ORIGINAL ARTICLE

Journal Section

Synoptic weather regimes over Aoteaora New Zealand

Jonny Willams, PhD[1*] | James Renwick, PhD[2†]

[1]NIWA - Taihoro Nukurangi, 301 Evans Bay Parade, Akautangi, Hataitai, Wellington, Aotearoa New Zealand

[2]School of Geography, Environmental and Earth Science, Victoria University of Wellington - Te Herenga Waka, Kelburn, Wellington, Aotearoa New Zealand

Correspondence
Jonny Williams PhD, NIWA - Taihoro Nukurangi, 301 Evans Bay Parade, Akautangi, Hataitai, Wellington, Aotearoa New Zealand
Email: jonny.williams@niwa.co.nz

Funding information
JW is funded through the Deep South National Science Challenge, under project C01X1902 from the New Zealand Government Department for Business, Innovation and Employment (MBIE).

This work provides an updated set of 12 dominant geopotential height fields - or 'regimes' - over Aotearoa New Zealand. These regimes were initially produced by Kidson (2000) and have provided the basis for many other subsequent studies. These maps provide a guide to the prevailing weather due to the broad equivalence of 1000hPa geopotential height and mean-sea-level pressure. The results presented here are broadly in agreement with previous work but with some important differences. The most notable of these being the need to average two blocking regimes together to provide good agreement between this work and Kidson (2000). These differences are attributed to the software used, improvements to the underlying dataset itself and to the 'mixing' of statistically indistinguishable empirical orthogonal functions - EOFs - in different linear combinations. All data and code used in this work is publicly accessible and it is hoped that this will provide a catalyst for open discussions on this topic, particularly with relation to future perturbations to these regimes under climate change.

KEYWORDS
Synoptic, Regimes, Blocking

Abbreviations: K2K, Kidson Int. J. Climatol. (2000); UTC, Coordinated Universal Time

[*]JW performed the data analysis, plotted the figures and wrote the text.

[†]JR provided extensive guidance on the mathematical formulation used in previous studies, provided the data from K2K plotted in blue contours in figure 3 and contributed the section of the manuscript on EOF mixing.

1 | INTRODUCTION

The world's meteorology and climatology is extraordinarily complex in its spatio-temporal variability. Because of this, simplifications and methodologies for reducing and communicating this complexity are essential if we are to understand it better. Examples are many and various; e.g. the Beaufort scale in wind speed, rainfall probability in a weather forecast, cloud clustering in climate models (Williams and Webb (2009)) and classification of El Niño or La Niña (e.g. Trenberth (1997) for a review).

Although these methodologies are of course hugely different in their scope, they all have one fundamental thing in common; that of the reduction in the complexity of a system such that it can be better understood by the relevant audience. This study is limited only to the region surrounding Aotearoa New Zealand and to one meteorological variable, the 1000hPa geopotential height, z, at 0000 and 1200 Coordinated Universal Time, UTC.

The dataset used is identical to that used in the seminal paper by Kidson (Kidson (2000)) - 'K2K' in this work - which uses 28,852 fields of z between January 1958 and June 1997 to just 12 dominant synoptic weather regimes. These are frequently referred to as 'Kidson types', given their wide use in modern (Parsons et al. (2014)) and paleo (Ackerley et al. (2011)) studies. The same methods have also been used to study weather regimes in other parts of the world such as South America (Solman and Menéndez (2003)).

Synoptic weather types split the tens of thousands of input data points into 3 regimes; trough, zonal and blocking. These are further split into 4, 3 and 5 regimes respectively, making up the final 12. Together, these regimes express the dominant weather types over Aotearoa New Zealand. This is possible since the shape of the geopotential height field is practically the same as mean sea level pressure in this kind of analysis (Kidson (1994a)) and therefore gives a good approximation of the prevailing weather.

This study gives a more detailed account of the derivation of these synoptic regimes than given previously and provides an update to the regimes' occurrences compared to K2K, which itself builds on many other previous studies, for example Kidson (1999), Kidson (1997), Kidson (1994a), Kidson (1994b), Kidson and Watterson (1995) and Ward Jr (1963). The new results presented are broadly in agreement with those from K2K, however there are some notable differences.

The K2K methodology is widely cited, however the available details on the regimes' calculation are somewhat opaque, particularly to those readers less familiar with statistical analysis and clustering techniques. To aid future work, the code used is freely available and a step by step guide is given below for:

- How the weather types themselves are calculated from a reference dataset, in this case the NCEP/NCAR reanalysis.
- How to assign a particular synoptic type to a new observation or model output of z.
- How to interpret the meaning and derivation of the types from mathematical and geometrical arguments.

2 | METHODOLOGY

The most widely referenced work in the literature on this subject is K2K. This builds on earlier work (e.g. Kidson (1997) and Kidson (1994a)) in order to construct 12 dominant synoptic weather regimes. This section gives a step-by-step guide to reproducing these weather types, i.e. Figure 2 in Kidson (2000), which was replotted as Figure 1 in Ackerley et al. (2011) using a different map projection.

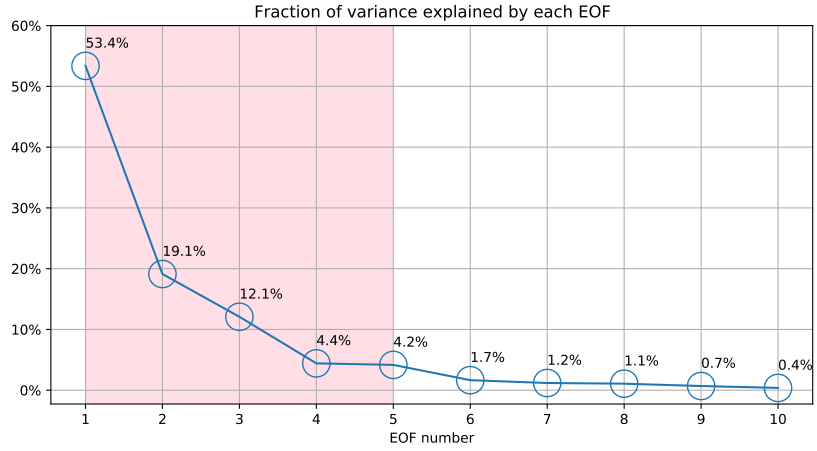

FIGURE 1 Fraction of variance explained by the first 5 EOFs in the time series of z_s. Although the first 10 EOFs are shown, only 5 are used in this analysis (shown by the shaded region).

2.1 | Data and software

We use the same input data as used in K2K, that is 1000hPa geopotential height data (z) from the NCEP/NCAR reanlysis (Kalnay et al. (01 Mar. 1996)) for January 1958 to June 1997 inclusive at 0000 and 1200 UTC.

We use the Python programming language exclusively for this work and make use of the open source `eofs` package (Dawson and Wales (2019), Dawson (2016)) to calculate the principal components (PCs) and empirical orthogonal functions (EOFs). K-means clustering analysis is then carried out on the PCs to obtain the dominant weather regimes using the `scikit-learn` package (Pedregosa et al. (2011)), which relies on `NumPy` (Harris et al. (2020)) and `SciPy` (Virtanen et al. (2020)) for its underlying operation and includes many more functions other than k-means clustering. The final clusters themselves are of equal variance, obtained by finding a minimum of the intra-cluster internal coherence or 'inertia'. A more detailed explanation of the mathematical basis of these packages and methods is beyond the scope of this paper.

2.2 | Mathematical basis

Firstly, height anomalies are calculated by removing the time mean of the heights ($\overline{z}$):

$$z_s = z - \overline{z}. \tag{1}$$

In the literature, this quantity is sometimes referred to as the 'standardised' height, hence the subscript s. Now the EOFs are calculated and the first 5 are retained (as in K2K). The fractions of variance explained by the first 5 EOFs are given in Figure 1 and together they account for 93.2% of the observed variability. The first 5 EOFs are shown in Figure 2 .

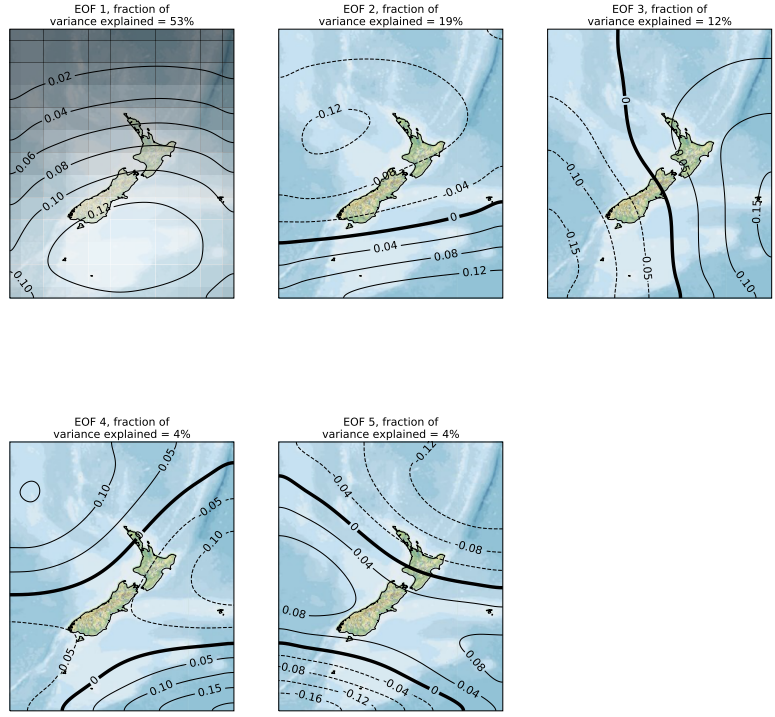

FIGURE 2 The first 5 EOFs, $\mathcal{E}$, of z_s. The contour lines are smoothed and the gridscale of the data used is shown in the subfigure for EOF 1. The background of the figures shows the local relief (Commons (2020)).

Next the PCs, P_n are calculated and normalised to give P_n:

$$\hat{P_n} = \frac{\left(P_n - \overline{P_n}\right)}{\sigma_{P_n}}, \tag{2}$$

where σ_P is the standard deviation of P across time and $1 < n < 5$.

The fundamental operation here involves projecting the height field onto the EOFs and this is discussed further in the context of assigning regimes to arbitrary datasets in Section 4.

Now, the k-means clustering assigns each z_s field to one of 12 clusters. This order of these clusters is arbitrary and is chosen purely to match Kidson's original work. Once each z_s field has been assigned a value between 1 and

12, the final cluster means, C_z are the time mean of the z_s fields assigned to each cluster, that is

$$C_{z,i} = \overline{z_i}, \tag{3}$$

where $1 < i < 12$.

The first 10 of these final clusters are shown in Figure 3, along with their equivalents from Kidson (2000). The associated winds from the same reanalysis product at the same twice daily sampling frequency are shown in Figure 4. Although the agreement between the 2 analyses is generally good for the 10 clusters shown in Figure 3, the 2 remaining ones are not shown due to their pronounced differences with the HW and R blocking clusters from K2K. These differences are discussed in the next section.

3 | HOW DO THE NEW SYNOPTIC TYPES DIFFER FROM K2K AND WHY?

It is clear from Figure 3 that 10 of the synoptic types obtained here are generally in good agreement with those of K2K. That being said, the HNW regime is considerably more zonal in the new case, especially at southernmost latitudes. The SW regime from K2K is displaced north of the equivalent one from this analysis by approximately 100km and the opposite for W. This likely accounts for the factor of approximately 0.5 and 2 difference in occurrence frequency in the new case with respect to K2K.

There are however notable structural differences between the remaining 2 clusters obtained in this work and the HW and R blocking clusters in K2K. So much so in fact, that it is dubious to assign the same synoptic weather type in these cases. Figure 5 shows the HW and R blocking regimes from K2K along with the 2 remaining clusters from this work. The averages of the two remaining regimes are also shown.

The 2 new regimes shown in Figure 5 (a) and (b) are quite different to the HW and R regimes in K2K (Figure 5 (d) and (e)) yet their mean is strikingly similar, as is their combined fractional occurrence of 10.6% versus 10.1%.

As to why the new regimes found here are somewhat different to those of K2K, we have performed sensitivity analysis of the parameters used in the k-means clustering and have found the fractional occurrence of the regimes obtained here to be robust. For example, by default the k-means solver is run 10 times using different initial estimates and each of these estimates is run through 300 iterations. Figure 6 shows the relationship between the maximum number of iterations and the deviation of the final fractional occurrence of the regimes shown in Figures 3 and 5.

The largest deviation across any of the clusters is approximately 2% and therefore cannot account for the larger differences in the SW and W regimes found here compared to K2K (Figure 3 (b) and (g)).

It is also possible that the dataset used in K2K was affected by the assimilation of incorrect pseudo-observations - or PAOBs - in early versions of the reanalysis (e.g. Kidson (1999)).

It should also be acknowledged that no calculation is perfect and that different implementations of common algorithms will inevitably lead to some element of disagreement. This is discussed further with regard to EOF analysis in Dawson (2016) and is noted in the documentation for the k-means software used here - `https://scikit-learn.org/stable/modules/clustering.html#k-means` - "*Given enough time, K-means will always converge, however this may be to a local minimum.*".

The result shown in Figure 5 is reminiscent of the 'mixing' of EOFs or principal components subject to high sampling variability (e.g. Cheng et al. (01 Jun. 1995)). When the eigenvalues of consecutive EOFs are not statistically separate, different linear combinations of those EOFs can appear, as sample size changes. In this case, a similar thing

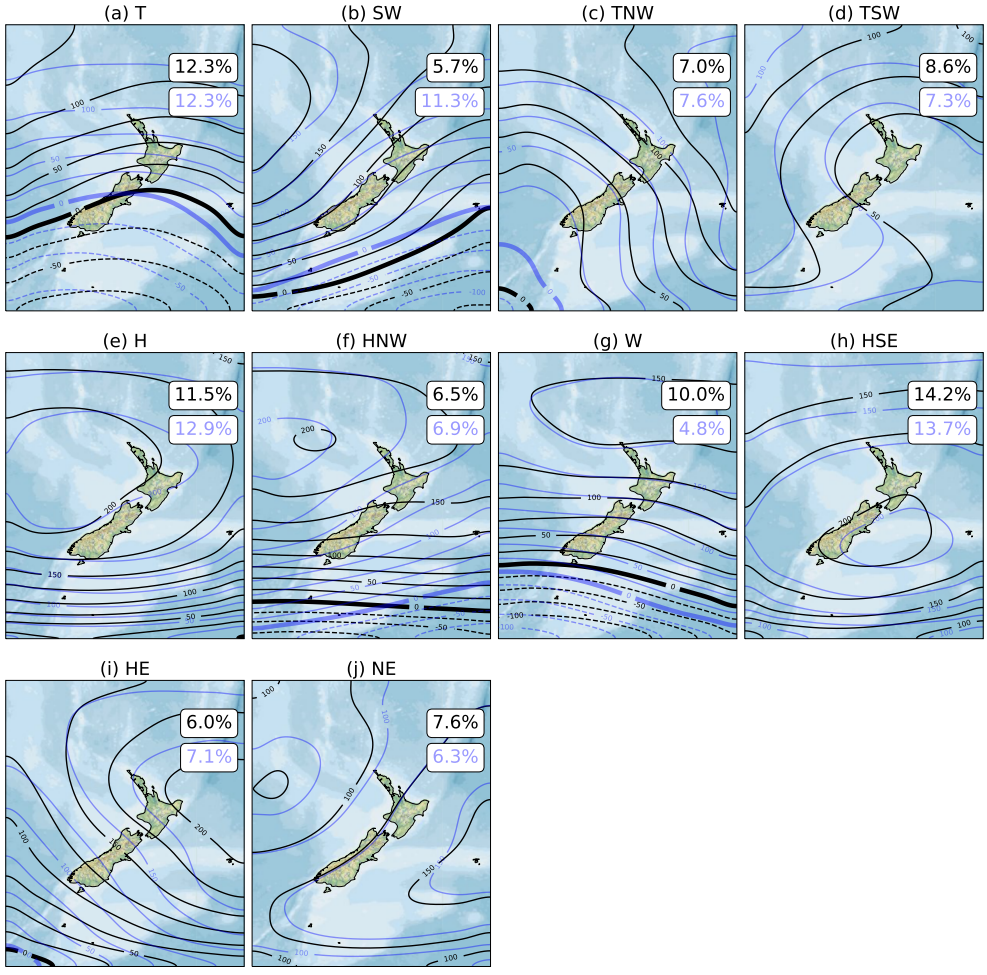

FIGURE 3 The 10 clusters which are deemed to be in close enough agreement with those of K2K to be assigned the same label (in black). Again the contour lines are smoothed and the original K2K clusters are shown in blue. The inset boxes show the fraction of time spent in each regime with the font colours matching the contour line colours. The background of the figures shows the local relief (Commons (2020)).

seems to have occurred within the clustering algorithm and the definition of the cluster means.

4 | HOW TO MANUALLY ASSIGN NEW DATA TO THE CLUSTER SET

In this section we use the term 'observation' to encompass any new dataset to which clusters are assigned.

There have been many studies based on the Kidson types and several of these involve fitting new datasets to the Kidson types (e.g. Ackerley et al. (2011), Parsons et al. (2014)). In this section the mathematical basis for this assignment is given.

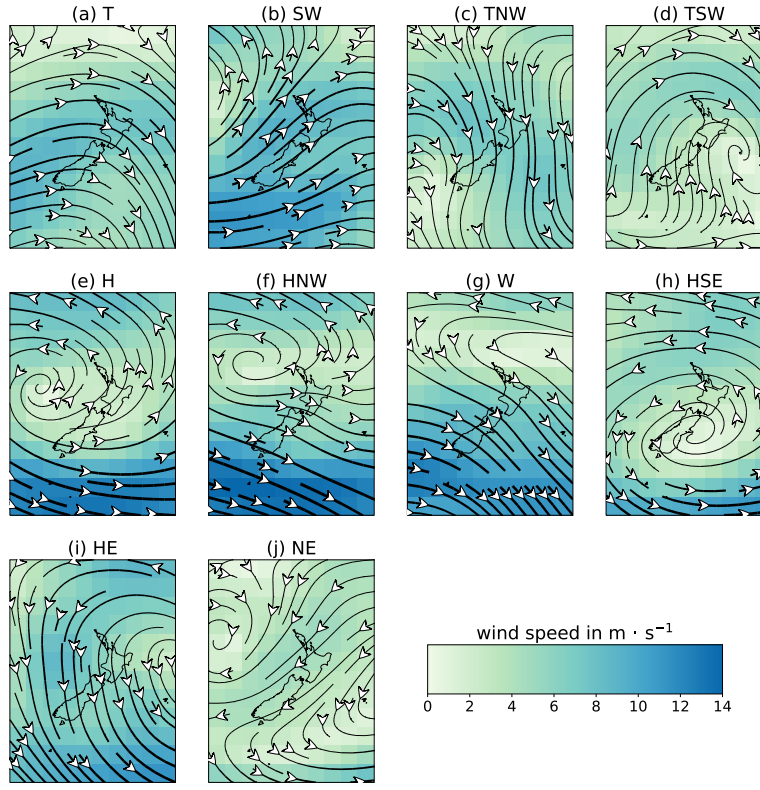

FIGURE 4 Wind streamlines and speeds (m·s^{-1}) for the synoptic regimes shown in Figure 3. The colours show the windspeed at the gridscale. The thickness of the lines is proportional to the local speed and is consistent across subfigures.

The first step is to calculate the 'projection' of the z_s onto the individual EOFs. The dimensions of the EOFs are 5 $\times$ 13 $\times$ 11, that is, 5 EOFs over a region with 13 latitude values and 11 longitude values. The z_s values have dimensions of $N_t \times$ 13 $\times$ 11, where N_t is the number of timesteps considered. The projection, $\mathcal{P}$, is defined as,

$$\mathcal{P} = \sum_{jk} z_{s,jk} \mathcal{E}_{jk} \tag{4}$$

and therefore, the dimensions of $\mathcal{P}$ are $N_t \times$ 5.

For each observation, we now have a 5 element array ($\mathcal{P}$) and a 12 $\times$ 5 element array representing the time mean

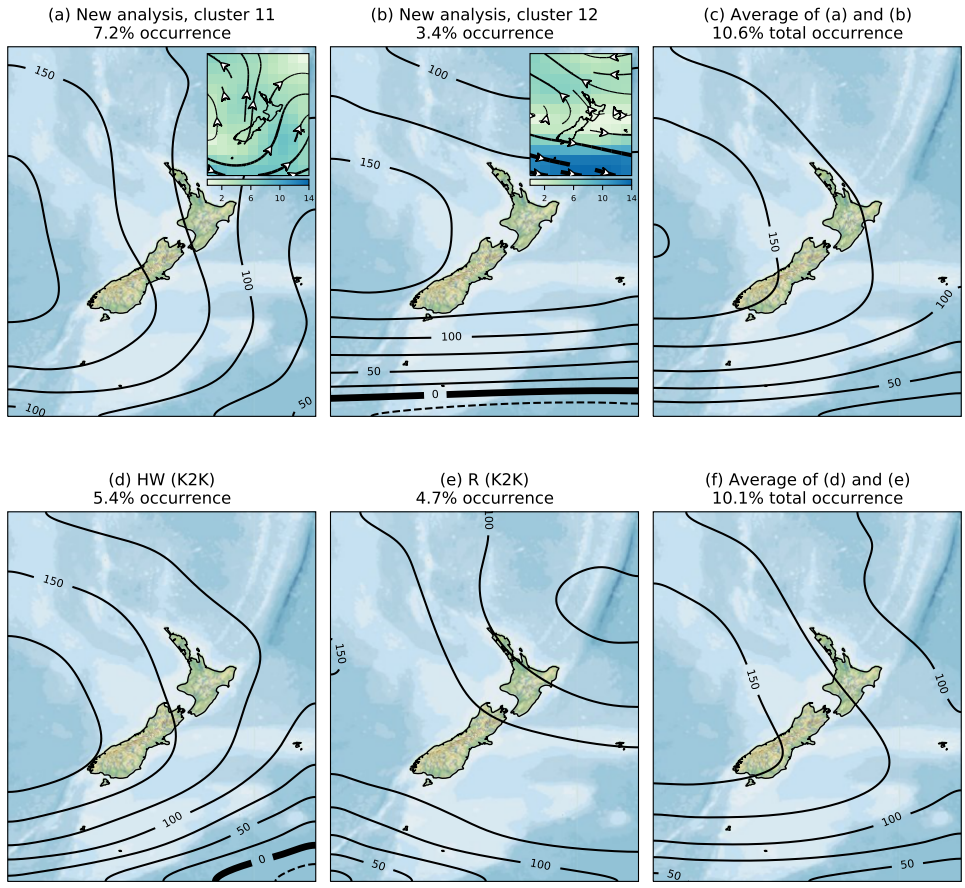

FIGURE 5 (a)-(b) The remaining 2 regimes from the new analysis which are not shown in Figure 3. Subfigures (a) and (b) also show the wind fields for the relevant cluster shown. The colour scale is the same as for Figure 4 and the horizontal density of the streamlines is halved to improve legibility. Subfigure (c) is the average of (a) and (b). (d)-(e) the `HW` and `R` regimes from K2K. Subfigure (f) shows the average of (d) and (e).

of the PCs (P_n) for each index calculated by the k means clustering algorithm,

$$C_{P,i} = \overline{P_i}, \tag{5}$$

where $1 < i < 12$. The $C_{P,i}$ are often referred to in the literature as 'cluster means'.

To find out which of the 12 regimes the projection should be assigned to for each observation, the Euclidean distance, d, between the projection and each of the 12 height clusters is calculated. The projection and PC clusters arrays are normalised and are given by $\hat{\mathcal{P}}$ and $\hat{C}_{P,n}$ respectively. The minimum of these 12 numbers (i.e. the Euclidean

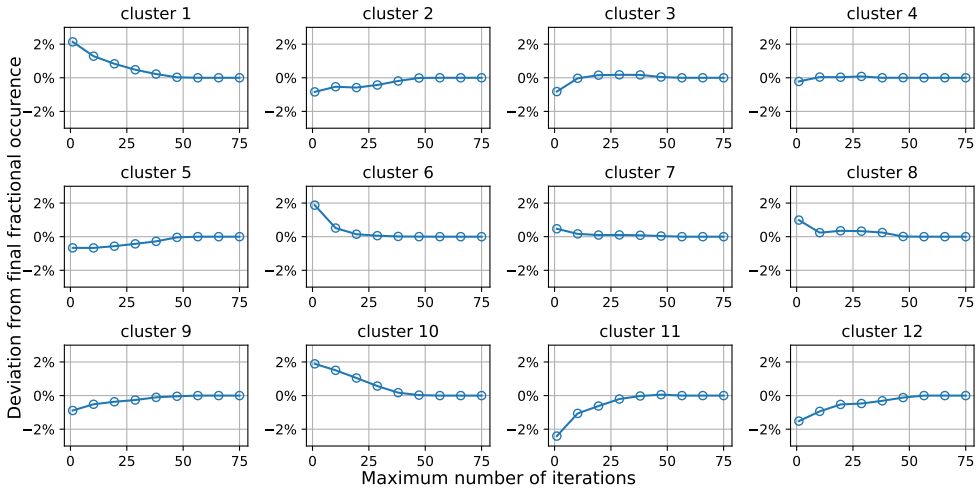

FIGURE 6 Deviation of the fractional occurrence of each regime as a function of the maximum number of iterations used in each pass of the k-means clustering algorithm. The largest difference is substantially smaller than the differences seen in Figure 3 (b) and (g) and therefore cannot account for the differences seen in the occurrences fractions.

distance in principal component space) gives the index and therefore the weather regime of each observation and its fractional occurrence. This can then be directly compared with the values obtained in Figure 3.

The Euclidean distances d_n are given by,

$$d_i = \sqrt{\sum_{n=1}^{N_{\mathcal{E}}} \left|\hat{\mathcal{P}}_{n,i} - \hat{C}_{n,i}\right|^2}, \tag{6}$$

where each of $\hat{\mathcal{P}}_n$ and $\hat{C}_{P,n}$ are 1×5 arrays and $N_{\mathcal{E}}$ is the number of EOFs; i.e. 5 in this work.

A geometrical illustration of what the minimum Euclidean distance means is shown in Figure 7 for an arbitrary observation of z in the NCEP/NCAR reanalysis. It is clear that the figure showing the lowest d value - highlighted lines in Figure 7(g) - has lines of $\hat{\mathcal{P}}$ and $\hat{C}_P$ which are closest to one another. The value of d can therefore be interpreted as a measure of the 'similarity' of $\hat{\mathcal{P}}_n$ and $\hat{C}_{P,n}$.

5 | CONCLUSIONS

In this work we have sought to provide not only a reassessment of dominant synoptic weather types over Aotearoa New Zealand from Kidson (2000), but also to provide detailed information on the calculation of these types, and how to assign arbitrary new data to them. Widely-used peer reviewed software packages in the Python programming language were used to calculate the EOFs and k means clusters in the manner of K2K. Our results are broadly in line with K2K but with some notable differences. The key findings of this work are:

- Although 10 of the 12 clusters identified in K2K are well reproduced in 'shape', 2 of the blocking regimes from

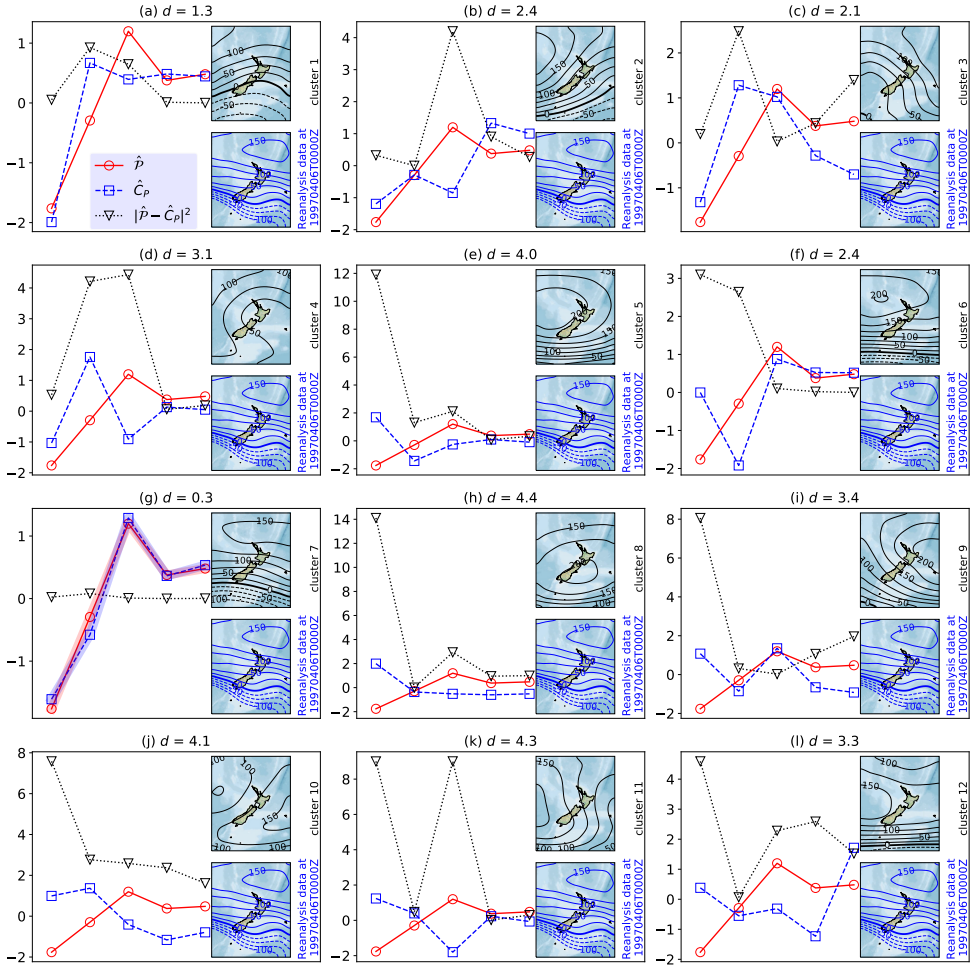

FIGURE 7 Geometrical illustration of how an arbitrary observation of z (here at 0000 UTC on the 6th of April 1997) can be fitted to pre-existing clusters. Each subfigure shows the same arbitrary observation of geopotential height in the blue contours and the black contours show the 12 separate clusters obtained from this analysis (Figure 3 and Figures 5 (a) and (b)). As shown in the text, we want to find the smallest Euclidean distance, d, between the projection of the observation onto the EOFs, $\hat{\mathcal{P}}$ (○), and the cluster means, $\hat{C}_P$ (□) as given by Equation 6. The individual $|\hat{\mathcal{P}} - \hat{C}_P|^2$ values are also shown (▽). The square root of the sum of the $|\hat{\mathcal{P}} - \hat{C}_P|^2$ values gives the Euclidean distance, d, shown in the subfigure titles and (g) highlights the similarity of the blue and red lines and hence the lowest d value.

K2K (HW and R) do not have recognisable analogues in the cluster found in this work. See Figure 3.

- Two of the regimes from K2K differ by factors of approximately 0.5 and 2 respectively with their spatially similar regimes from this work. This is attributed to meridional shifts of about 100km in both cases, although in opposite directions.
- The average of the HW and R clusters from K2K are in striking agreement with the equivalent average of cluster

11 and 12 from this work, see Figure 5. This is attributed to different levels of EOF mixing in this work and in K2K and the sensitivity of the k-means clustering algorithm.

Future work will apply this methodology to differences in synoptic weather regimes over Aotearoa New Zealand in the UK Earth System Model (UKESM, e.g. Sellar et al. (2020)) and in the NZESM (e.g. Behrens et al. (2020)). This will be especially pertinent with regards to climate change and its affect on the dominant weather types that can be expected to occur in the future. Work is also underway to apply this methodology to the new ERA-5 analysis (Hersbach et al. (2019)). This has ~100 times the areal resolution of NCEP/NCAR and is therefore able to explicitly resolve orographic features which are absent in this work.

6 | ACKNOWLEDGEMENTS

JW would like to thank Dr Nicolas Fauchereau of NIWA for his assistance in using the `scikit-learn` package at the outset of this work and Dr Andrew Lorrey, also of NIWA, for useful discussions on the use and interpretation of synoptic regimes. JW is funded through the Deep South National Science Challenge, under project C01X1902 from the New Zealand Government Department for Business, Innovation and Employment (MBIE). JW acknowledge the use of the high performance computing environment made available through the New Zealand eScience Infrastructure organisation, NeSI.

7 | CODE AND DATA AVAILABILITY

The code used here is publicly available as a Jupyter notebook via `GitHub` at `https://github.com/jonnyhtw/weather-types` and the NCEP/NCAR reanalysis data is also freely available at `https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.pressure.html`.

references

Ackerley, D., Lorrey, A., Renwick, J. A., Phipps, S. J., Wagner, S., Dean, S., Singarayer, J., Valdes, P., Abe-Ouchi, A., Ohgaito, R. and Jones, J. M. (2011) Using synoptic type analysis to understand new zealand climate during the mid-holocene. *Climate of the Past*, **7**, 1189–1207. URL: `https://cp.copernicus.org/articles/7/1189/2011/`.

Behrens, E., Williams, J., Morgenstern, O., Sutton, P., Rickard, G. and Williams, M. J. (2020) Local grid refinement in new zealand's earth system model: Tasman sea ocean circulation improvements and super-gyre circulation implications. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS001996.

Cheng, X., Nitsche, G. and Wallace, J. M. (01 Jun. 1995) Robustness of low-frequency circulation patterns derived from eof and rotated eof analyses. *Journal of Climate*, **8**, 1709 – 1713. URL: `https://journals.ametsoc.org/view/journals/clim/8/6/1520-0442_1995_008_1709_rolfcp_2_0_co_2.xml`.

Commons, W. (2020) File:equirectangular-projection-topographic-world.jpg — wikimedia commons, the free media repository. URL: `https://commons.wikimedia.org/w/index.php?title=File:Equirectangular-projection-topographic-world.jpg&oldid=502090647`. [Online; accessed 28-January-2021].

Dawson, A. (2016) eofs: A library for eof analysis of meteorological, oceanographic, and climate data. *Journal of Open Research Software*, **4**. URL: `http://doi.org/10.5334/jors.122`.

Dawson, A. and Wales, S. (2019) ajdawson/eofs: Version 1.4.0. URL: `https://doi.org/10.5281/zenodo.2661604`.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. and Oliphant, T. E. (2020) Array programming with NumPy. *Nature*, **585**, 357–362. URL: https://doi.org/10.1038/s41586-020-2649-2.

Hersbach, H., Bell, W., Berrisford, P., Horányi, A., J., M.-S., Nicolas, J., Radu, R., Schepers, D., Simmons, A., Soci, C. and Dee, D. (2019) Global reanalysis: goodbye era-interim, hello era5. 17–24. URL: https://www.ecmwf.int/node/19027.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. and Joseph, D. (01 Mar. 1996) The ncep/ncar 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437 – 472. URL: https://journals.ametsoc.org/view/journals/bams/77/3/1520-0477_1996_077_0437_tnyrp_2_0_co_2.xml.

Kidson, J. W. (1994a) An automated procedure for the identification of synoptic types applied to the new zealand region. *International Journal of Climatology*, **14**, 711–721. URL: https://doi.org/10.1002/joc.3370140702.

— (1994b) Relationship of new zealand daily and monthly weather patterns to synoptic weather types. *International Journal of Climatology*, **14**, 723–737. URL: https://doi.org/10.1002/joc.3370140703.

— (1997) The utility of surface and upper air data in synoptic climatological specification of surface climatic variables. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **17**, 399–413. URL: https://doi.org/10.1002/(SICI)1097-0088(19970330)17:4%3C399::AID-JOC108%3E3.0.CO;2-M.

— (1999) Principal modes of southern hemisphere low-frequency variability obtained from ncep–ncar reanalyses. *Journal of Climate*, **12**, 2808–2830. URL: https://doi.org/10.1175/1520-0442(1999)012%3C2808:PMOSHL%3E2.0.CO;2.

— (2000) An analysis of new zealand synoptic types and their use in defining weather regimes. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **20**, 299–316. URL: https://doi.org/10.1002/(SICI)1097-0088(20000315)20:3%3C299::AID-JOC474%3E3.0.CO;2-B.

Kidson, J. W. and Watterson, I. G. (1995) A synoptic climatological evaluation of the changes in the csiro nine-level model with doubled co2 in the new zealand region. *International journal of climatology*, **15**, 1179–1194. URL: https://doi.org/10.1002/joc.3370151102.

Parsons, S., McDonald, A. J. and Renwick, J. A. (2014) The use of synoptic climatology with general circulation model output over new zealand. *International journal of climatology*, **34**, 3426–3439. URL: https://doi.org/10.1002/joc.3919.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830. URL: https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

Sellar, A. A., Walton, J., Jones, C. G., Wood, R., Abraham, N. L., Andrejczuk, M., Andrews, M. B., Andrews, T., Archibald, A. T., de Mora, L. et al. (2020) Implementation of uk earth system models for cmip6. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS001946.

Solman, S. and Menéndez, C. (2003) Weather regimes in the south american sector and neighbouring oceans during winter. *Climate Dynamics*, **21**, 91–104. URL: https://doi.org/10.1007/s00382-003-0320-x.

Trenberth, K. E. (1997) The definition of el niño. *Bulletin of the American Meteorological Society*, **78**, 2771 – 2778. URL: https://journals.ametsoc.org/view/journals/bams/78/12/1520-0477_1997_078_2771_tdoeno_2_0_co_2.xml.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P. and SciPy 1.0 Contributors (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, **17**, 261–272. URL: https://doi.org/10.1038/s41592-019-0686-2.

Ward Jr, J. H. (1963) Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, **58**, 236–244. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10500845.

Williams, K. D. and Webb, M. (2009) A quantitative performance assessment of cloud regimes in climate models. *Climate dynamics*, **33**, 141–157. URL: https://doi.org/10.1007/s00382-008-0443-1.