A History of Open Weather in New Zealand (HOWNZ): an open access 1-km resolution monthly 1910-2019 time-series of interpolated temperature and rainfall grids with associated uncertainty

Thomas R. Etherington[1], George L. W. Perry[2], Janet M. Wilmshurst[1]

5   [1]Manaaki Whenua – Landcare Research, Lincoln 7608, New Zealand
[2]School of Environment, University of Auckland, Private Bag 92019, Auckland, New Zealand

Correspondence to: Thomas R. Etherington (etheringtont@landcareresearch.co.nz)

Abstract. Long time-series of weather grids are fundamental to understanding how weather affects environmental or ecological patterns and processes such as plant distributions, plant and animal phenology, wildfires, and hydrology. Ideally

10   such weather grids should be openly available and be associated with uncertainties so that users can understand any data quality issues. We present a History of Open Weather in New Zealand (HOWNZ) that uses climatological aided natural neighbour interpolation to provide monthly 1-km resolution grids of total rainfall, mean air temperature, mean daily maximum air temperature, and mean daily minimum air temperature across New Zealand from 1910 to 2019. HOWNZ matches the best available temporal extent and spatial resolution of any open weather grids that include New Zealand, and is

15   unique in providing associated spatial uncertainty in appropriate units of measurement. The HOWNZ weather and uncertainty grids capture the dynamic spatial and temporal nature of the monthly weather variables and the uncertainty associated with the interpolation. We also demonstrate how to quantify and visualise temporal trends across New Zealand that recognise the temporal and spatial variation of uncertainties in the HOWNZ data. The HOWNZ data is openly available at   https://doi.org/10.7931/zmvz-xf30 (Etherington et al., 2021).

20

1 Introduction

Climatologies such as WorldClim2 (Fick and Hijmans, 2017) and Chelsa (Karger et al., 2017) provide spatial grids of climatic variables such as temperature and rainfall that have underpinned thousands of environmental and ecological studies. These climatologies represent long-term ≈ 30-year climate averages, but long time-series of weather grids are also highly

25   desirable as they can be correlated with long-term environmental or ecological data to explore how weather influences such processes. For example, monthly time-series data of weather conditions can be used to improve the understanding of plant distributions (Stewart et al., 2021), plant and animal phenology (Gordo and Sanz, 2005), wildfire histories (Girardin and Wotton, 2009), and hydrology (Remesan et al., 2014).

30   Meteorologists and climatologists are recognising the importance of implementing open access methodologies, as openly sharing data, source code, and knowledge provides exciting opportunities for scientific discovery (de Vos et al., 2020). In New Zealand there are limited open access options for national-scale weather grids (Table 1), and none of these meet all the currently optimal criteria amongst these open access options of spatial (≤ 1 km^2) and temporal (monthly) resolution over the last century. Producing such historical weather data is challenging because weather station data are often sparse, especially

35   in remote areas and for earlier time periods. However, this challenge should not preclude the creation of historical weather data, if we accept that we cannot be consistently successful for all locations and dates and we prioritise providing weather data with associated uncertainty data.

Much emphasis has been placed on quantifying (Foley, 2010) and visualising (Retchless and Brewer, 2016) the uncertainty

40   of future climates. Quantifying and visualising uncertainty is also critical for the judicious use of historic weather data, as even when following open science practices that involve sharing data and code it can still be difficult to ensure that an end-user is aware of and understands the quality of the data (de Vos et al., 2020). Historical climate and weather data are often used predictively, but these predictions must recognise the underlying uncertainties so that their reliability can be understood and communicated. The uncertainty (or reliability) of weather and climate spatial data is often quantified with metrics such

45   as the root-mean-square error (RMSE) or the mean absolute error (MAE) derived from cross-validation that iteratively excludes each weather station and then interpolates a value for the weather station using the remaining data (Willmott and Matsuura, 2006). However, these are global metrics that describe uncertainty with a single spatially averaged value and provide no information about how the uncertainty will vary across space (Zhang and Goodchild, 2002). When estimates of spatial uncertainty accompany climate and weather data, they typically simply indicate areas that are reliable (Abatzoglou et

50   al., 2018;Harris et al., 2020b), which limits how uncertainty can be incorporated into analyses as estimates from individual cells are not matched with an uncertainty in the units of measurement.

To facilitate understanding of how long-term weather patterns have been changing and potentially affecting environmental and ecological processes in New Zealand, we have produced a History of Open Weather in New Zealand (HOWNZ).

55   HOWNZ is a new openly available history of temperature and rainfall weather in New Zealand that matches the best available spatial and temporal extent and resolution of any currently available open data and is unique in providing associated spatial uncertainty grids in appropriate units of measurement (Table 1).

2. Weather and climatology data

60   Monthly weather station data are freely available from New Zealand's National Climate Database (NIWA, 2020). For New Zealand's three main islands and their associated near-shore islands we compiled all reliable data with sub-kilometre spatial

precision from 1900 to 2019 for monthly statistics of: total rainfall, mean air temperature, mean daily maximum air temperature, and mean daily minimum air temperature. The amount of data available in any given month varied through time, across space, and between weather variables, but there were some consistencies. There was much more rainfall data

65   than temperature data, there was generally more data in recent times (with a peak around 1980), and there was always less data in mountainous interiors of both islands and in the more remote southerly and westerly regions of New Zealand (Figure 1). Openly available New Zealand climatology grids for each weather variable giving the average weather in each month for the period 1950-1980 at 100 m grid cell resolution (McCarthy et al., 2021;Leathwick et al., 2002) were reprojected and aggregated to 1-km^2 grid cell resolution.

70

3. Interpolation with uncertainty

Climatologies have formed part of successful previous weather interpolations in New Zealand (Tait et al., 2006). Therefore, we used a climatologically aided interpolation approach by interpolating monthly weather as an anomaly from a climatic normal (Willmott and Robeson, 1995) as this technique has been successfully applied elsewhere (Abatzoglou et al.,

75   2018;Harris et al., 2020b;Hofstra et al., 2008). The basic premise of climatologically aided interpolation is that rather than directly interpolating weather station data in each month, by using an underlying long-term climatology grid for that month, monthly anomalies can be calculated as the difference between the weather station value and the climatology grid value. These anomalies are then interpolated and added to the climatology grid to estimate how the weather in that month deviated from the climatological normal.

80

We selected natural neighbour (or Sibson) interpolation (Sibson, 1981) to interpolate the anomalies. This decision was based on evidence that natural neighbour interpolation performs well for interpolating weather data (Hofstra et al., 2008;Keller et al., 2015;Lyra et al., 2018), but more specifically because natural neighbour interpolation is: (i) an exact interpolator, meaning it will retain the original data values in the interpolated grid and will only interpolate within the range

85   of the original data and so cannot produce wildly unrealistic interpolations; (ii) a local method, which interpolates for a location only using data from that location's immediate surrounds; (iii) spatially adaptive, so automatically adapting to localised data distribution and density; (iv) not based on fitting statistical trends and so does not require large sample sizes (Etherington, 2020). These properties of natural neighbour interpolation are desirable given our interpolations will need to adapt to the increasingly sparse and irregular data that occurs further back in New Zealand history.

90

We applied a discrete (or digital) form of natural neighbour interpolation (Park et al., 2006) that simultaneously calculates uncertainty as a cross-validation error-distance field (Etherington, 2020). This interpolation method works by defining a grid of cells for which interpolated values will be calculated; data cells are given the mean value of any weather stations they

95   contain, and all other cells are given the value of the nearest data cell (Figure 2a). The interpolated value for a grid cell is the mean of the grid cell values that are as close or closer to the interpolation cell than a data cell (Figure 2b), which, when repeated across all grid cells, results in a smooth interpolation grid (Figure 2c). Uncertainty is calculated by using the same discrete natural neighbour interpolation process to interpolate the distances to the data cells to produce natural neighbour distances (Figure 2d) and to interpolate the cross-validated error rate, as the ratio of the cross-validated absolute error to natural neighbour distance, for each data cell (Figure 2e). The product of the natural neighbour distances and cross-validated

100   error rates produces a cross-validation error-distance field (Figure 2f) that yields a grid of interpolation uncertainties in the same units as the interpolation and is highest in areas that are more distant from data cells and are in areas that areas that are harder to interpolate accurately.

      The resulting weather grids appeared to capture the heterogeneous and dynamic nature of the monthly weather variables and

105   the uncertainty associated with the interpolation. For example, the May total rainfalls and uncertainties (Figure 3) illustrate the dynamic nature of both the monthly weather variables and the uncertainty. There are clear differences in total rainfall in May over time and shifts in the location of the wettest regions. Our emphasis on quantifying uncertainty is justified by the uncertainty grids being extremely dynamic, with the magnitude and location of uncertainty changing dramatically through time. We interpret this dynamic uncertainty as being due to the data limitations of interpolation, as higher uncertainty occurs

110   in locations more distant from weather data (Figure 1). However, uncertainty can be high in regions with weather data available, which we interpret as arising from the spatial variability of individual monthly weather patterns, as when spatial variability of weather patterns increases over shorter distances, interpolation becomes increasingly difficult (Etherington, 2020).

115   **4. Interpolation validation**

      The MAE provides a reasonable estimate of the actual error rates for discrete natural neighbour interpolation (Etherington, 2020). Therefore, even though each monthly weather grid has a matching uncertainty grid, we also used cross-validation to calculate the MAE for each monthly interpolation to validate the method and facilitate comparisons with other weather interpolations (Willmott and Matsuura, 2006).

120

      While the number of data cells (which equates closely to the number of weather stations) used during interpolation varied over time, the MAEs associated with all the weather variables remained reasonably constant around 20 mm for total rainfall and 1 °C for all temperature variables (Figure 4). This consistency suggests that the interpolation methodology was robust to reductions in available data. As might be expected, the MAEs were lowest when interpolating data during 1950-1980

125   which is the period the climatologies aiding the interpolation were created for. There was also an increase in MAE moving

back in time from 1950 to 1900 that becomes far more pronounced pre-1910 as weather data availability becomes extremely limited. Therefore, we conclude that the data are most reliable from 1910 to 2019, representing the four decades either side of the 1950-1980 climatologies used in the interpolations.

130 The spatial variability in monthly uncertainty (Figure 3) explains the variable pattern of MAE over time (Figure 4); while the number and location of weather stations may vary from month to month, the weather patterns may change dramatically making some months easier to interpolate than others. This variability of uncertainty further emphasises the need to provide spatially and temporally relevant estimates of uncertainty for all interpolations, and that globally aggregated uncertainty estimates may not provide users with a full picture of the data's limitations.

135

5. Using the uncertainty data

When using interpolated weather data users need to be able to make decisions specific to their requirements (de Vos et al., 2020). Therefore, in HOWNZ we have endeavoured to match every interpolation with an individual measure of uncertainty in the relevant units. However, we recognise that it is unusual for historical weather data to be presented alongside such
140 detailed uncertainty estimates. These uncertainty estimates provide a new analytical opportunity and challenge for potential data users, so to enable data users to leverage the uncertainty data we feel it is important to demonstrate how to incorporate the uncertainty data into an analytical workflow.

A simple example of weather time-series analysis would be to use a Spearman's rank (r_s) correlation (Gregory,
145 1978;Spearman, 1904) to detect the directionality of any trends in weather patterns over time (Girardin and Wotton, 2009). To incorporate uncertainty into this process, we adopt a Monte Carlo approach in which we produce many equally possible weather histories by randomly sampling each month's weather as a random value from a uniform distribution with a range equal to the interpolation uncertainty (though limiting rainfall to a minimum of 0 mm). Many trends can then be calculated, with their distribution used to infer the reliability of the analysis. The importance of incorporating uncertainty is evident
150 when comparing locations that have differing uncertainty levels. For a location with high uncertainty the possible weather histories can vary widely around the single interpolated weather history resulting in a wide distribution of possible trends (Figure 5a). In other instances, the trend may be stronger than the uncertainty meaning that while the strength of the trend varies its direction can be clearly established (Figure 5b). At locations where there is little uncertainty and hence minimal variation in possible weather histories the trend can be established precisely (Figure 5c).

155

If this trend analysis with uncertainty process is repeated for every location, spatial trends can be analysed. The challenge then becomes how to best visualise the spatial pattern of uncertainty. Based on guidance relating to the cartographic

visualisation of uncertainty (Kaye et al., 2012;Retchless and Brewer, 2016), we selected a diverging colour scheme to show trends as the median r_s of the possible weather histories, and a value-by-alpha approach (Roth et al., 2010) to mask locations

160   of increasing uncertainty measured as the 5th to 95th percentile range of the r_s of the possible weather histories. The resulting map indicates that in some regions there are clear trends in weather patterns over time, but in other regions the uncertainty is too large to reliably make inferences from them (Figure 6).

These results clearly demonstrate that it is possible to produce a long-term history of interpolated weather and emphasise the

165   importance of quantifying the uncertainty of all interpolations. It is obviously impossible for us to give guidance on how to incorporate uncertainty into all possible applications, but the Monte Carlo approach we present could be adapted to many situations. For example, the trend analysis shown here could be extended for those users interested in comparing long-term environmental data to weather data to identify associations between an environmental process and weather (Girardin and Wotton, 2009).

170

6. Limitations and future recommendations

Some aspects of the temporal and spatial scales of HOWNZ could be improved in subsequent refinements of this dataset. One limitation is that the monthly temporal resolution does not capture extreme but short duration weather events. For example, in July of 1996 there was an extreme cold snap that had significant effects on vegetation in southern New Zealand

175   (Bannister, 2003) but is not evident in our data that averages minimum temperatures over the whole month. We only had access to monthly climatologies on which to base our interpolations, but if openly available weekly, or even daily, climatologies were created then it would be possible to interpolate historical weather at a finer temporal resolution to better capture extreme weather events of short durations.

180   While improving temporal resolution would require the creation of new climatologies, the spatial resolution could be improved up to 100 m with the climatologies used here. This improvement could be beneficial in the mountainous areas of New Zealand where temperatures can vary considerably within the 1-km resolution of our grids. Future improvements in temporal and spatial resolution would benefit from a more efficient computational workflow. Our computational workflow limited our processing to a 1-km resolution, but future versions could either leverage high performance computing, or to

185   maintain a high degree of openness could continue to use desktop computing but with discrete natural neighbour interpolation leveraging the power of graphics processing units that are well suited to this method (Park et al., 2006).

We do not feel there is much point in trying to extend the temporal extent of the data set. The growth of MAE associated with the reduction of weather data pre-1910 (Figure 4) indicates the available weather station data available is insufficient to

190   reliably estimate historical weather using our approach before the 20th century, and that palaeo-environmental techniques may provide a better option for longer-term weather, or perhaps more realistically, climate information (Cook et al., 2006;Duncan et al., 2010). There was a subtle increase in MAE towards the present that may be a function of a reduction in weather station data, particularly rainfall data, and perhaps a growing temporal mismatch between the 1950-1980 climatologies used to aid the interpolation (Figure 4). This indicates that continuing to use interpolation methods may
195   become less feasible in the future. More recent climatologies could be produced to provide more temporally relevant climate data, but satellite data may provide a more useful data source for more recent weather history (Funk et al., 2015). We believe our interpolation approach is most valid for the pre-satellite era, and that by producing data that span part of the satellite era we provide a useful overlap for comparative purposes that could allow for a transition between historical weather data sources.

200

7. Data availability

The resulting HOWNZ data (Etherington et al., 2021) are openly available in non-proprietary file formats under a Creative Commons by Attribution 4.0 Licence and are archived at the Manaaki Whenua – Landcare Research DataStore https://doi.org/10.7931/zmvz-xf30.

205

8. Code availability

All HOWNZ analyses were conducted using an open-source software Python computational framework (Pérez et al., 2011) in conjunction with the gdal (GDAL/OGR contributors, 2021), pyproj (Snow et al., 2020), NumPy (Harris et al., 2020a), SciPy (Virtanen et al., 2020), numba (Lam et al., 2015), and Matplotlib (Hunter, 2007) packages. All the resulting code used
210   to process data and plot figures presented here (Etherington, 2021) is openly available under an MIT Licence from the Manaaki Whenua – Landcare Research DataStore https://doi.org/10.7931/yk7g-vz81.

9. Conclusions

We view HOWNZ as *a,* rather than *the,* history of New Zealand weather; it would be possible to repeat the process using
215   equally defensible quantitative methods and obtain different results. Likewise, changes in the spatial or temporal resolution will result in different patterns, but there is no single *best* resolution as this will vary depending on the desired application. Nevertheless, as HOWNZ matches the best available spatial and temporal extent and resolution of any currently available

open weather grids (Table 1), we believe that in creating HOWNZ we have significantly improved the ability for environmental and ecological scientists in New Zealand to understand how changing weather patterns have affected various environmental and ecological processes. Even with the spatially and temporally complex patterns of uncertainty, that are sometimes large, it is still possible to find consistent trends in weather (Figure 6). We also hope our efforts to produce interpolation estimates with associated uncertainty, and examples of how to build that uncertainty into any analyses, will encourage the quantification and visualisation of uncertainty in other weather and climate data sets.

Author contributions

TRE, GLWP, and JMW conceived and developed the idea. TRE developed the code and performed the data processing. TRE, GLWP, and JMW wrote the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by: the Strategic Science Investment Funding for Crown Research Institutes from the New Zealand Ministry of Business, Innovation and Employment's Science and Innovation Group; and The University of Auckland Faculty Research Development Fund 3702237.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., and Hegewisch, K. C.: TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015, Scientific Data, 5, 170191, 10.1038/sdata.2017.191, 2018.

Bannister, P.: Are frost hardiness ratings useful predictors of frost damage in the field? A test using damage records from the severe frost in South Otago and Southland, New Zealand, July 1996, N Z J Bot, 41, 555-569, 10.1080/0028825X.2003.9512869, 2003.

Cook, E. R., Buckley, B. M., Palmer, J. G., Fenwick, P., Peterson, M. J., Boswijk, G., and Fowler, A.: Millennia-long tree-ring records from Tasmania and New Zealand: a basis for modelling climate variability and forcing, past, present and future, J Quat Sci, 21, 689-699, 10.1002/jqs.1071, 2006.

245 de Vos, M. G., Hazeleger, W., Bari, D., Behrens, J., Bendoukha, S., Garcia-Marti, I., van Haren, R., Haupt, S. E., Hut, R., Jansson, F., Mueller, A., Neilley, P., van den Oord, G., Pelupessy, I., Ruti, P., Schultz, M. G., and Walton, J.: Open weather and climate science in the digital era, Geoscience Communication, 3, 191-201, 10.5194/gc-3-191-2020, 2020.

Duncan, R. P., Fenwick, P., Palmer, J. G., McGlone, M. S., and Turney, C. S. M.: Non-uniform interhemispheric temperature trends over the past 550 years, Climate Dynamics, 35, 1429-1438, 10.1007/s00382-010-0794-2, 2010.

250 Etherington, T. R.: Discrete natural neighbour interpolation with uncertainty using cross-validation error-distance fields, PeerJ Computer Science, 6, e282, 10.7717/peerj-cs.282, 2020.

Etherington, T. R.: Code for A History of Open Weather in New Zealand (HOWNZ), DataStore [code], https://doi.org/10.7931/yk7g-vz81, 2021.

Etherington, T. R., Perry, G. L. W., and Wilmshurst, J. M.: Data for A History of Open Weather in New Zealand (HOWNZ), DataStore
255 [data set], https://doi.org/10.7931/zmvz-xf30, 2021.

Fick, S. E., and Hijmans, R. J.: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas, International Journal of Climatology, 37, 4302-4315, 10.1002/joc.5086, 2017.

Foley, A. M.: Uncertainty in regional climate modelling: a review, Prog Phys Geogr, 34, 647-670, 10.1177/0309133310375654, 2010.

Funk, C., Verdin, A., Michaelsen, J., Peterson, P., Pedreros, D., and Husak, G.: A global satellite-assisted precipitation climatology, Earth
260 System Science Data, 7, 275-287, 10.5194/essd-7-275-2015, 2015.

GDAL/OGR Geospatial Data Abstraction software Library: https://gdal.org, 2021.

Girardin, M. P., and Wotton, B. M.: Summer moisture and wildfire risks across Canada, Journal of Applied Meteorology and Climatology, 48, 517-533, 10.1175/2008JAMC1996.1, 2009.

Gordo, O., and Sanz, J. J.: Phenology and climate change: a long-term study in a Mediterranean locality, Oecologia, 146, 484-495,
265 10.1007/s00442-005-0240-z, 2005.

Gregory, S.: Statistical Methods and the Geographer, 4th ed., Longman, London, 1978.

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.: Array programming with NumPy, Nature, 585, 357-
270 362, 10.1038/s41586-020-2649-2, 2020a.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Scientific Data, 7, 109, 10.1038/s41597-020-0453-3, 2020b.

Hofstra, N., Haylock, M., New, M., Jones, P., and Frei, C.: Comparison of six methods for the interpolation of daily, European climate data, Journal of Geophysical Research, 113, D21110, 10.1029/2008JD010100, 2008.

275 Hunter, J. D.: Matplotlib: a 2D graphics environment, Computing in Science & Engineering, 9, 90-95, 10.1109/MCSE.2007.55, 2007.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M.: Climatologies at high resolution for the earth's land surface areas, Scientific Data, 4, 170122, 10.1038/sdata.2017.122, 2017.

Kaye, N. R., Hartley, A., and Hemming, D.: Mapping the climate: guidance on appropriate techniques to map climate variables and their uncertainty, Geoscientific Model Development, 5, 245-256, 10.5194/gmd-5-245-2012, 2012.

280 Keller, V. D. J., Tanguy, M., Prosdocimi, I., Terry, J. A., Hitt, O., Cole, S. J., Fry, M., Morris, D. G., and Dixon, H.: CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications, Earth System Science Data, 7, 143-155, 10.5194/essd-7-143-2015, 2015.

Lam, S. K., Pitrou, A., and Seibert, S.: Numba: A LLVM-Based Python JIT Compiler, Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, 7, 1-6, 10.1145/2833157.2833162, 2015.

285 Leathwick, J., Morgan, F., Wilson, G., Rutledge, D., McLeod, M., and Johnson, K.: Land environments of New Zealand: a technical guide, Ministry for the Environment, Wellington, 2002.

Lyra, G. B., Correia, T. P., de Oliveira-Júnior, J. F., and Zeri, M.: Evaluation of methods of spatial interpolation for monthly rainfall data over the state of Rio de Janeiro, Brazil, Theoretical and Applied Climatology, 134, 955-965, 10.1007/s00704-017-2322-3, 2018.

McCarthy, J. K., Leathwick, J. R., Roudier, P., Barringer, J. R. F., Etherington, T. R., Morgan, F. J., Odgers, N. P., Price, R. H., Wiser, S.

290 K., and Richardson, S. J.: New Zealand Environmental Data Stack (NZEnvDS): A standardised collection of spatial layers for environmental modelling and site characterisation, N Z J Ecol, 45, 3440, 10.20422/nzjecol.45.31, 2021.

Muñoz Sabater, J.: ERA5-Land monthly averaged data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), [data set], https://doi.org/10.24381/cds.68d2bb3, 2019.

CliFlo: NIWA's National Climate Database on the Web: https://cliflo.niwa.co.nz/, access: 15 June 2020, 2020.

295 Park, S. W., Linsen, L., Kreylos, O., Owens, J. D., and Hamann, B.: Discrete Sibson interpolation, IEEE Transactions on Visualization and Computer Graphics, 12, 243-253, 10.1109/TVCG.2006.27, 2006.

Pérez, F., Granger, B. E., and Hunter, J. D.: Python: an ecosystem for scientific computing, Computing in Science & Engineering, 13, 13-21, 10.1109/MCSE.2010.119, 2011.

Remesan, R., Bellerby, T., and Frostick, L.: Hydrological modelling using data from monthly GCMs in a regional catchment, Hydrological

300 Processes, 28, 3241-3263, 10.1002/hyp.9872, 2014.

Retchless, D. P., and Brewer, C. A.: Guidance for representing uncertainty on global temperature change maps, International Journal of Climatology, 36, 1143-1159, 10.1002/joc.4408, 2016.

Roth, R. E., Woodruff, A. W., and Johnson, Z. F.: Value-by-alpha maps: an alternative technique to the cartogram, The Cartographic Journal, 47, 130-140, 10.1179/000870409x12488753453372, 2010.

305 Sibson, R.: A brief description of natural neighbour interpolation, in: Interpreting Multivariate Data, edited by: Barnett, V., Wiley, Chichester, 21-36, 1981.

Snow, A. D., Whitaker, J., Cochran, M., Van den Bossche, J., Mayo, C., de Kloe, J., Karney, C., Ouzounoudis, G., Dearing, J., Lostis, G., Heitor, Filipe, May, R., Itkin, M., Couwenberg, B., Berardinelli, G., The Gitter Badger, Eubank, N., Dunphy, M., Brett, M., Raspaud, M., Aurélio da Costa, M., Evers, K., Ranalli, J., de Maeyer, J., Popov, E., Gohlke, C., Willoughby, C., Barker, C., and Wiedemann, B. M.:

310 pyproj4/pyproj: 2.6.1 Release, [code], https://doi.org/10.5281/zenodo.3783866, 2020.

Spearman, C.: The proof and measurement of association between two things, The American Journal of Psychology, 15, 72-101, 10.2307/1412159, 1904.

Stewart, S. B., Elith, J., Fedrigo, M., Kasel, S., Roxburgh, S. H., Bennett, L. T., Chick, M., Fairman, T., Leonard, S., Kohout, M., Cripps, J. K., Durkin, L., and Nitschke, C. R.: Climate extreme variables generated using monthly time-series data improve predicted distributions

315 of plant species, Ecography, 44, 626-639, 10.1111/ecog.05253, 2021.

Tait, A., Henderson, R., Turner, R., and Zheng, X. G.: Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface, International Journal of Climatology, 26, 2097-2115, 10.1002/joc.1350, 2006.

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.,

320 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., Vijaykumar, A., Bardelli, A. P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C. N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D. A., Hagen, D. R., Pasechnik, D. V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G. A., Ingold, G.-L., Allen, G. E., Lee, G. R., Audren, H., Probst, I., Dietrich, J. P., Silterra, J., Webber, J. T., Slavič, J., Nothman, J., Buchner, J., Kulick, J.,

325 Schönberger, J. L., de Miranda Cardoso, J. V., Reimer, J., Harrington, J., Rodríguez, J. L. C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N. J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P. A., Lee, P., McGibbon, R. T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T. J., Robitaille, T. P., Spura, T., Jones, T. R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y. O., and Vázquez-Baeza, Y.: SciPy 1.0: fundamental algorithms for scientific computing in Python, Nature Methods, 17, 261-272,

330 10.1038/s41592-019-0686-2, 2020.

Willmott, C. J., and Robeson, S. M.: Climatologically aided interpolation (CAI) of terrestrial air temperature, International Journal of Climatology, 15, 221-229, 1995.

Willmott, C. J., and Matsuura, K.: On the use of dimensioned measures of error to evaluate the performance of spatial interpolators, International Journal of Geographical Information Science, 20, 89-102, 10.1080/13658810500286976, 2006.

335 Zhang, J., and Goodchild, M.: Uncertainty in Geographical Information, Taylor & Francis, London, 2002.

Table 1: Open access datasets providing a time-series of spatial grids that estimate precipitation and temperature that include New Zealand.

Dataset	Temporal extent	Temporal resolution	Spatial extent	Spatial resolution	Uncertainty grid
Worldclim 2 (Fick and Hijmans, 2017)	1960-2018	Monthly	Global	≈ 4 km	No
Chelsa (Karger et al., 2017)	1979-2013	Monthly	Global	≈ 1 km	No
TerraClimate (Abatzoglou et al., 2018)	1958-2015	Monthly	Global	≈ 4 km	No [1]
ERA5-Land (Muñoz Sabater, 2019)	1981-2021	Monthly	Global	≈ 9 km	No [1]
CRU TS version 4 (Harris et al., 2020b)	1901-2018	Monthly	Global	≈ 50 km	No [1]
HOWNZ	1910-2019	Monthly	New Zealand	1 km	Yes

340   [1] An uncertainty grid is provided to aid a user to make a personal uncertainty assessment but it is not in the units of measurement.

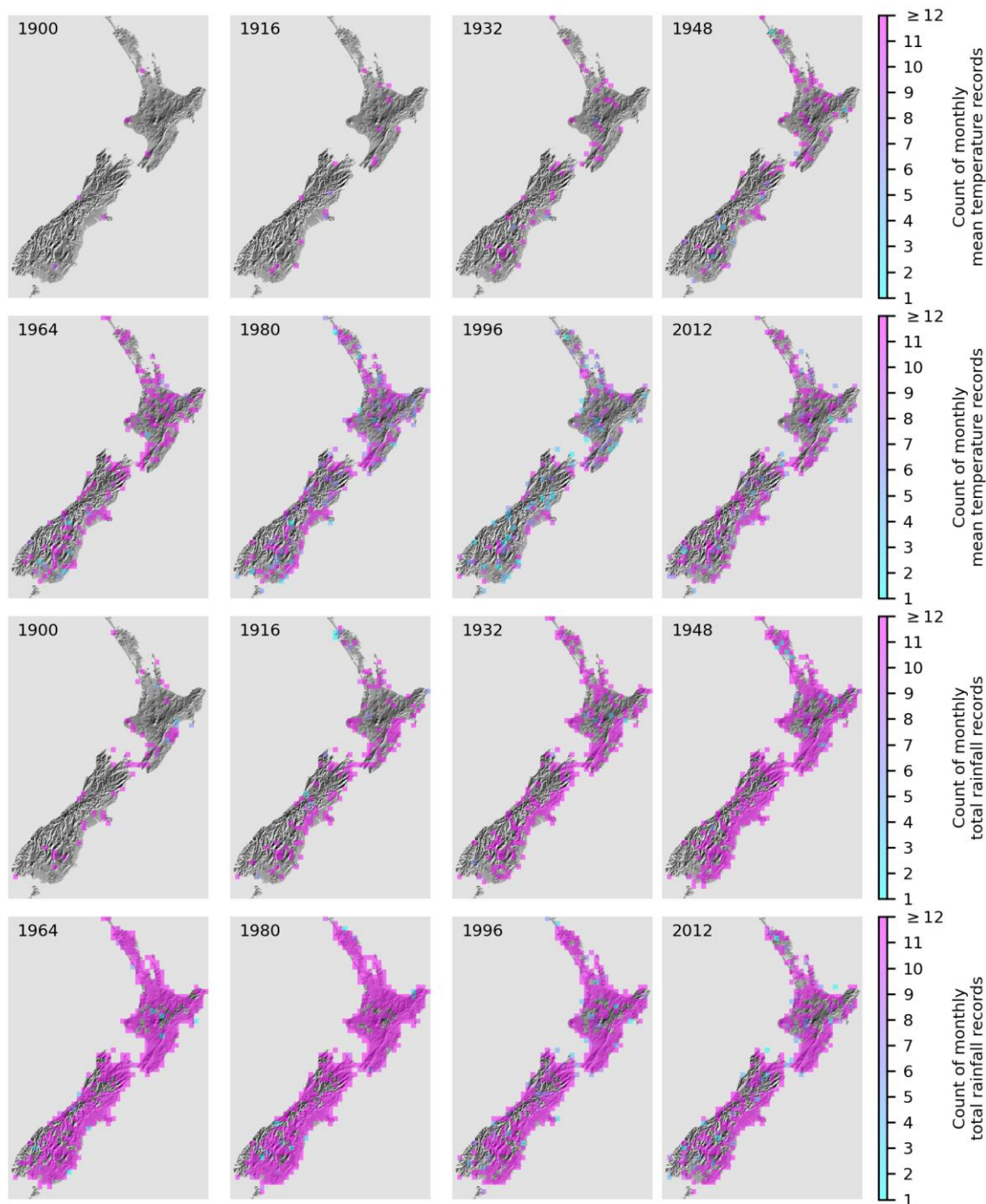



Figure 1: A time-series of the total number of monthly weather records for each year within 25 × 25 km grid cells for mean air temperature (that are similar to mean daily maximum air temperature and mean daily minimum air temperature) and total rainfall. Cells with ≥ 12 records will usually mean at least one weather station is present will records for all 12 months of the year, but in some instances many weather stations may be present resulting in tens or hundreds of records per grid cell.

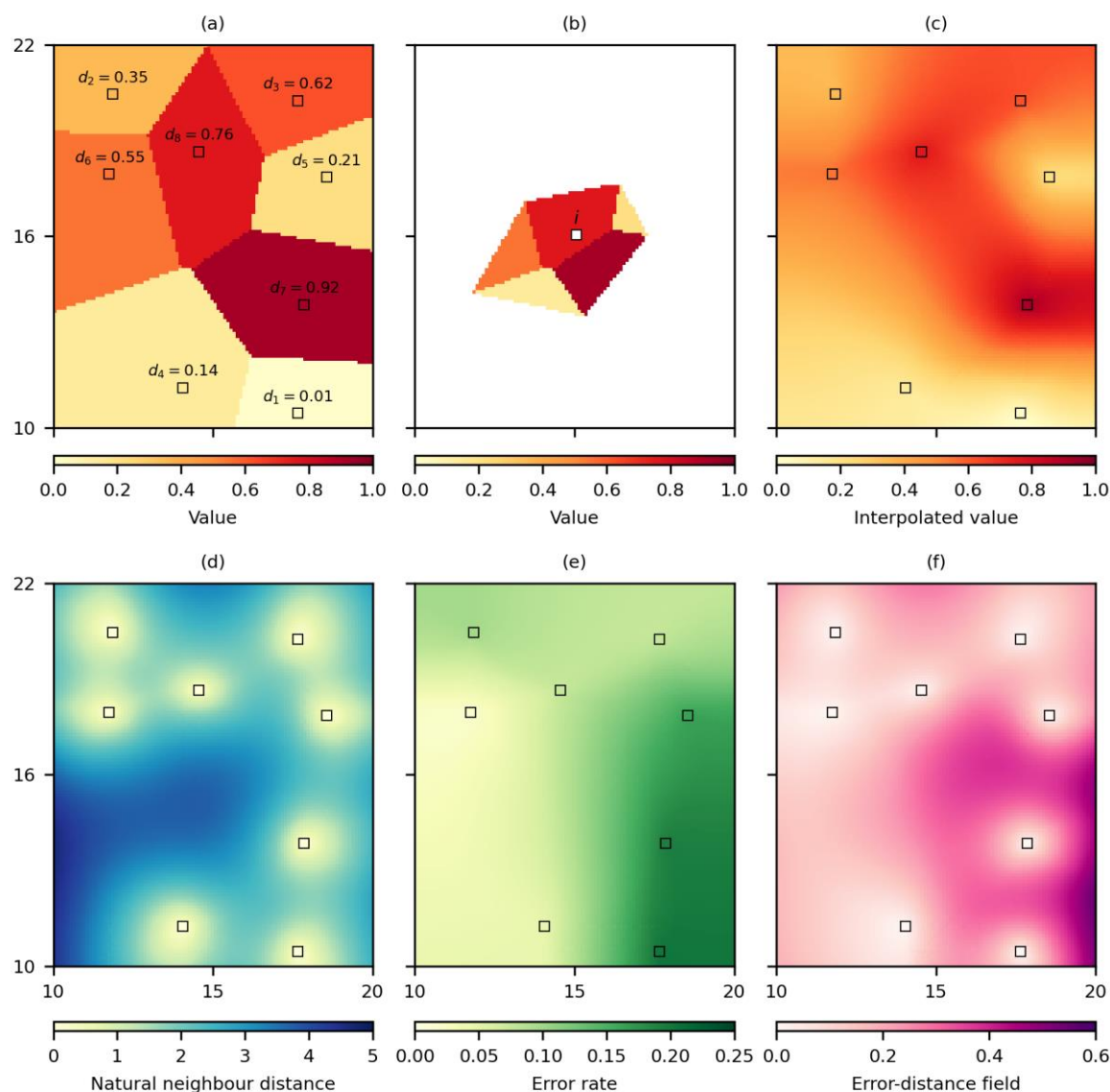



Figure 2: An example of discrete natural neighbour interpolation with a cross-validation error-distance field. Having defined a grid of cells (a) each cell is given the value of the nearest data cell *d*, (b) for any interpolation cell *i* the interpolated value is the mean of cell values that are as close or closer to the interpolation cell than a data cell, that (c) when repeated for all grid cells produces a natural neighbour interpolation. Natural neighbour interpolation is also used to interpolate (d) the distances to the data cells, and (e) the cross-validated error rate of each data cell. The interpolation uncertainty is then (f) the cross-validation error distance field that is the product of the natural neighbour distances and error rates.

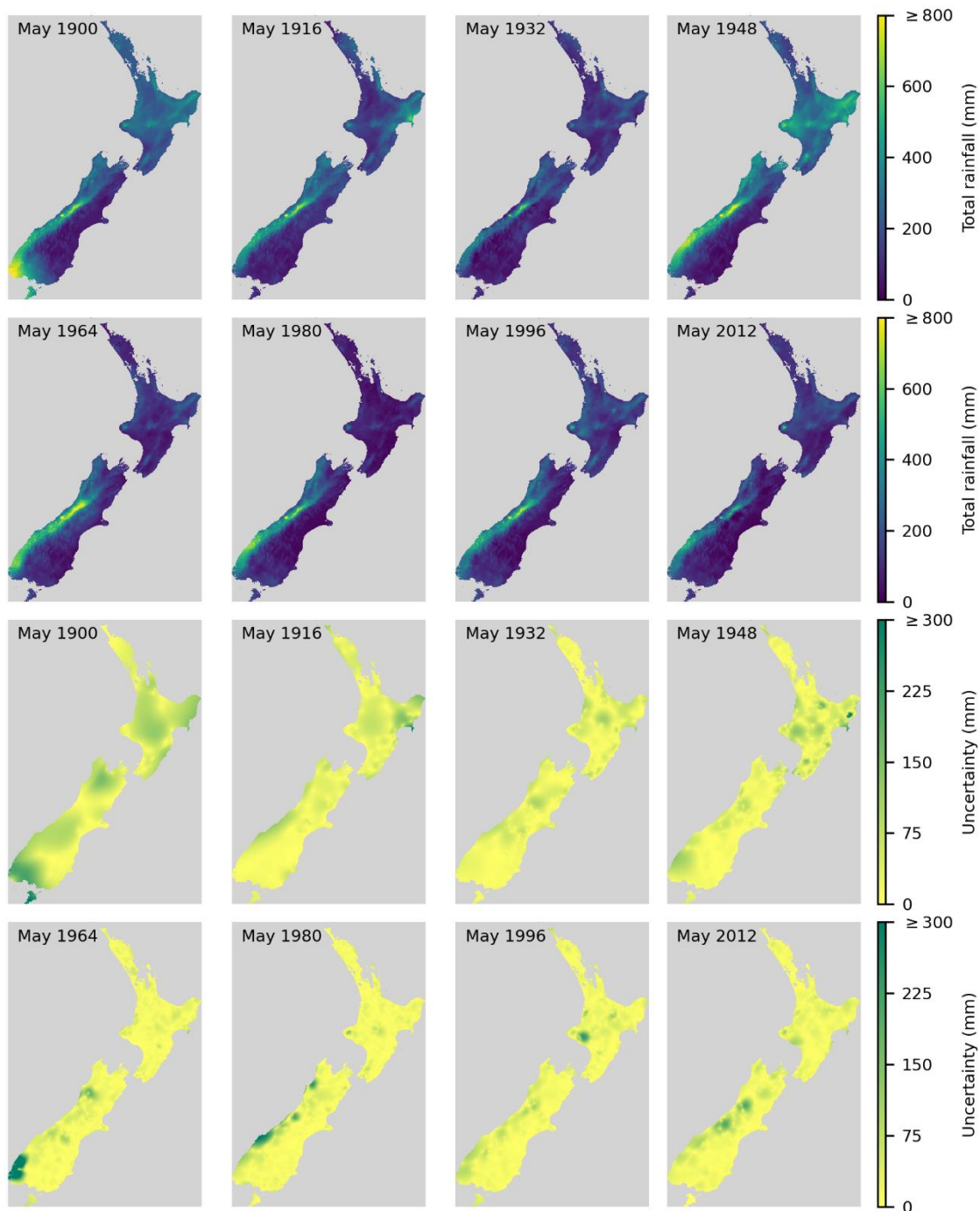



Figure 3: Example time-series of climatologically aided discrete natural neighbour interpolation of May total rainfall with associated uncertainty.

360

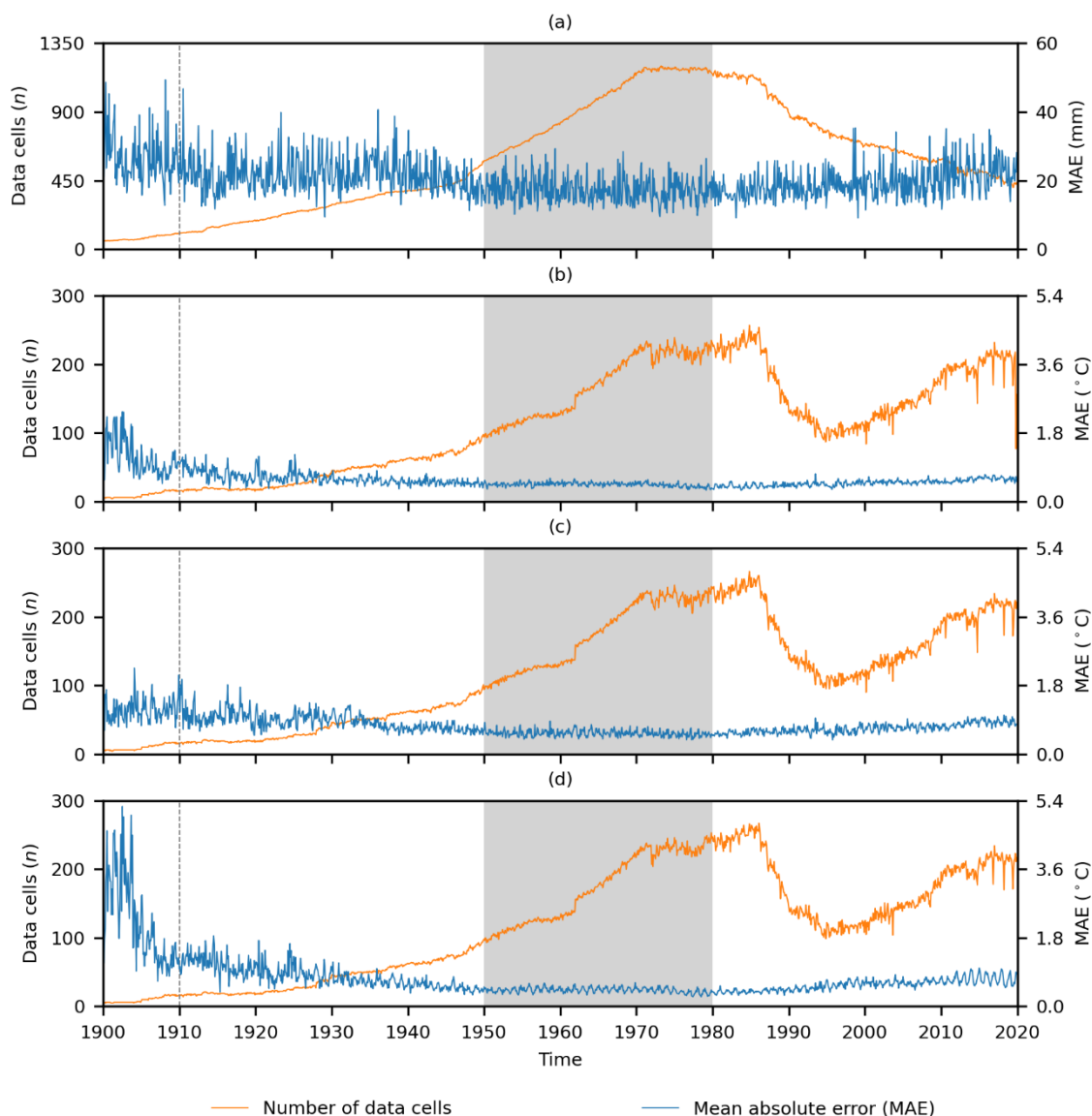



Figure 4: Monthly time-series of the number *n* of data cells (that relates closely to the number of weather stations) contributing to, and the cross-validated mean absolute error (MAE) of, climatologically aided natural neighbour interpolation across New Zealand for (a) total rainfall (mm), (b) mean air temperature (°C), (c) mean daily minimum air temperature (°C), and (d) mean daily maximum air temperature (°C). The grey areas show the period (1950-1980) over which the climatologies used to aid interpolation apply, and the dashed lines indicate the temporal limit of reliable data.

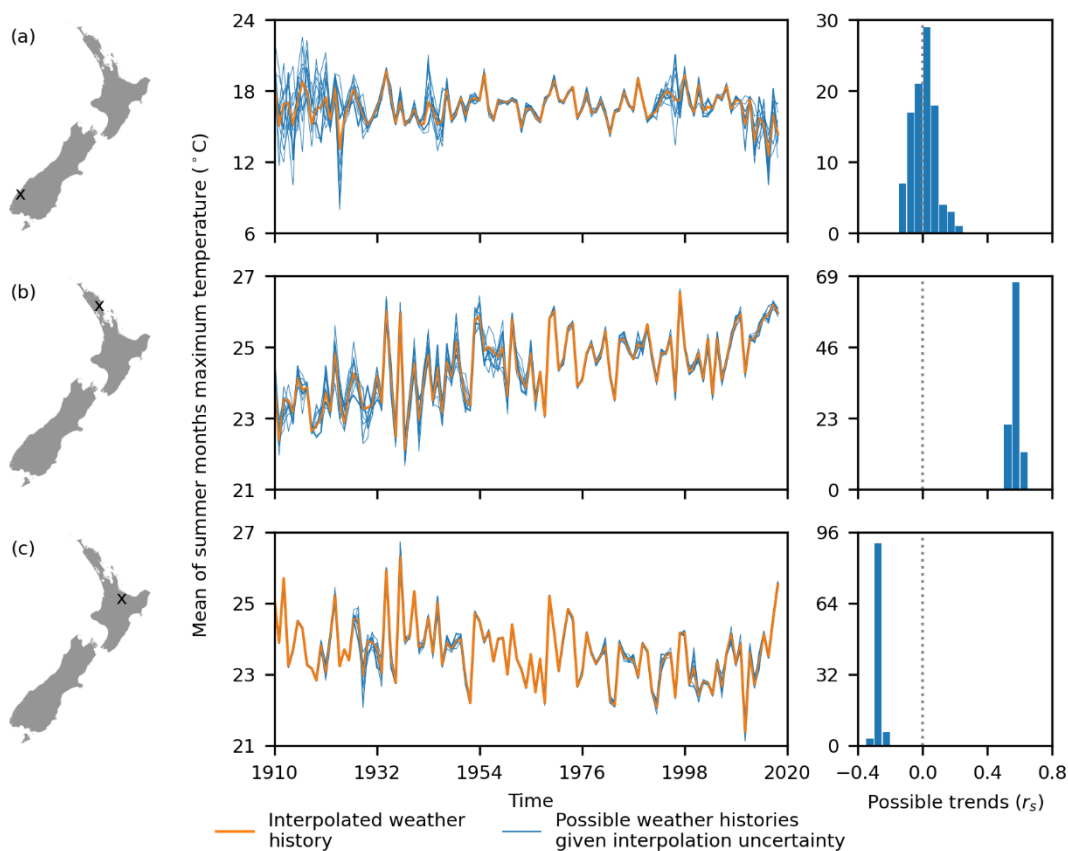



Figure 5: Examples of mean of summer months (December, January, and February) maximum temperature trend analysis that
incorporates interpolation uncertainty for three locations in New Zealand with (a) high uncertainty and no reliable trend, (b) some
uncertainty but a consistent positive trend, and (c) little uncertainty and a confident negative trend. For each location, the weather
history for the interpolated values is shown along with the first 10 of 100 possible weather histories that are within the range of
uncertainty for the interpolated value for each month. The distribution of trends for the 100 possible weather histories are shown
as Spearman's Rank (r_s) correlations between temperature and time.

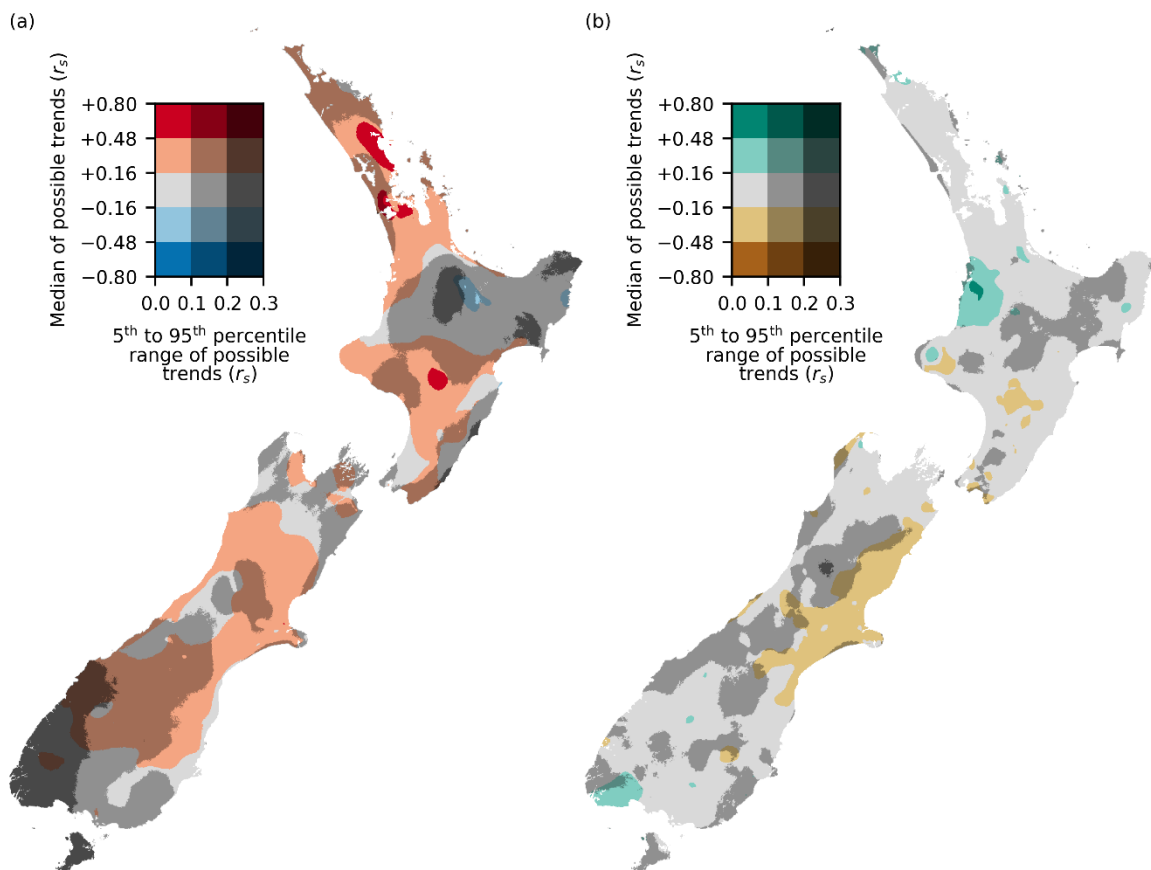
370

375

Figure 6: Spatial uncertainty analysis for the summer months (December, January, and February) trends in (a) mean maximum temperature, and (b) total rainfall. Trends were calculated as a Spearman's Rank (r_s) between the weather and time from 1910 to 2019 for 100 possible weather histories that were within the range of uncertainty for the interpolated values for each month. The median r_s value is visualised along with 5th to 95th percentile range of r_s values as a measure of uncertainty.

380