Gaussian process regression-based Bayesian optimisation (G-BO) of model parameters – a WRF model case study of southeast Australia heat extremes

*P. Jyoteeshkumar Reddy[1, *], Sandeep Chinta[2], Harish Baki[3], Richard Matear[1], John Taylor[4, 5]*

[1] Commonwealth Scientific and Industrial Research Organisation Environment, Hobart, TAS, Australia

[2] Center for Global Change Science, Massachusetts Institute of Technology, Cambridge, MA, USA

[3] Faculty of Civil Engineering and Geosciences, TU Delft, Delft, The Netherlands

[4] Commonwealth Scientific and Industrial Research Organisation Data61, Canberra, ACT, Australia

[5]Australian National University, Canberra, ACT, Australia

[*] Corresponding author: P. Jyoteeshkumar Reddy (jyoteesh.papari@csiro.au)

Key Points

- Our study optimises WRF model parameters for Southeast Australia heat extremes, enhancing the accuracy of the model simulation.
- G-BO method finds optimal parameter ranges, substantially improving the simulation of temperature and humidity.
- Results suggest updating WRF model's default settings for better extreme heat event simulations.

Abstract

In Numerical Weather Prediction (NWP) models, such as the Weather Research and Forecasting (WRF) model, parameter uncertainty in physics parameterization schemes significantly impacts model output. Our study adopts a Bayesian probabilistic approach, building on prior research that identified temperature (T) and relative humidity (Rh) as sensitive to three key WRF parameters during southeast Australia's extreme heat events. Using Gaussian process regression-based Bayesian Optimisation (G-BO), we accurately estimated the optimal distributions for these parameters. Results show that the default values are outside their corresponding optimal distribution bounds for two of the three parameters, suggesting the need to reconsider these default values. Additionally, the robustness of the optimal parameter distributions is validated by their application to an independent extreme heat event, not included in the optimisation process. In this test, the optimised parameters substantially improved the simulation of T and Rh, highlighting their effectiveness in enhancing simulation accuracy during extreme heat conditions.

Plain Language Summary

This study aims to enhance the accuracy of a numerical weather model called the Weather Research and Forecasting (WRF) model, especially for simulating extreme heat events in Southeast Australia. Typically, the accuracy of such models depends on specific settings, which are often set to default values. Our research used a method known as Gaussian process regression-based Bayesian Optimisation (G-BO) to determine the best range of values for these settings. We found that the default settings were not optimal. By applying G-BO, we identified more effective values that substantially improved the model's simulations of temperature and humidity during heat extremes. This improvement was consistent even when tested on an independent extreme heat event. These advancements are vital for more accurate weather forecasting, which is essential for emergency services, electricity management, and agriculture planning during extreme heat conditions.

1 Introduction

Recent studies highlight a significant increase in extreme weather globally, including intensified heatwaves that impact human and natural systems, especially in Southeast Australia (Reddy et al., 2021a; Masson-Delmotte et al., 2021; Perkins-Kirkpatrick and Lewis, 2020). Accurate heatwave simulations using Numerical Weather Prediction (NWP) models are essential in this context. The success of these models depends on their initial conditions and the representation of atmospheric processes, despite computational limitations (Bjerknes, 1910). Parameterisation in NWP models is a technique used to represent complex atmospheric processes that are too small-scale or intricate to directly resolve by the model. It involves simplifying these processes into manageable mathematical forms, often employing empirical or theoretical relationships. For example, processes like cloud formation and convection are

represented through parameterization schemes, which use a set of tuneable parameters. These parameters, often constants or exponents in model equations, are critical for the accuracy of simulations (Di et al., 2015; Yang et al., 2012). The Weather Research and Forecasting (WRF) model, noted for its adaptability and high-resolution capabilities (Evans et al., 2014; Skamarock et al., 2021), is widely used in Southeast Australia for forecasting and simulating extreme events. While the sensitivity of various physics parameterization schemes in these simulations has been explored (Evans et al., 2012; Ji et al., 2022), the specific influence of parameter values within these schemes is an area of active research, with the potential to further refine and improve model simulations.

Parameter optimisation is a process in which the model parameters are tuned to match the simulated output variables with respective observations. One of the primary challenges with optimisation is the exponential increase in complexity with an increase in the number of tuneable parameters, resulting in a "curse of dimensionality" (Duan et al., 2017, 2006). Another challenge is the number of output variables considered in the optimisation's objective function. These complexities make the optimisation process computationally demanding by making observational constraints inconsistent, by causing the parameters to be correlated and making the parameters poorly constrained (Matear, 1995). Therefore, several studies (Baki et al., 2022a; Chinta et al., 2021; Di et al., 2017, 2015; Ji et al., 2018; Quan et al., 2016; Yang et al., 2012) first performed a sensitivity analysis to identify the sensitive parameters that influence the output variables of interest. This helps reduce the number of parameters optimise, thereby reducing the computational costs.

Several studies (Baki et al., 2022b; Chinta and Balaji, 2020; Di et al., 2018) performed optimisation of WRF model parameters either for a single variable (single objective) using adaptive surrogate model-based optimisation (ASMO) (Wang et al., 2014) or for multiple variables using Multi-Objective ASMO (Gong et al., 2016) and knee point-based multi-objective optimisation (KMO) (Wang et al., 2023) algorithms. The main goal of these studies was to identify a single optimum value for each parameter that minimizes the simulation error with respect to observations. However, this approach often overlooks the natural predictive uncertainties and erroneously presumes that a unique, ideal set of parameter values is always applicable for all scenarios (Hoversten et al., 2006). It's important to recognize that a single optimal parameter value might not always be attainable but even when it is, the uncertainties involved could be substantial. Moreover, while approaches like Pareto front analysis in multi-objective optimisation reveal multiple near-optimal solutions, they too have limitations. Specifically, Pareto optimality focuses on finding a balance among competing objectives, which might not effectively capture the underlying uncertainties or the complexity of the parameter space (Gupta et al., 1998; Van Straten and Keesman, 1991). In this context, Bayesian optimisation offers a significant advantage. It provides a probabilistic framework that accounts for uncertainties and explores the parameter space more comprehensively, offering a range of solutions with quantified uncertainties (Beven and Binley, 1992). This approach not only acknowledges the complexity inherent in such models but also adapts more fluidly to varying scenarios, making it a more robust and flexible method for parameter optimisation.

Bayesian optimisation employs probabilistic methods to account for parameter uncertainties in models (Issan et al., 2023; Reiker et al., 2021; Xu et al., 2022; Chinta et al., 2023a). This approach represents input parameters as probability distributions from which multiple samples are drawn. These samples facilitate ensemble simulations, allowing the model to generate a

range of predictions. The model's outputs are then compared with actual observations using an objective function, refining the parameter distributions into more accurate posterior distributions. Subsequently, simulations based on these refined distributions align more closely with observed data. However, this method demands significant computational resources, as it involves numerous simulations of the WRF model. To address this, machine learning (ML) strategies, particularly ML-based surrogate models, are increasingly vital (Chinta et al., 2023b; Reddy et al., 2024; Wang et al., 2020). Once trained on a subset of existing simulations to understand the complex relationships between input parameters and outputs, surrogate models efficiently help explore the parameter space for Bayesian optimisation.

This study aims to optimise the WRF model parameters that influence different output variables corresponding to heat extremes using Bayesian optimisation. We do this by focussing on Southeast Australia during two extreme heat events. This study is organized as follows: Section 2 describes the data, events selected, WRF model configuration, introduces how surrogate models were developed, and presents the methodology of Bayesian optimisation. Section 3 presents the results and a detailed discussion of the optimised parameters. Section 4 summarises the conclusions from this study.

2 Methods

2.1 WRF model configuration and selected extreme heat events

In the present study, the WRF model v4.4 (Skamarock et al., 2021) is adopted for the numerical simulations. The simulation domain is configured with a single domain (d01) across southeast Australia, as shown in Figure S1. The domain consists of 206×181 grid points in the horizontal direction, with a horizontal resolution of 12 km and 40 terrain-following σ vertical levels reaching up to the 50 hPa atmospheric level. The simulations are integrated with a time step of 72 seconds. For initial and lateral boundary conditions, the European Centre for Medium-Range Weather Forecast Reanalysis 5th generation data set (ERA5) (Hersbach et al., 2020) at a horizontal resolution of $0.25°$ and a six-hourly interval is employed. The WRF model output variables, namely temperature at 2m height (T) and relative humidity at 2m height (Rh), are obtained at hourly intervals. This work extends our previous study (Reddy et al., 2024), where only three model parameters were identified to influence meteorological variables significantly during extreme heat events over southeast Australia. Consistent with our previous work, we adopt the same physics schemes as described in Table 1 of (Reddy et al., 2024). The description of three sensitive parameters is presented in the supplementary table S1.

The present study selected two southeast Australian extreme heat events like the previous study (Reddy et al., 2024) for the parameter optimisation. The first event spans 13 days, encompassing the heatwave period from January 26th, 12 UTC to February 8th, 12 UTC of 2009. The second extreme heat event simulation covers 15 days from December 16th, 12 UTC to December 31st, 12 UTC of 2019. Further, to assess the robustness of the optimised parameters, we consider an additional extreme heat event of 2013 covering the heatwave from 01st Jan 12 UTC to 18th Jan 12 UTC over southeast Australia. For all the selected events, a 36-hour model spin-up is considered. The simulation results are compared against hourly data

from the Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia (BARRA2; (Su et al., 2022)) at 12 km horizontal resolution.

2.2 Gaussian Process Regression-based Bayesian Optimisation (G-BO) using Markov Chain Monte Carlo sampling

We employ the Gaussian Process Regression-based Bayesian Optimisation (G-BO) methodology to obtain the optimal parameter distributions of sensitive WRF model parameters in simulating the critical meteorological variables of extreme heat events, such as temperature (T) and relative humidity (Rh) at 2m height. In this approach, first, we generate 128 parameter samples across the parameter space of three sensitive parameters utilizing the Quasi Monte-Carlo (QMC) Sobol sequence design, facilitated by the Uncertainty Quantification Python Laboratory (UQ-PyL) package (Wang et al., 2020). Then, the 128 WRF simulations were performed based on the generated parameter samples. Next, we compute the mean absolute error (MAE) of T and Rh between the WRF simulations and BARRA2 data. The MAE is normalized with respect to the default WRF simulation MAE as follows:

$$Normalised\ MAE\ (nMAE) = \frac{MAE\left(WRF_{p-runs}, BARRA2\right)}{MAE\left(WRF_{default}, BARRA2\right)} \quad (1)$$

where WRF_{p-runs} is each of the 128 parameter sample WRF runs, and $WRF_{default}$ is the default parameter WRF simulation. Any value of nMAE < 1 implies that the parameter sample is better than default.

We then train a surrogate model based on the generated parameter sample WRF simulations. Following the previous studies, we considered the Gaussian Process Regression (GPR; (Williams and Rasmussen, 1995, 2006)) as a surrogate model training with parameter samples as input and nMAE as target. The accuracy of the trained GPR model is evaluated through K-fold cross-validation (here, K=8), and the dependence on sample size is illustrated in Figure S2. The results indicate that the 128 samples are adequate for GPR training in achieving good accuracy with a goodness of fit (R^2) value of 0.99. Subsequently, the trained GPR is used to estimate the objective function (nMAE) in performing the optimisation of model parameters.

Parameter optimisation is broadly classified into the frequentist deterministic approach and the Bayesian probabilistic approach. In the frequentist deterministic approach, the goal of optimisation is to find a single optimal parameter value; however, Bayesian optimisation estimates the optimal distribution providing the uncertainty associated with the model parameters. Bayesian parameter optimisation is a process of learning the optimal distributions of model parameters based on Bayes' theorem, given the observational data. In the Bayesian approach, first, we choose the prior distribution; here, we consider it to be a uniform distribution that provides equal importance to all the values in the parameter range. Smith (2013) suggests using the non-informative prior (such as uniform distribution) if there isn't accurate prior information. Next, the selection of likelihood function, here, it is the normalized mean absolute error (nMAE) based on the previous studies (Wang et al., 2023). Finally, the posterior distribution of parameters is estimated by Bayes' theorem:

$$P(x/z) = \frac{p(z/x)\, p(x)}{p(z)} \quad (2)$$

where $p(x)$ is the prior, $p(z/x)$ is the likelihood, $p(z)$ is the marginal likelihood or normalising constant, x is the parameter sample of the random variable X, and z is the observation sample of the random variable Z. In the Bayesian framework, directly computing the marginal likelihood, $p(z)$, is often impractical due to its complexity, but this does not compromise the estimation of the posterior distribution. The focus, instead, is on employing Markov Chain Monte Carlo (MCMC) sampling algorithms. These methods effectively estimate the posterior distribution $P(x/z)$ by bypassing the explicit calculation of the marginal likelihood. This approach avoids the potential biases that can arise from improper definition or calculation of the marginal likelihood (Issan et al., 2023).

MCMC sampling systematically draws a representative set of samples from the target posterior distribution by constructing the Markov Chain. Here, the drawn sample from the probability distribution depends on the previously drawn sample. As the number of samples increases, the chain converges to the desired target posterior distribution (Roberts and Rosenthal, 2004). There are many MCMC algorithms, each considering different ways of constructing the Markov Chain while sampling, such as Gibbs sampling, Metropolis-Hastings algorithm, and Affine invariant ensemble sampling. Out of these sampling algorithms, previous studies recommended the Affine invariant ensemble sampling because it reaches faster convergence by considering the ensemble of chains in parallel, invariant to the affine transformations of parameters, enabling easy sampling from anisotropic probability distributions and has only two hyperparameters (one is number of walkers (i.e., ensemble of chains) and the other is stretch move (updates the next step of a given walker)) (Goodman and Weare, 2010; Issan et al., 2023).

Affine invariant ensemble sampling is an ensemble-based extension of the most widely used Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953). Briefly, Metropolis–Hastings algorithm involves iteratively drawing samples from a chosen prior-based proposal density centred (Gaussian distribution) around the previous sample, accepting or rejecting these samples based on defined probabilities to build the Markov chain. The use of a single chain in the Metropolis–Hastings algorithm is extended to an ensemble of chains in the Affine invariant ensemble sampling technique, which can be run in parallel for faster convergence. In this study, we implemented the Affine invariant ensemble sampling using the "emcee" Python package by choosing 50 walkers and stretch move as 2 (for more detailed description, refer to Mackey et al. (2013)). The MCMC sampling is sensitive to the initial point, where a low probable initial condition could be considered, which might not be representative of the target posterior distribution. Hence, the few initial samples were disregarded until the chain reached the stationary distribution, which is referred to as burn-in. In this study, an initial 1000 steps were chosen as burn-in, after which the chains converge (Fig. S3). Following the Mackey et al. (2013), we run the chains to 3000 steps (i.e., around 50 times the integrated autocorrelation time (which is around 50)) to ensure the convergence of chains to the target distribution (Fig. S3). More information about the autocorrelation times can be found in Goodman & Weare (2010) and Mackey et al. (2013).

3 Results and discussion

3.1 Gaussian process regression based-Bayesian Optimisation (G-BO) of parameters for improving temperature and relative humidity

G-BO results of the three sensitive parameters in calibrating the hourly temperature (T) and relative humidity (Rh) individually and the T and Rh combined are presented in figure 1. The most probable optimal values of the scattering tuning parameter (P14) are towards the lower end of the parameter range for all the three cases of optimisation (only T (Fig. 1(a)), only Rh (Fig. 1(d)), and T and Rh combined (Fig. 1(g))). Here, the default value of the P14 is outside the 2σ interval of the optimal posterior, which clearly suggests that the default value is not the best for providing accurate information of T and Rh during extreme heat events over southeast Australia. Multiplier for the saturated soil water content parameter (P17) posterior distribution resembles a Gaussian for all three optimisation cases (T (Fig. 1(b)), Rh (Fig. 1(e)), and both T and Rh (Fig. 1(h))), with a mean value around 0.42, 0.37, and 0.40 (normalised values) when optimised for only T, only Rh, and both T and Rh, respectively. Here, the default P17 value is outside the 2σ of posterior for T; however, it is within the 2σ interval when optimised for only Rh and on the lower end of the posterior with less probability when optimised T and Rh combined. This suggests that the default value of P17 is less likely to accurately simulate the T and Rh during heat extremes in southeast Australia. Similar to P14, profile shape exponent for calculating the momentum diffusivity coefficient parameter (P22) posterior has high densities towards the lower bound of the parameter range when optimised for T (Fig. 1(c)), Rh (Fig. 1(f)) individually, and T and Rh combined (Fig. 1(i)). The default value of P22 is not in the 2σ interval of optimal posterior, suggesting the default value should be reconsidered for this parameter.

The mean (and 2σ confidence interval) nMAE of G-BO posterior distribution of optimised parameter combinations for T and Rh combined case is 0.867 (0.863, 0.874) for T and 0.928 (0.924, 0.934) for Rh. Further, we compare the MAE spatial patterns of T and Rh between the default and optimised parameter distribution combinations, as shown in figure 2. The MAE of default or optimised parameter combination is calculated with respect to BARRA2 data. We randomly sample 10 parameter sets from the G-BO posterior distribution (T and Rh combined) as the representative sample (see Table S2), and the MAE of the ensemble mean of 10 runs computed with respect to BARRA2 data is shown for the spatial comparison. Figures 2(a) and 2(d) show the average T and Rh, respectively, during all days of both events (2009 and 2019) using the BARRA2 data. The optimised ensemble mean reduced the MAE of T mostly across the domain compared to the default simulation (compare Figs. 2(b) vs. 2(c)). Particularly, the substantial reductions were seen in the regions of high average temperatures (greater than 33 °C, see in Fig. 2(a)) and in the northeast parts of the domain (Fig. 2(c)). Similar to T, optimised ensemble mean improved the simulation of Rh compared to the default over the regions of the low average Rh i.e., central parts of the domain and across the regions of average high Rh i.e., northeast coast of the domain (compare Figs. 2(e) vs. 2(f)). Previous studies have also shown that the default parameter set has a substantial (cold) temperature and (wet) precipitation bias over southeast Australia, broadly consistent with the current results (Di Virgilio et al., 2019; Ji et al., 2022; Kala et al., 2015). The optimised ensemble mean improves the prediction of T and Rh mostly across the domain, particularly over the east coast and the northeast parts of the domain where the substantial biases observed in the default simulation.

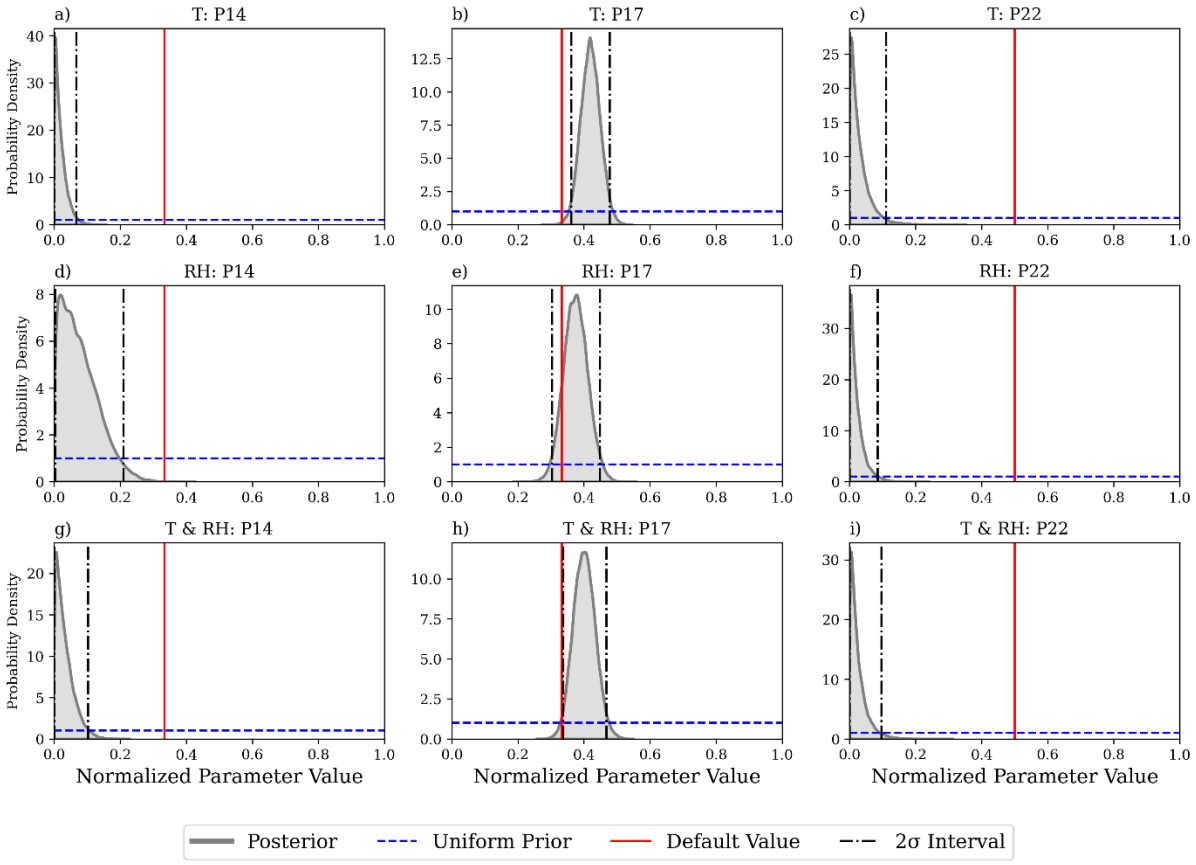

Fig. 1 *Bayesian optimised posterior distribution (grey shading) of sensitive parameters (presented as normalised values) for hourly temperature (T) (a-c) and relative humidity (Rh) (d-f) individually and for both T and Rh (g-i) combined. The red and blue lines show the default and uniform prior distribution, respectively. The grey dashed lines show the 2σ interval (95%) of the optimised posterior distribution values.*

Daily maximum temperature (T_{max}) and daily minimum relative humidity (Rh_{min}) are the critical meteorological variables considered for identifying and quantifying the heat extremes, particularly the dry heat, over southeast Australia (Abram et al., 2021; Reddy et al., 2021b). Hence, we compare the spatial patterns of the T_{max} and Rh_{min} mean during all days of the two selected events (Fig. 3(a-c and f-h)) and only on the extremely hot days of each event (2009 (Fig. S4) and 2019 (Fig. S5)). The optimised ensemble mean simulation improved the realism of the mean T_{max} by around 2.5 °C (compare Fig. 3(b) vs. (c)) and Rh_{min} (compare Fig. 3(g) vs. (h)) by approximately 0.1 over the domain compared to the default parameter values. Further, we compare the domain area-average time series of daily maximum temperature (T_{max}) during all days of both the events (2009 and 2019) between BARRA2, default, and optimised ensembles (Fig. 3(d)). Results show that the optimised ensemble clearly improved the accuracy of T_{max} across all days of the two selected events compared to the default (Fig. 3(d)). Further, the ensemble spread of domain average bias of T_{max} and Rh_{min} across all the days of both events (2009 and 2019) is compared with the default domain average bias (Fig. 3(e) and (j)). This shows that all the 10 optimised ensemble members clearly improved the cold T_{max} and wet Rh_{min} bias compared to the default. Furthermore, the optimised ensemble accurately simulated

the hot region (region of T_{max} greater than 44 °C) on the extremely hot day of both the selected events (2009 (compare Fig. S4(b) vs (c)) and 2019 (compare Fig. S5(b) vs (c))) compared to the default. This suggests that the optimised ensemble simulations better capture the extremeness of the extreme heat events over southeast Australia compared to the default. The accurate information on extremeness of the extreme heat events is critical for planning emergency services, electricity demand management, and cattle safety and crop management (Asseng et al., 2011; Lindstrom et al., 2013; Loridan et al., 2016).

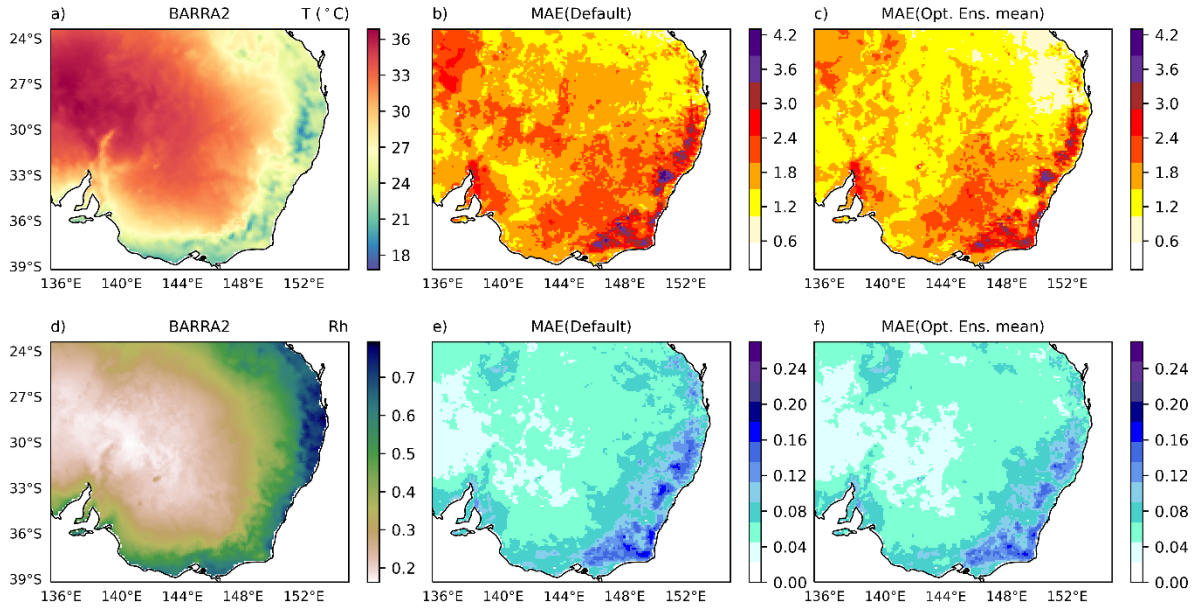

Fig. 2 *Spatial plot of average hourly temperature (T; °C) (a) and hourly relative humidity (Rh) (d) during all days of both selected events (2009 and 2019) using the BARRA2 data. MAE of the WRF default parameters run (default) and optimised ensemble mean (of randomly drawn 10 parameter combinations from the optimal posterior distribution of both T and Rh combined) parameters run (Opt. Ens. mean) with respect to BARRA2 data for the considered meteorological variables. The MAE of T (b-c) and Rh (e-f) for default and Opt. Ens. mean runs with respect to BARRA2.*

The G-BO optimised parameter distributions are further tested on an independent extreme heat event (2013 event) not used in the optimisation. Similar to the 2009 and 2019 events, the 2013 event optimised ensemble mean improves the simulation of T and Rh mostly across the domain, particularly over the east coast and the northeast regions of the domain, compared to the default parameters (Figs. S6 and S7). Further, the T_{max} and Rh_{min} results of the 2013 event show that the optimised ensemble improves the simulation compared to the default, which is consistent with the two optimised events (Fig. S7). This further testing help demonstrates the robustness of the G-BO results. Overall, this study's G-BO methodology improved the simulation accuracy of T and Rh during heat extremes, specifically bettered the extremeness of the extreme heat information.

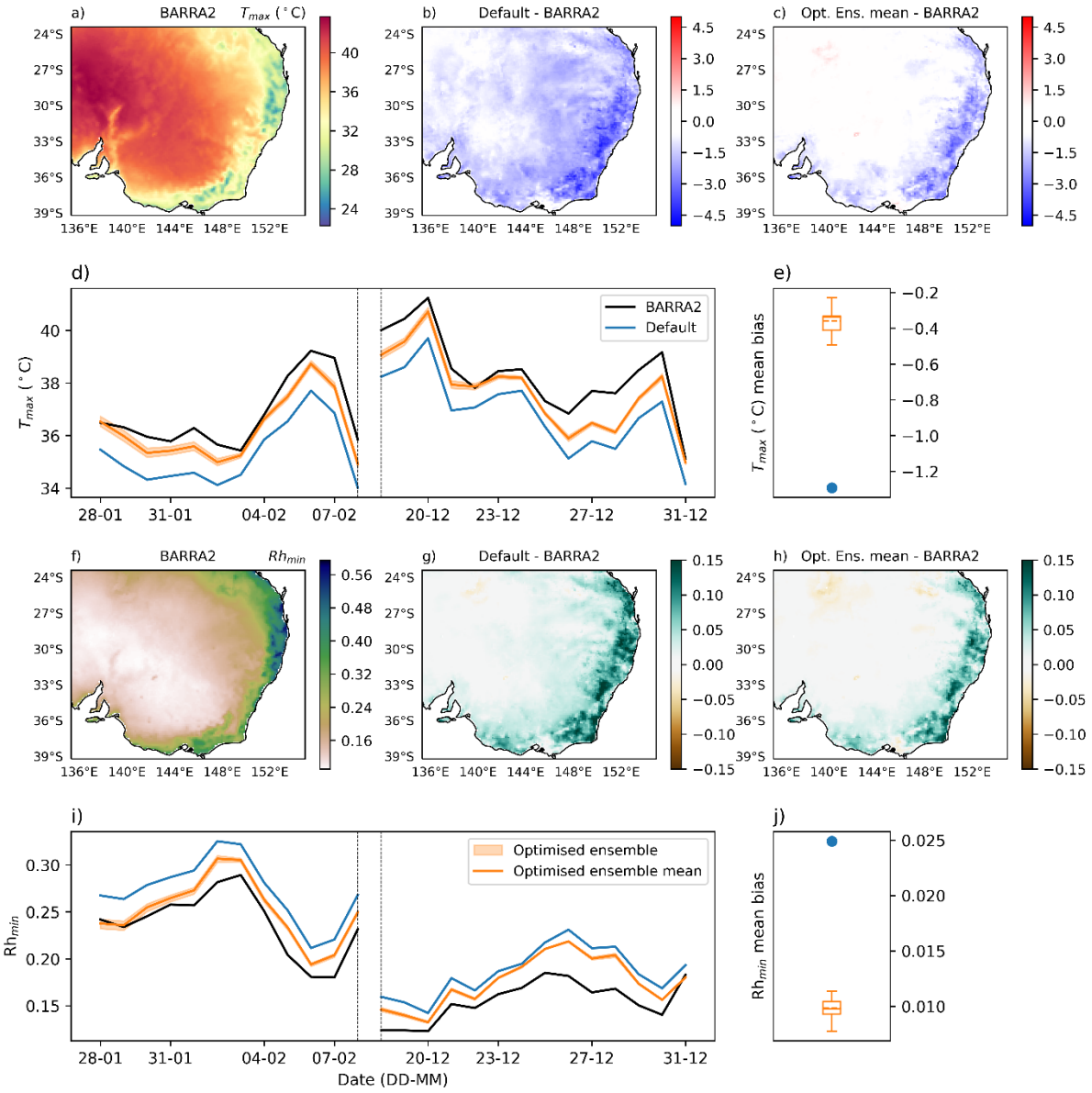

Fig. 3 *Spatial plot of average daily maximum temperature (T_{max}; °C) (a) and daily minimum relative humidity (Rh_{min}) (f) during all days of both selected events (2009 and 2019) using the BARRA2 data. Comparison of the WRF default parameters run (default) and optimised ensemble mean (Opt. Ens. mean) (of randomly drawn ten parameter combinations from the optimal posterior distribution of both T and Rh combined) with respect to BARRA2 data for the considered meteorological variables. The mean bias of T_{max} (b-c) and Rh_{min} (g-h) between default and Opt. Ens. mean runs with respect to BARRA2. Domain average temporal comparison of daily maximum temperature (T_{max}; °C) (d), and daily minimum relative humidity (Rh_{min}) (i) of BARRA2 (black line), default (blue line), optimised ensemble (orange shading), and optimised ensemble mean (orange line) during all days of 2009 and 2019 events (events are separated with dotted vertical lines). Box plots of domain average bias of optimised ensemble with respect to BARRA2 and the default domain average bias value is shown as a blue dot (T_{max}(e) and Rh_{min}(j)).*

3.2 Physical understanding of optimal distribution of parameters

The scattering tuning parameter (P14) optimal posterior is towards the lower bound of its range with a maximum likelihood value around 0.5×10^{-5} to 0.6×10^{-5}, which is lower than the default value (1×10^{-5}). The lower the P14, the weaker the scattering, leading to more incoming solar radiation, which increases the surface heating and can amplify the daytime surface temperature (Dudhia, 1989; Montornès Torrecillas et al., 2015). This supports our result of lower P14 compared to the default as the optimal, which improves the temperature cold bias of default parameter simulation (Fig. 3(b) vs. (c)). This is more specifically seen in the T_{max} because it is much affected by the P14 (Fig. 3(b) vs. (c)) (Reddy et al., 2024). Low P14 values favour low Rh_{min}; our results agree with this and show that the positive Rh_{min} bias in default simulation is improved in the optimised run with the P14 value lower than the default one (Fig. 3(g) vs. (h)).

The multiplier for saturated soil water content or soil porosity P17 in the land surface scheme is optimised to the posterior, with the maximum likelihood value ranging between 1.03-1.18, which is slightly higher than the default value (1). Our G-BO results show a higher T_{max} with the calibrated parameter set compared to the default parameter simulation. Consistent with the results, previous studies suggest that low P17 favours a decrease in surface temperature, particularly during the daytime (Fonseca et al., 2019; Reddy et al., 2024; Temimi et al., 2020). Next is parameter P22, which is the profile shape exponent in the momentum diffusivity coefficient of the planetary boundary layer scheme. P22 optimised posterior maximum probability value is around 1.0 to 1.14, which is lower than the default value (2). Previous studies suggest that the low P22 weakens the turbulent mixing below the maximum height of momentum diffusivity, which may moderate the convective mass flux, leading to a lower Rh (Hong et al., 2006; Oke, 2002).

Our study has focused on the critical task of quantifying parameter uncertainty and optimising parameter values relative to observations, which is fundamental for enhancing model reliability. It is important to recognize, however, that there are additional sources of uncertainty that also affect model accuracy. These include uncertainties in initial and boundary conditions, the accuracy of observational data, and inherent limitations within the model structure. While our results provide valuable insights for parameter optimisation, future studies could further improve model simulations by exploring these additional sources of uncertainty, thereby offering a more holistic approach to model accuracy and reliability.

4 Conclusions

We used the G-BO methodology to estimate the optimal distribution of the three WRF model parameters previously identified as the most important to simulate extreme heat conditions in SE Australia (Reddy et al., 2024). The parameters, scattering tuning parameter (P14), the multiplier for saturated soil water content (P17), and the profile shape exponent for calculating the momentum diffusivity coefficient (P22), have the produced the greatest sensitivity to the simulated hourly temperature (T) and relative humidity (Rh) during the two considered southeast Australian extreme heat events (2009 and 2019). Unlike the previous studies, which focus on identifying single optimum parameter values, our methodology provides optimal

parameter distributions, which allows us to quantify the parameter uncertainty and parameter correlations.

The key results from the parameter optimisation are: 1) for two of the three parameters optimised the default WRF default parameter values lie outside the optimal range suggesting the need to reconsidering the parameter values for simulating heat extremes in this region. 2) Randomly drawing ten parameter samples from the optimal distributions improved the MAE of the simulated T and Rh by 11.2-12% and 5.4-6.8 %, respectively. 3) The mean spatial pattern of ten optimal parameter simulations improves the default simulation negative bias of T and positive bias of Rh mostly across the domain. Most importantly optimal parameter sample substantially improved critical variables of the dry heat, such as daily maximum temperature (T_{max}) and daily minimum relative humidity (Rh_{min}), compared to the default parameters over the domain. The changes in the optimised parameters from the default values are physically plausible and explainable from a physical parameterization perspective. An investigation of the optimal parameter distributions shows no correlations between the optimal parameters. A small spread in the ensemble from the optimised model indicates constrained parameter uncertainty (Fig. 3(d-e) and (i-j)). However, the discrepancies between the ensemble predictions and the observed data suggest that additional uncertainties from other sources are present within our model.

Further, to demonstrate the robustness of the optimal parameter distributions we use them to simulate a heat extreme event in 2013 in southeast Australia. The simulations improved representation of the T_{max} and Rh_{min}, over the default parameter values. Overall, G-BO methodology improved the simulation of T and Rh during heat extremes, specifically bettered the extremeness of the extreme heat information, which have significant implications for emergency services management and cattle and crop productivity. The present study results may quite not be applicable to wet extremes, which needs to be further explored and is clearly outside the scope of this study. Future studies can apply the present study's methodology to other extreme events such as extreme rainfall and tropical cyclones, to name a few. Further, this study's approach can be applicable to other dynamical models in the atmospheric, oceanic, and biological sciences, to name a few.

Acknowledgements

The authors would like to thank National Computing Infrastructure (NCI) Australia for providing computational resources. We would like to thank Chun-Hsu Su for providing the BARRA2 reanalysis data. We acknowledge the funding support of the CSIRO and Australian Climate Service.

Open Research

The ERA5 data is openly available at https://doi.org/10.24381/cds.adbb2d47. The source code of WRF v4.4 is openly available at https://github.com/wrf-model/WRF/releases/tag/v4.4. The

BARRA2 dataset is available from the NCI THREDDS data server https://dapds00.nci.org.au/thredds/catalogs/ob53/catalog.html. All the figures are generated with Python.

References

Abram, N.J., Henley, B.J., Sen Gupta, A., Lippmann, T.J.R., Clarke, H., Dowdy, A.J., Sharples, J.J., Nolan, R.H., Zhang, T., Wooster, M.J., Wurtzel, J.B., Meissner, K.J., Pitman, A.J., Ukkola, A.M., Murphy, B.P., Tapper, N.J., Boer, M.M., 2021. Connections of climate change and variability to large and extreme forest fires in southeast Australia. Commun Earth Environ 2, 8. https://doi.org/10.1038/s43247-020-00065-8

Asseng, S., Foster, I.A.N., Turner, N.C., 2011. The impact of temperature variability on wheat yields. Glob Chang Biol 17, 997–1012.

Baki, H., Chinta, S., Balaji, C., Srinivasan, B., 2022a. Determining the sensitive parameters of the Weather Research and Forecasting (WRF) model for the simulation of tropical cyclones in the Bay of Bengal using global sensitivity analysis and machine learning. Geosci Model Dev 15, 2133–2155.

Baki, H., Chinta, S., Balaji, C., Srinivasan, B., 2022b. Parameter Calibration to Improve the Prediction of Tropical Cyclones over the Bay of Bengal Using Machine Learning–Based Multiobjective Optimization. J Appl Meteorol Climatol 61, 819–837. https://doi.org/10.1175/JAMC-D-21-0184.1

Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol Process 6, 279–298.

Bjerknes, V., 1910. Dynamic Meteorology and Hydrography: Part [1]-2,[and atlas of plates]. Carnegie Institution of Washington.

Chinta, S., Balaji, C., 2020. Calibration of WRF model parameters using multiobjective adaptive surrogate model-based optimization to improve the prediction of the Indian summer monsoon. Clim Dyn 55, 631–650. https://doi.org/10.1007/s00382-020-05288-1

Chinta, S., Gao, X., Zhu, Q., 2023a. Machine Learning Assisted Bayesian Calibration of Model Physics Parameters for Wetland Methane Emissions: A Case Study at a FLUXNET-CH4 Site, in: NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning.

Chinta, S., Gao, X., Zhu, Q., 2023b. Machine Learning Driven Sensitivity Analysis of E3SM Land Model Parameters for Wetland Methane Emissions. arXiv preprint arXiv:2312.02786.

Chinta, S., Yaswanth Sai, J., Balaji, C., 2021. Assessment of WRF Model Parameter Sensitivity for High-Intensity Precipitation Events During the Indian Summer Monsoon. Earth and Space Science 8, e2020EA001471. https://doi.org/10.1029/2020EA001471

Di Virgilio, G., Evans, J.P., Di Luca, A., Olson, R., Argüeso, D., Kala, J., Andrys, J., Hoffmann, P., Katzfey, J.J., Rockel, B., 2019. Evaluating reanalysis-driven CORDEX regional climate models over Australia: model performance and errors. Clim Dyn 53, 2985–3005. https://doi.org/10.1007/s00382-019-04672-w

Di, Z., Duan, Q., Gong, W., Wang, C., Gan, Y., Quan, J., Li, J., Miao, C., Ye, A., Tong, C., 2015. Assessing WRF model parameter sensitivity: A case study with 5 day summer precipitation forecasting in the Greater Beijing Area. Geophys Res Lett 42, 579–587. https://doi.org/10.1002/2014GL061623

Di, Z., Duan, Q., Gong, W., Ye, A., Miao, C., 2017. Parametric sensitivity analysis of precipitation and temperature based on multi-uncertainty quantification methods in the Weather Research and Forecasting model. Sci China Earth Sci 60, 876–898. https://doi.org/10.1007/s11430-016-9021-6

Di, Z., Duan, Q., Wang, C., Ye, A., Miao, C., Gong, W., 2018. Assessing the applicability of WRF optimal parameters under the different precipitation simulations in the Greater Beijing Area. Clim Dyn 50, 1927–1948. https://doi.org/10.1007/s00382-017-3729-3

Duan, Q., Di, Z., Quan, J., Wang, C., Gong, W., Gan, Y., Ye, A., Miao, C., Miao, S., Liang, X., Fan, S., 2017. Automatic Model Calibration: A New Way to Improve Numerical Weather Forecasting. Bull Am Meteorol Soc 98, 959–970. https://doi.org/10.1175/BAMS-D-15-00104.1

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V, Gusev, Y.M., Habets, F., Hall, A., Hay, L., 2006. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. J Hydrol (Amst) 320, 3–17.

Dudhia, J., 1989. Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model. Journal of Atmospheric Sciences 46, 3077–3107. https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2

Evans, J.P., Ekström, M., Ji, F., 2012. Evaluating the performance of a WRF physics ensemble over South-East Australia. Clim Dyn 39, 1241–1258. https://doi.org/10.1007/s00382-011-1244-5

Evans, J.P., Ji, F., Lee, C., Smith, P., Argüeso, D., Fita, L., 2014. Design of a regional climate modelling projection ensemble experiment – NARCliM. Geosci Model Dev 7, 621–629. https://doi.org/10.5194/gmd-7-621-2014

Fonseca, R., Zorzano-Mier, M.-P., Azua-Bustos, A., González-Silva, C., Martín-Torres, J., 2019. A surface temperature and moisture intercomparison study of the Weather Research and Forecasting model, in-situ measurements and satellite observations over the Atacama Desert. Quarterly Journal of the Royal Meteorological Society 145, 2202–2220. https://doi.org/10.1002/qj.3553

Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Ye, A., Miao, C., Dai, Y., 2016. Multiobjective adaptive surrogate modeling-based optimization for parameter estimation of large, complex geophysical models. Water Resour Res 52, 1984–2008.

Goodman, J., Weare, J., 2010. Ensemble samplers with affine invariance. Comm App Math Comp Sci 5, 65–80.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resour Res 34, 751–763.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J.,

519 Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P.,
520 Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.N., 2020. The ERA5 global reanalysis. Quarterly
521 Journal of the Royal Meteorological Society 146, 1999–2049. https://doi.org/10.1002/QJ.3803

522 Hong, S.-Y., Noh, Y., Dudhia, J., 2006. A New Vertical Diffusion Package with an Explicit Treatment of
523 Entrainment Processes. Mon Weather Rev 134, 2318–2341.
524 https://doi.org/10.1175/MWR3199.1

525 Hoversten, G.M., Cassassuce, F., Gasperikova, E., Newman, G.A., Chen, J., Rubin, Y., Hou, Z., Vasco, D.,
526 2006. Direct reservoir parameter estimation using joint inversion of marine seismic AVA and
527 CSEM data. Geophysics 71, C1–C13.

528 Issan, O., Riley, P., Camporeale, E., Kramer, B., 2023. Bayesian Inference and Global Sensitivity
529 Analysis for Ambient Solar Wind Prediction. Space Weather 21, e2023SW003555.
530 https://doi.org/10.1029/2023SW003555

531 Ji, D., Dong, W., Hong, T., Dai, T., Zheng, Z., Yang, S., Zhu, X., 2018. Assessing Parameter Importance of
532 the Weather Research and Forecasting Model Based On Global Sensitivity Analysis Methods.
533 Journal of Geophysical Research: Atmospheres 123, 4443–4460.
534 https://doi.org/10.1002/2017JD027348

535 Ji, F., Nishant, N., Evans, J.P., Di Virgilio, G., Cheung, K.K.W., Tam, E., Beyer, K., Riley, M.L., 2022.
536 Introducing NARCliM1.5: Evaluation and projection of climate extremes for southeast Australia.
537 Weather Clim Extrem 38, 100526. https://doi.org/10.1016/j.wace.2022.100526

538 Kala, J., Evans, J.P., Pitman, A.J., 2015. Influence of antecedent soil moisture conditions on the
539 synoptic meteorology of the Black Saturday bushfire event in southeast Australia. Quarterly
540 Journal of the Royal Meteorological Society 141, 3118–3129. https://doi.org/10.1002/qj.2596

541 Lindstrom, S.J., Nagalingam, V., Newnham, H.H., 2013. Impact of the 2009 M elbourne heatwave on a
542 major public hospital. Intern Med J 43, 1246–1250.

543 Loridan, T., Coates, L., Argueso, D., Perkins-Kirkpatrick, S.E., McAneney, J., 2016. The Excess Heat
544 Factor as a metric for heat-related fatalitis: Defining heatwave risk categories. Australian Journal
545 of Emergency Management, The 31, 31–37.

546 Mackey, D.F., Hogg, D.W., Lang, D., Goodman, J., 2013. emcee: the MCMC hammer. Publ. Astron. Soc.
547 Pac 125, 306.

548 Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y.,
549 Goldfarb, L., Gomis, M.I., 2021. Climate change 2021: the physical science basis. Contribution of
550 working group I to the sixth assessment report of the intergovernmental panel on climate
551 change 2.

552 Matear, R.J., 1995. Parameter optimization and analysis of ecosystem models using simulated
553 annealing: A case study at Station P. J Mar Res 53, 571–607.

554 Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state
555 calculations by fast computing machines. J Chem Phys 21, 1087–1092.

556 Montornès Torrecillas, A., Codina, B., Zack, J.W., 2015. A discussion about the role of the shortwave
557 schemes on real WRF-ARW simulations. Two case studies: cloudless and cloudy sky. Tethys:
558 Journal of Mediterranean Meteorology & Climatology, 2015, num. 12, p. 13-31.

559 Oke, T.R., 2002. Boundary layer climates, 2nd Editio. ed. Routledge, London.

560 Perkins-Kirkpatrick, S.E., Lewis, S.C., 2020. Increasing trends in regional heatwaves. Nat Commun 11,
561 3357. https://doi.org/10.1038/s41467-020-16970-7

562 Quan, J., Di, Z., Duan, Q., Gong, W., Wang, C., Gan, Y., Ye, A., Miao, C., 2016. An evaluation of
563 parametric sensitivities of different meteorological variables simulated by the WRF model.
564 Quarterly Journal of the Royal Meteorological Society 142, 2925–2934.
565 https://doi.org/10.1002/qj.2885

566 Reddy, P.J., Perkins-Kirkpatrick, S.E., Sharples, J.J., 2021a. Intensifying Australian Heatwave Trends
567 and Their Sensitivity to Observational Data. Earths Future 9, e2020EF001924.
568 https://doi.org/10.1029/2020EF001924

569 Reddy, P.J., Sharples, J.J., Lewis, S.C., Perkins-Kirkpatrick, S.E., 2021b. Modulating influence of
570 drought on the synergy between heatwaves and dead fine fuel moisture content of bushfire
571 fuels in the Southeast Australian region. Weather Clim Extrem 31, 100300.
572 https://doi.org/10.1016/j.wace.2020.100300

573 Reddy, P.J., Chinta, S., Matear, R., Taylor, J., Baki, H., Thatcher, M., Kala, J., Sharples, J., 2024. Machine
574 learning based parameter sensitivity of regional climate models—a case study of the WRF
575 model for heat extremes over Southeast Australia. Environmental Research Letters 19, 14010.
576 https://doi.org/10.1088/1748-9326/ad0eb0

577 Reiker, T., Golumbeanu, M., Shattock, A., Burgert, L., Smith, T.A., Filippi, S., Cameron, E., Penny, M.A.,
578 2021. Emulator-based Bayesian optimization for efficient multi-objective calibration of an
579 individual-based model of malaria. Nat Commun 12, 7212.

580 Roberts, G.O., Rosenthal, J.S., 2004. General state space Markov chains and MCMC algorithms.

581 Skamarock, W.C.;, Klemp, J.B., Dudhia, J.B., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X.-Y.;, Wang,
582 W., Powers, J.G., 2021. A Description of the Advanced Research WRF Model Version 4.3. NCAR
583 Technical Note TN–556+STR, 1–165.

584 Smith, R.C., 2013. Uncertainty quantification: theory, implementation, and applications. Siam.

585 Su, C.-H., Rennie, S., Dharssi, I., Torrance, J., Smith, A., Le, T., Steinle, P., Stassen, C., Warren, R.A.,
586 Wang, C., Marshall, J. Le, 2022. BARRA2: Development of the next-generation Australian
587 regional atmospheric reanalysis.

588 Temimi, M., Fonseca, R., Nelli, N., Weston, M., Thota, M., Valappil, V., Branch, O., Wizemann, H.-D.,
589 Kondapalli, N.K., Wehbe, Y., Al Hosary, T., Shalaby, A., Al Shamsi, N., Al Naqbi, H., 2020.
590 Assessing the Impact of Changes in Land Surface Conditions on WRF Predictions in Arid Regions.
591 J Hydrometeorol 21, 2829–2853. https://doi.org/10.1175/JHM-D-20-0083.1

592 Van Straten, G.T., Keesman, K.J., 1991. Uncertainty propagation and speculation in projective
593 forecasts of environmental change: A lake-eutrophication example. J Forecast 10, 163–190.

594 Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., Miao, C., 2014. An evaluation of adaptive surrogate
595 modeling based optimization with two benchmark problems. Environmental Modelling &
596 Software 60, 167–179.

597 Wang, C., Qian, Y., Duan, Q., Huang, M., Berg, L.K., Shin, H.H., Feng, Z., Yang, B., Quan, J., Hong, S.,
598 Yan, J., 2020. Assessing the sensitivity of land-atmosphere coupling strength to boundary and

surface layer parameters in the WRF model over Amazon. Atmos Res 234, 104738. https://doi.org/10.1016/j.atmosres.2019.104738

Wang, H., Mo, H., Di, Z., Liu, R., Lang, Y., Duan, Q., 2023. Knee Point-Based Multiobjective Optimization for the Numerical Weather Prediction Model in the Greater Beijing Area. Geophys Res Lett 50, e2023GL104330.

Williams, C., Rasmussen, C., 1995. Gaussian processes for regression. Adv Neural Inf Process Syst 8.

Williams, C.K.I., Rasmussen, C.E., 2006. Gaussian processes for machine learning. MIT press Cambridge, MA.

Xu, D., Bisht, G., Sargsyan, K., Liao, C., Leung, L.R., 2022. Using a surrogate-assisted Bayesian framework to calibrate the runoff-generation scheme in the Energy Exascale Earth System Model (E3SM) v1. Geosci Model Dev 15, 5021–5043.

Yang, B., Qian, Y., Lin, G., Leung, R., Zhang, Y., 2012. Some issues in uncertainty quantification and parameter tuning: a case study of convective parameterization scheme in the WRF regional climate model. Atmos Chem Phys 12, 2409–2427. https://doi.org/10.5194/acp-12-2409-2012