degree of similarity to others within the same cluster, and their low similarity to items in different clusters. This method is valuable for exploratory data analysis because it reveals patterns within the unlabelled dataset though the objective categorisation of data into similar groupings (Aghabozorgi, Seyed Shirkhorshidi & Ying Wah 2015). Time series data, a form of temporal data, inherently exhibits high dimensionality and encompasses a large volume of data (Rani & Sikka 2012). Clustering complex entities is especially beneficial as it facilitates the uncovering of intriguing patterns within time series datasets. The use of time series clustering in this report is for the purpose of pattern discovery, to generate insights into how climate variations may be influencing plant phenology.

While there are many time series clustering methods that can be employed, this report focuses on the use of non-probabilistic partitional clustering algorithms. As outlined by

Holder et al. (2024) this algorithm begins by selecting example cases whose purpose is to characterise a cluster, which is commonly known as the initialisation stage. Following the selection of these example cases, an updating process begins, where the examples are refined in an iterative manner until a convergence condition has been met. The algorithm that is used in this report is the k-medoids clustering algorithm. The k-medoids algorithm will carry out the partitional clustering steps until the value of the clustering objective function becomes stable (Han & Kamber 2012). The clustering objective function that aims to minimise Total Deviation (TD) is given as follows:

(1)

$$\sum_{i=1}^{k} \sum_{x_c \in C_i} d(x_c, e_i)$$

(2)

$$s = 40 + 2k$$

where *k* is the number of clusters, C_i is the set of cases in the *ith* cluster, *d* is the dissimilarity measure, x_c is a case in cluster C_i and e_i is the exemplar (representative) of cluster C_i (Holder, Guijo-Rubio & Bagnall 2023).

The initialisation technique used in this report is known as *random initialisation*, which works by choosing the initial cluster centres from the dataset at random. The decision to implement this technique was due to its popularity (Holder, Middlehurst & Bagnall 2024). The distance measure, used in this report is the move-split-merge (MSM) distance measure as research has shown it to be the most effective distance function (Holder, Middlehurst & Bagnall 2024). The number of clusters selected for the clustering model in this report is 5. Only the flowering data is used to create the model, with the climate variables being excluded.

Rank-Biased Overlap (RBO)

Rank-Biased Overlap (RBO) is a method that allows the comparison of two ranked lists (Webber, Moffat & Zobel 2010). A numerical value between zero and one is generated to quantify the similarity between the ranked lists. A value of zero is expected if each list does not share any elements, while a value of one, indicates that each list shares the same elements, and they are all in the same order. In this case, lists ordered from earliest to latest median FFD for each cluster is generated, with RBO metrics for all ten cluster combinations determined. The purpose of doing this is to quantify the amount of change in flowering days across clusters which in the case of this report, are groups of years that exhibit similar FFD patterns.

Software Tools

To perform these tasks, the Python programming language is utilised. Within the Python programming language, the *aeon* library is used to create the time series clustering model, while the *pandas* library is used to create tables. The Python libraries *matplotlib* and *seaborn* were used to create visualisations.

Results

In this section the results of the time series clustering and RBO metrics will be displayed and interpreted. As mentioned previously, the focus will be on the Millennium Drought period of approximately 1996/7 – 2009/10.

Time Series Clustering

Table 1 displays the results of the cluster model, which has allocated each year of the FFD data to a cluster. Cluster 0 (Red) appears frequently in the early 1990s and again in 2000, 2001, and 2010. Cluster 1 (Green) is seen in consecutive years from 1985 to 1986. Cluster 2 (Blue) appears throughout the time span including in consecutive years from 2002 to 2005, during the Millennium Drought. Cluster 3 (Orange) is most frequently observed in consecutive years from 2006 to 2009, towards the end of the Millennium drought. Cluster 4 is relatively rare, only appearing in 1996, and 2001. It is clear from this output that the Millennium Drought period coincides with a shift in clusters. Cluster 0 is prevalent in the years preceding the drought while Clusters 2 and 3 are more prevalent in the middle and end of the drought.

Table 1: Assignment of Years to Clusters Based on Time Series Clustering of First Flowering Day (FFD) Data from 1984 to 2014

Year	cluster_labels
1984	2
1985	1
1986	1
1987	1
1988	2
1989	0
1990	2
1991	0
1992	0
1993	0
1994	0
1995	3
1996	4
1997	0
1998	2
1999	3
2000	0
2001	4
2002	2
2003	2
2004	2
2005	2
2006	3
2007	3
2008	3
2009	3
2010	2
2011	3
2012	2
2013	3
2014	2

Rank-Biased Overlap (RBO)

Table 2 displays the results of the RBO metrics for the median FFD values of each cluster ranked from earliest FFD to latest FFD. The closer to 1 the RBO value is the higher its concordance. Table 2 shows a RBO value of 0.907 which indicates a high concordance. This value signifies a significant maintenance of species rank order between Cluster 0 (predominantly spanning the years of 1991-1994, before the drought) and Cluster 2 (predominantly spanning the years 2002-2005, six years into the drought). Furthermore, there is a relatively good concordance of RBOs, indicating the maintenance of species rank order between Cluster 2 (predominantly spanning the years 2002-2005, six years into the drought) and Cluster 3 (predominantly 2006-2009, the latter part of the drought), with a value of 0.891. A similar pattern of relatively good concordance in RBOs is observed on the maintenance of species rank order between Cluster 2 and Cluster 4 (years 1996 and 2001), with a value of 0.873. However, it is important to note that the RBO values suggest that the ranks of Cluster 1 (years 1985-1987), before the drought) are not highly concordant with Clusters 2, 3, and 4. This discrepancy illustrates the disruption caused by the drought.

Table 2 : Rank-Biased Overlap (RBO) Metrics Comparing the Median First Flowering Day Sequences Between Clustered Years

Column 1	**Column 2**	**RBO**
Cluster 0	Cluster 1	0.868821
Cluster 0	Cluster 2	0.906980
Cluster 0	Cluster 3	0.866367
Cluster 0	Cluster 4	0.864934
Cluster 1	Cluster 2	0.848207
Cluster 1	Cluster 3	0.821773
Cluster 1	Cluster 4	0.819109
Cluster 2	Cluster 3	0.889872
Cluster 2	Cluster 4	0.873155
Cluster 3	Cluster 4	0.831436

Discussion

This report examined the first flowering day (FFD) of 81 Australian flower species over 31 years, employing time series clustering and Rank Biased Overlap (RBO) to elucidate the effects of climate change with a focus on the Millennium Drought. The results unveiled distinctive clusters, indicating variability in FFD patterns, and RBO analysis suggested relative stability in FFD patterns when clusters made up of years from or close to drought periods were compared with each other, while there were clear shifts when comparing clusters made up of years far away from drought periods when compared to clusters made up of years from drought periods.

The time series clustering results indicated periods of phenological congruity and divergence. Notably, years within the Millennium Drought (1997-2009) were often grouped into distinct clusters compared to the pre- and post- drought years, implying a potential reshaping of flowering patterns by extended arid conditions. The RBO scores highlighted a moderate similarity between flowering ranks, with the highest similarity between the clusters that consist of drought years and the lowest similarity between clusters made up of non-drought years and clusters containing drought years. This suggests that climatic influences such as droughts may cause ecological disruptions such as when flowers bloom. It may also be inferred that while the general ranking of the flowering days may be conserved, individual species' responses to climate stressors may lead to notable shifts in phenological synchrony.

Shifts in flowering time can have profound impacts on society, particularly in agriculture. For instance, the control of flowering time strongly influences grain yield in crops (Wu et al. 2023). This indicates that there is a great need to understand how and why flowering time is changing. This report is limited in its reliance on historical data, with the most recent year in the dataset being 2014. The lack of recent data means there may be further insights missing regarding the way flower species are responding to the change in climate. Additionally, the analytical methods used, while robust, offer a simplified view on what are complex ecological interactions.

Conclusion

This investigation into the first flowering day (FFD) of various Australian flower species over a span of 31 years has illuminated the subtle yet significant influences of climate change on phenological events. Through the application of time series clustering, distinct periods were identified, reflecting changes in FFD patterns potentially driven by climatic factors such as the Millennium Drought. Rank-Biased Overlap (RBO) analysis further underscored these findings, revealing variation in flowering species across certain different clusters of years. Collectively, these results underscore the dynamic interplay between climate and phenological timings.

Future research should focus on expanding the scope of phenological studies to include a broader range of species and ecosystems. This, combined with the addition of more recent data would allow the creation of sophisticated predictive models that incorporate a multitude of factors would allow for better preparation for the societal changes that will be brought by climate change. It should also be noted that future work on investigating the maintenance or otherwise on the rank order of FFD over time will involve upper and lower copula threshold methods (Hudson et al. 2022; Tursunavelia et al 2019) Ultimately, this research highlights the urgency for the scientific community and policymakers to address the repercussions of climate change on natural systems.

References

Aghabozorgi, S, Seyed Shirkhorshidi, A & Ying Wah, T 2015, 'Time-series clustering – A decade review', Information Systems, vol. 53, pp. 16–38.

Amano, T, Smithers, RJ, Sparks, TH & Sutherland, WJ 2010, 'A 250-year index of first flowering dates and its response to temperature changes', Proceedings of the Royal Society B: Biological Sciences, vol. 277, no. 1693, pp. 2451–2457.

Bock, A, Sparks, TH, Estrella, N, Jee, N, Casebow, A, Schunk, C, Leuchner, M & Menzel, A 2014, 'Changes in first flowering dates and flowering duration of 232 plant species on the island of Guernsey', Global Change Biology, vol. 20, no. 11, pp. 3508–3519.

Ge, Q, Wang, H, Zheng, J, This, R & Dai, J 2014, 'A 170 year spring phenology index of plants in eastern China', Journal of Geophysical Research: Biogeosciences, vol. 119, no. 3, pp. 301–311.

Gerst, KL, Crimmins, TM, Posthumus, EE, Rosemartin, AH & Schwartz, MD 2020, 'How well do the spring indices predict phenological activity across plant species?', International Journal of Biometeorology, vol. 64, no. 5, pp. 889–901.

Han, J & Kamber, M 2012, Data mining: concepts and techniques, 3rd ed, Elsevier, Burlington, MA.

Hassan, T, Gulzar, R, Hamid, M, Ahmad, R, Waza, SA & Khuroo, AA 2023, 'Plant phenology shifts under climate warming: a systematic review of recent scientific literature', Environmental Monitoring and Assessment, vol. 196, no. 1, p. 36.

Holder, C, Guijo-Rubio, D & Bagnall, A 2023, 'Clustering Time Series with k-Medoids Based Algorithms', in G Ifrim, R Tavenard, A Bagnall, P Schaefer, S Malinowski, T Guyet, & V Lemaire (eds), Advanced Analytics and Learning on Temporal Data, Springer Nature Switzerland, Cham, pp. 39–55.

Holder, C, Middlehurst, M & Bagnall, A 2024, 'A review and evaluation of elastic distance functions for time series clustering', Knowledge and Information Systems, vol. 66, no. 2, pp. 765–809.

Hudson, I 2011, 'Meta-analysis In Encyclopedia of Climate and Weather', Oxford University Press, no. 2, pp. 273–287.

Hudson, I 2018a, 'Keynote: Stochastic trend detection by a broad class of state-space models: application to phenology', in Melbourne, Australia.

Hudson, I 2018b, 'Keynote: Tropical Phenology Functional Change in Ecosystems Across Space and Time, What are the analytical challenges and possible solutions?', in Chicheley Hall, Newport Pagnell, Buckinghamshire, United Kingdom.

Hudson, I, Tursunalieva, A & Chase, JG 2022, 'Copula Modelling of Agitation-Sedation (A-S) in ICU: Threshold Analysis of Nurses' Scores of A-S and Automated Drug Infusions by Protocol', in Recent Advances in Medical Statistics, IntechOpen.

Hudson, IL & Keatley, MR (eds) 2010, Phenological research: methods for environmental and climate change analysis,.

Hudson, IL, Keatley, MR & Kang, I 2011, 'Wavelet characterization of eucalypt flowering and the influence of climate', Environmental and Ecological Statistics, vol. 18, no. 3, pp. 513–533.

Keatley, MR, Hudson, I & Leemaqz, S 2018, 'Species on the move: Rank-biased overlap (RBO) metrics for indefinite rankings (adaptive pruning)', in Melbourne, Australia.

Marchin, RM, Salk, CF, Hoffmann, WA & Dunn, RR 2015, 'Temperature alone does not explain phenological variation of diverse temperate plants under experimental warming', Global Change Biology, vol. 21, no. 8, pp. 3138–3151.

Prevéy, JS 2020, 'Climate Change: Flowering Time May Be Shifting in Surprising Ways', Current Biology, vol. 30, no. 3, pp. R112–R114.

Rai, P & Singh, S 2010, 'A Survey of Clustering Techniques', International Journal of Computer Applications, vol. 7, no. 12, pp. 1–5.

Rani, S & Sikka, G 2012, ‘Recent Techniques of Clustering of Time Series Data: A Survey’, International Journal of Computer Applications, vol. 52, no. 15, pp. 1–9.

Schwartz, MD (ed.) 2003, Phenology: an integrative environmental science, Kluwer Academic Publishers, Dordrecht ; Boston.

Tursunalieva, A, Hudson, I & Chase, G 2019, ‘Copula Modelling of Nurses’ Agitation-Sedation Rating of ICU Patients’, in H Nguyen (ed.), Statistics and Data Science, Springer Singapore, Singapore, pp. 148–161.

Van Dijk, AIJM, Beck, HE, Crosbie, RS, De Jeu, RAM, Liu, YY, Podger, GM, Timbal, B & Viney, NR 2013, ‘The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society’, Water Resources Research, vol. 49, no. 2, pp. 1040–1057.

Webber, W, Moffat, A & Zobel, J 2010, ‘A similarity measure for indefinite rankings’, ACM Transactions on Information Systems, vol. 28, no. 4, pp. 1–38.

Wu, L, Marcelis, LFM, Kong, F & Zhu, Y 2023, ‘Editorial: Flowering time control in agricultural and horticultural crops, volume II’, Frontiers in Plant Science, vol. 14, p. 1183355.