Geophysical Research Letters

RESEARCH LETTER
10.1029/2024GL111074

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], and John Taylor[4,5]

[1]Commonwealth Scientific and Industrial Research Organisation Environment, Hobart, TAS, Australia, [2]Center for Sustainability Science and Strategy, 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

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 optimized 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.

Key Points:

- Our study optimizes 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

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

P. J. Reddy,
jyoteesh.papari@csiro.au

Citation:

Reddy, P. J., Chinta, S., Baki, H., Matear, R., & Taylor, J. (2024). Gaussian process regression-based Bayesian Optimisation (G-BO) of model parameters—a WRF model case study of southeast Australia heat extremes. *Geophysical Research Letters*, *51*, e2024GL111074. https://doi.org/10.1029/2024GL111074

Received 29 JUN 2024
Accepted 22 AUG 2024

Author Contributions:

Conceptualization: P. Jyoteeshkumar Reddy, Sandeep Chinta, Harish Baki, Richard Matear
Formal analysis: P. Jyoteeshkumar Reddy, Sandeep Chinta, Harish Baki
Methodology: P. Jyoteeshkumar Reddy, Sandeep Chinta, Harish Baki, Richard Matear, John Taylor
Visualization: P. Jyoteeshkumar Reddy, Sandeep Chinta, Harish Baki
Writing – original draft: P. Jyoteeshkumar Reddy, Sandeep Chinta, Harish Baki
Writing – review & editing: Richard Matear, John Taylor

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 (Masson-Delmotte et al., 2021; Perkins-Kirkpatrick & Lewis, 2020; Reddy, Perkins-Kirkpatrick, & Sharples, 2021). 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). Parameterization 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., 2006, 2017). 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., 2015, 2017; 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 optimize, thereby reducing the computational costs.

Several studies (Baki et al., 2022b; Chinta & 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 & 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 & 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 (Chinta et al., 2023; Gong & Duan, 2017; Issan et al., 2023; Reiker et al., 2021; Xu et al., 2022). 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., 2024; Reddy, Chinta, Matear, 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 optimize the WRF model parameters that influence different output variables corresponding to heat extremes using Bayesian optimisation. We do this by focusing 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 optimized parameters. Section 4 summarizes 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 in

Supporting Information 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 s. For initial and lateral boundary conditions, the European Center for Medium-Range Weather Forecast Reanalysis fifth 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 2 m height (T) and relative humidity at 2 m height (Rh), are obtained at hourly intervals. This work extends our previous study (Reddy, Chinta, Matear, 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, Chinta, Matear, et al., 2024). The description of three sensitive parameters and the respective physics schemes is presented in the supplementary Table S1 in Supporting Information S1.

The present study selected two southeast Australian extreme heat events like the previous study (Reddy, Chinta, Matear, 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 optimized 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. More details about the selection of events are presented in the supplementary information. For all the selected events, a 36 hr 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 2 m 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:

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

where $WRF_{p\text{-}runs}$ 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 & Rasmussen, 1995, 2006)) as a surrogate model training with parameter samples as input and nMAE as target. We built the GPR model with a composite kernel that consists of three components: a constant kernel, a Radial Basis Function (RBF) kernel with a unit length scale, and a white noise kernel with a small noise value of 1^{-10}, which refers to prediction with noisy responses (regression). The GPR model provides the mean predictions and the variance; in this study, we only considered the mean predicted values. 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 in Supporting Information S1. 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.

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:

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

where $p(x)$ is the prior, $p(z/x)$ is the likelihood, $p(z)$ is the marginal likelihood or normalizing 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 & 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 (Hastings, 1970; Metropolis et al., 1953), 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 & Weare, 2010; Issan et al., 2023).

In this study, we implemented the Affine invariant ensemble sampling using the "emcee" Python package by choosing 50 walkers and stretch move as two (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 1,000 steps were chosen as burn-in, after which the chains converge (Figure S3 in Supporting Information S1). Following the Mackey et al. (2013), we run the chains to 3,000 steps (i.e., around 50 times the integrated autocorrelation time (which is around 50)) to ensure the convergence of chains to the target distribution (Figure S3 in Supporting Information S1). More information about the autocorrelation times can be found in Goodman and 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 toward the lower end of the parameter range for all the three cases of optimisation (only T (Figure 1a), only Rh (Figure 1d), and T and Rh combined (i.e., T and Rh are given equal weights in constructing a combined single objective function for optimisation) (Figure 1g)). 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 (Figure 1b), Rh (Figure 1e), and both T and Rh (Figure 1h)), with a mean value around 0.42, 0.37, and 0.40 (normalized values) when optimized 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 optimized for only Rh and on the lower end of the posterior with less probability when optimized T and Rh combined. This suggests that the default

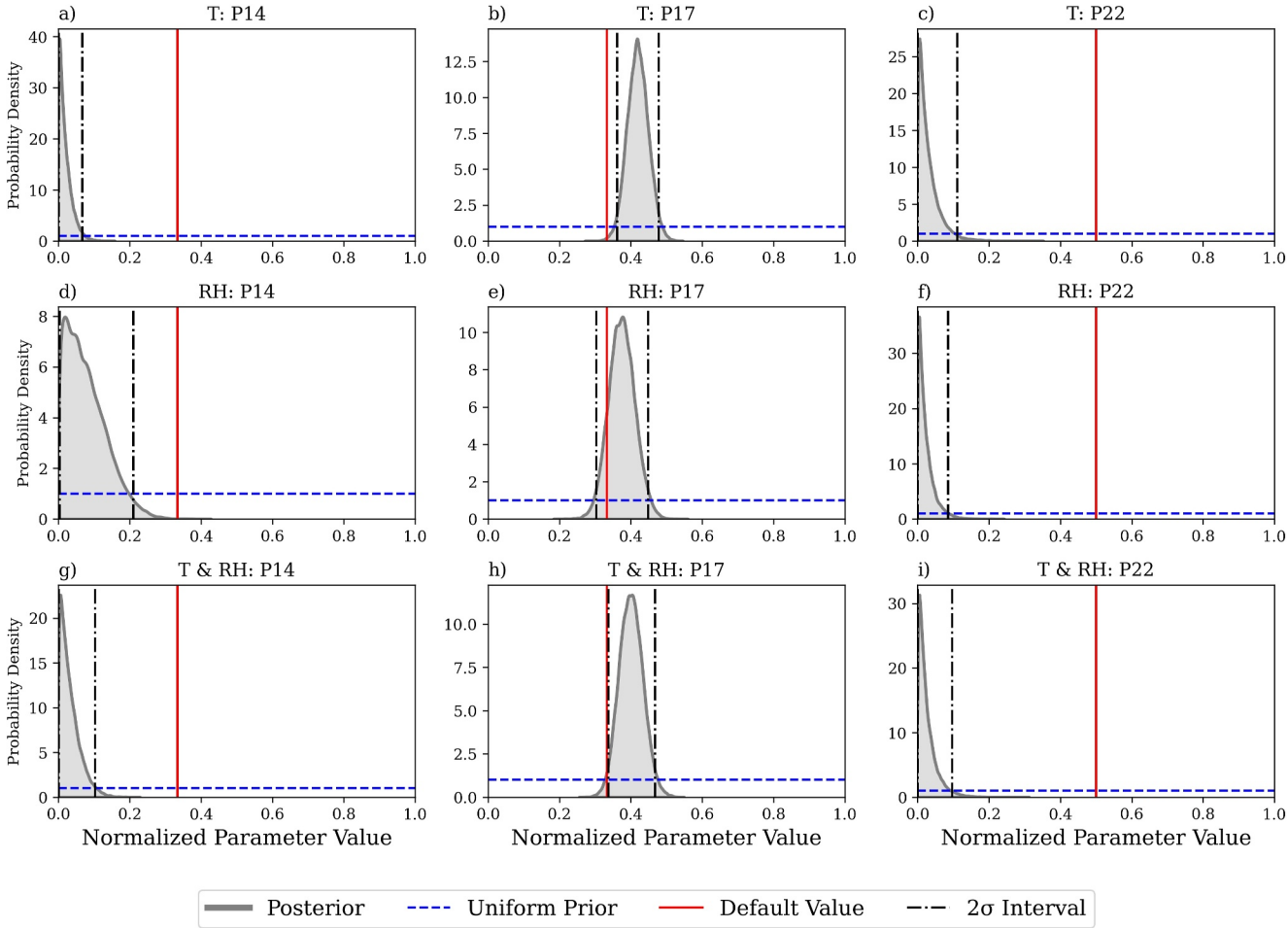

Figure 1. Bayesian optimized posterior distribution (gray shading) of sensitive parameters (presented as normalized values) for hourly temperature (*T*) (a–c) and relative humidity (Rh) (d–f) individually and for both *T* and Rh (g–i) combined. The red and blue lines show the default and uniform prior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

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 toward the lower bound of the parameter range when optimized for *T* (Figure 1c), Rh (Figure 1f) individually, and *T* and Rh combined (Figure 1i). 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 optimized 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 optimized parameter distribution combinations, as shown in Figure 2. The MAE of default or optimized 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 of optimal parameter values (see Table S2 in Supporting Information S1). These 10 parameter samples from the optimal distributions improved the MAE of the simulated *T* and Rh during all days of both events (2009 and 2019) by 11.2%–12% and 5.4%–6.8%, respectively. Further, MAE of the ensemble mean of 10 runs computed with respect to BARRA2 data is shown for the spatial comparison. Figures 2a and 2d show the average *T* and Rh, respectively, during all days of both events (2009 and 2019) using the BARRA2 data. The optimized ensemble mean reduced the MAE of *T* mostly across the domain compared to the default simulation (compare Figure 2b vs. Figure 2c). Particularly, the substantial reductions were seen in the regions of high average temperatures (greater than 33°C, see in Figure 2a) and in the northeast parts of the domain (Figure 2c). Similar to *T*, optimized ensemble mean improved the simulation of Rh compared to the default over

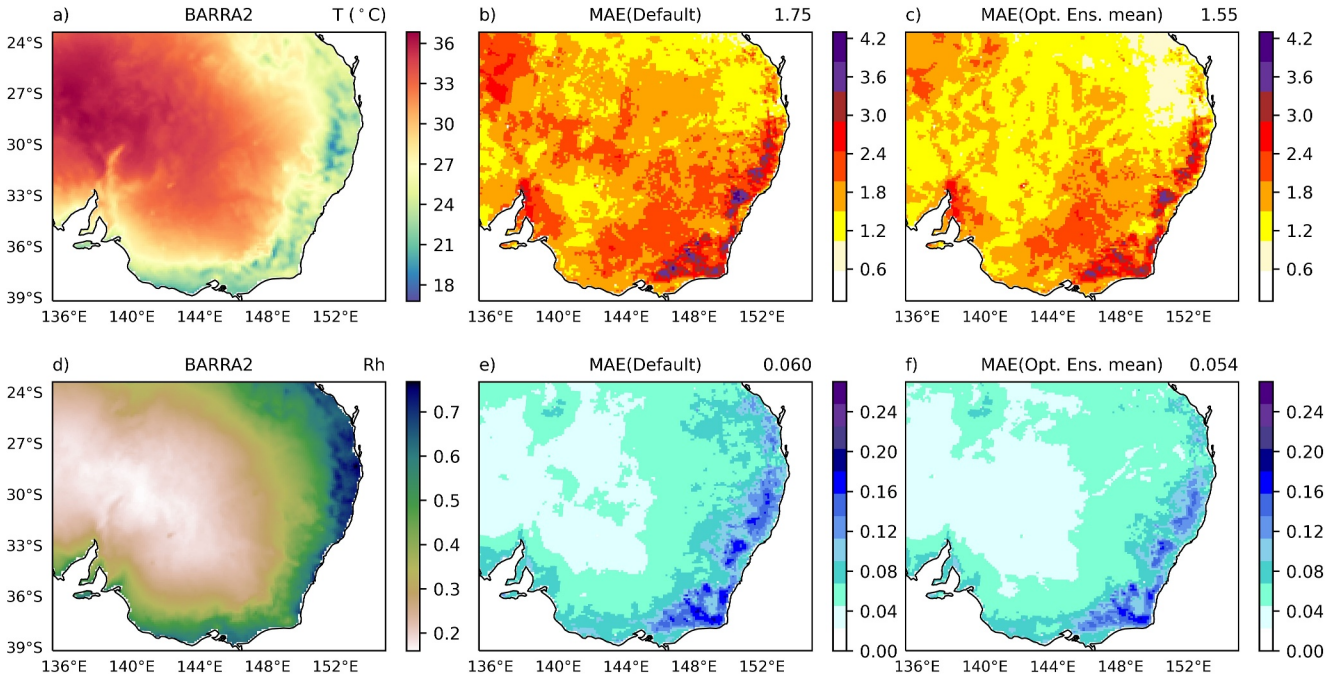

Figure 2. Spatial plot of average hourly temperature (T; °C) (a) and hourly relative humidity (Rh) (d) during all days of both selected events (2009 and 2019) using the BARRA2 data. MAE of the WRF default parameters run (default) and optimized ensemble mean (of randomly drawn 10 parameter combinations 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 spatial mean value of each subplot is shown in the top right corner of the subplot.

the regions of the low average Rh that is, central parts of the domain and across the regions of average high Rh that is, northeast coast of the domain (compare Figure 2e vs. Figure 2f). 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 optimized 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. Further, for a more thorough evaluation of optimisation results, following Wang et al. (2023), we analyzed the frequency distribution of bias improvements between the default and optimal simulations of T and Rh. Results show that 59.1%–59.8% of the total land grids show a reduction in T error, and 57.2%–58.8% of the land grids also exhibit a decrease in Rh error.

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, Sharples, et al., 2021). Hence, we compare the spatial patterns of the T_{max} and Rh_{min} mean during all days of the two selected events (Figures 3a–3c and 3f-3h) and only on the extremely hot days of each event (2009 (Figure S4 in Supporting Information S1) and 2019 (Figure S5 in Supporting Information S1)). The optimized ensemble mean simulation improved the realism of the mean T_{max} by around 2.5°C (compare Figure 3b vs. Figure 3c) and Rh_{min} (compare Figure 3g vs. Figure 3h) 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 optimized ensembles (Figure 3d). Results show that the optimized ensemble clearly improved the accuracy of T_{max} across all days of the two selected events compared to the default (Figure 3d). 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 (Figures 3e and 3j). This shows that all the 10 optimized ensemble members clearly improved the cold T_{max} and wet Rh_{min} bias compared to the default. Around 71.4%–74.5% and 65.4%–66.9% of total land grids observe an improvement in the T_{max} and Rh_{min}, respectively, with the optimal parameter simulations compared to the default run. Furthermore, the optimized ensemble accurately simulated the hot region

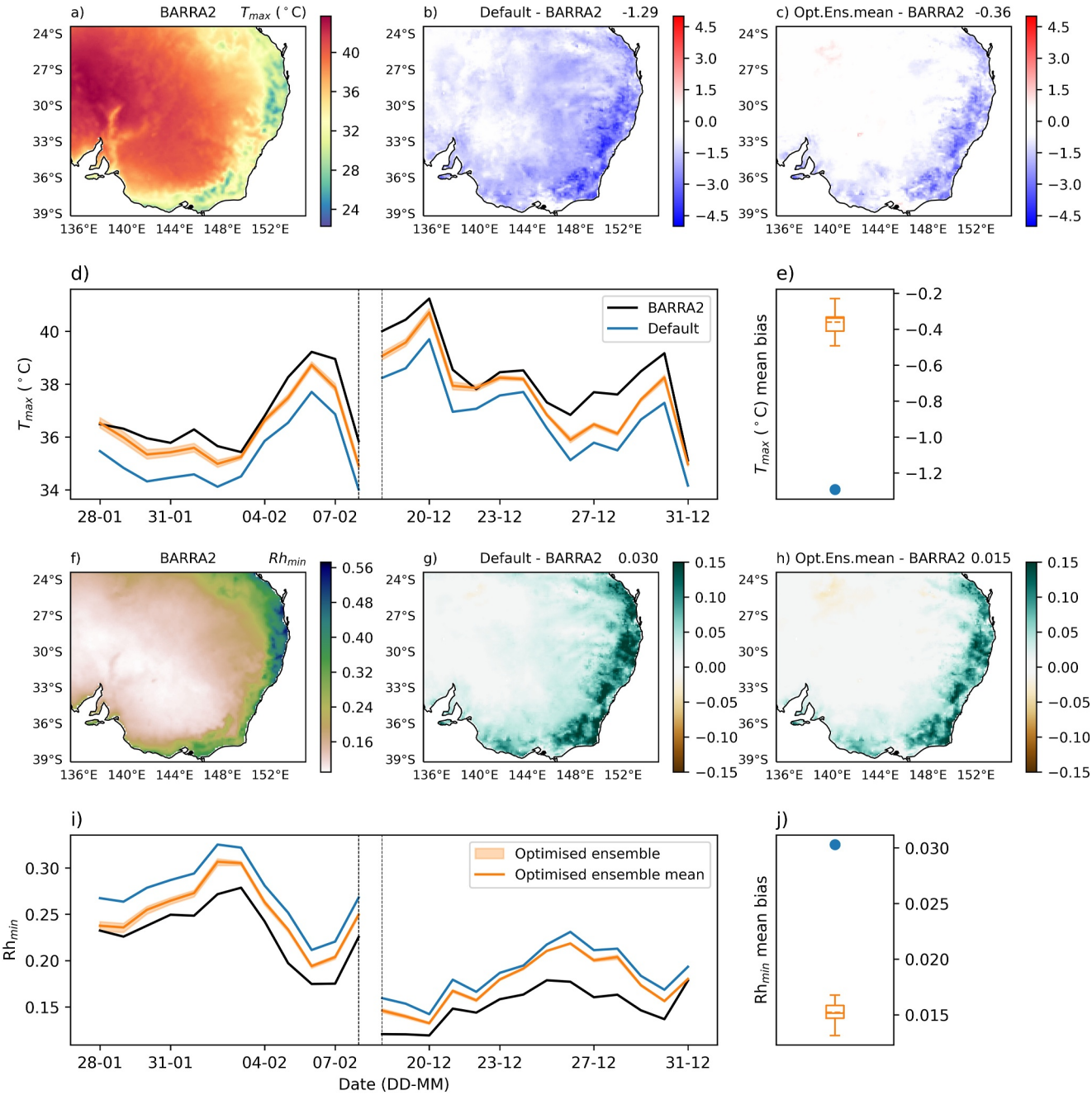

Figure 3. Spatial plot of average daily maximum temperature (T_{max}; °C) (a) and daily minimum relative humidity (Rh_{min}) (f) during all days of both selected events (2009 and 2019) using the BARRA2 data. Comparison of the WRF default parameters run (default) and optimized ensemble mean (Opt. Ens. mean) (of randomly drawn 10 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), optimized ensemble (orange shading), and optimized 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 optimized 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)). The spatial mean value of each subplot is shown in the top right corner of the subplot.

(region of T_{max} greater than 44°C) on the extremely hot day of both the selected events (2009 (compare Figure S4b vs. Figure S4c in Supporting Information S1) and 2019 (compare Figure S5b vs. Figure S5c in Supporting Information S1) compared to the default. This suggests that the optimized 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).

The G-BO optimized 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 optimized 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 (Figures S6 and S7 in Supporting Information S1). Further, the T_{max} and Rh_{min} results of the 2013 event show that the optimized ensemble improves the simulation compared to the default, which is consistent with the two optimized events (Figure S7 in Supporting Information S1). This further testing help demonstrates the robustness of the G-BO results. Additional experiments with respect to spatial resolution were performed at a 6 km horizontal grid. The optimal distribution range of the three key parameters improved the T and Rh of extreme heat events over southeast Australia on a 6 km grid, like the 12 km spatial resolution simulations (for more details, refer to supporting Text S2 in Supporting Information S1 and see Figures S8 and S9 in Supporting Information S1). This suggests the spatial resolution independence of the identified optimal parameter distributions. 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.

3.2. Physical Understanding of Optimal Distribution of Parameters

The scattering tuning parameter (P14) optimal posterior is toward the lower bound of its range with a maximum likelihood value around 0.5×10^{-5} to 0.6×10^{-5} $m^2\,kg^{-1}$, which is lower than the default value (1×10^{-5} $m^2\,kg^{-1}$). 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 (Figure 3b vs. Figure 3c). This is more specifically seen in the T_{max} because it is much affected by the P14 (Figure 3b vs. Figure 3c) (Reddy, Chinta, Matear, et al., 2024). Low P14 values favor low Rh_{min}; our results agree with this and show that the positive Rh_{min} bias in default simulation is improved in the optimized run with the P14 value lower than the default one (Figure 3g vs. Figure 3h).

The multiplier for saturated soil water content or soil porosity P17 in the land surface scheme is optimized to the posterior, with the maximum likelihood value ranging between 1.03 and 1.18, which is slightly higher than the default value (one). 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 favors a decrease in surface temperature, particularly during the daytime (Fonseca et al., 2019; Reddy, Chinta, Matear, 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 optimized posterior maximum probability value is around 1.0 to 1.14, which is lower than the default value (two). 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 optimizing 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. Our study has focused on improving the accuracy of T and Rh during extreme heat events over southeast Australia by optimizing the WRF model parameters. In particular, our results show significant improvements for extreme heat events, however, further research and optimisation would be necessary to determine the applicability and effectiveness of these parameters under different conditions.

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 Southeast Australia (Reddy, Chinta, Matear, 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 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: (a) for two of the three parameters optimized, the default WRF parameter values lie outside the optimal range suggesting the need to reconsidering the parameter values for simulating heat extremes in this region. (b) Randomly drawing 10 parameter samples from the optimal distributions improved the MAE of the simulated T and Rh during all days of both events (2009 and 2019) by 11.2%–12% and 5.4%–6.8%, respectively. (c) The mean spatial pattern of 10 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 optimized 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 (Figure S10 in Supporting Information S1). Further, Reddy, Chinta, Matear, et al. (2024) showed the sensitivities of the three optimized parameters are mostly first order, which indicates the negligible interaction effects of parameters in influencing the T and Rh. A small spread in the ensemble from the optimized model indicates constrained parameter uncertainty (Figures 3d, 3e and 3i-3j). 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 optimal parameter simulations improved the 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. Apart from G-BO, there are other approaches that utilize GPR and BO, such as the Adaptive Surrogate Modeling-Based Sampling Strategy for Parameter Optimisation and Distribution Estimation (ASMO-PODE) (Gong & Duan, 2017). Unlike G-BO, ASMO-PODE starts with a small set of initial samples and uses a resampling strategy to iteratively perform the optimisation. Future studies could focus on comparing these methods and assess the computational cost to obtain an acceptable solution. 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.

Data Availability Statement

The ERA5 data is openly available at https://doi.org/10.24381/cds.adbb2d47 (Hersbach et al., 2020). The source code of WRF v4.4 is openly available at https://github.com/wrf-model/WRF/releases/tag/v4.4 (Skamarock et al., 2021). The BARRA2 dataset is available from the NCI THREDDS data server https://dapds00.nci.org.au/thredds/catalogs/ob53/catalog.html (Su et al., 2022). The code used to perform the optimisation analysis is available at https://doi.org/10.5281/zenodo.12511781 (Reddy, Chinta, Baki, et al., 2024). All the figures are generated with Python.

Acknowledgments
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.

References

Abram, N. J., Henley, B. J., Sen Gupta, A., Lippmann, T. J. R., Clarke, H., Dowdy, A. J., et al. (2021). Connections of climate change and variability to large and extreme forest fires in southeast Australia. *Commun Earth Environment*, *2*(1), 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. *Global Change Biology*, *17*(2), 997–1012. https://doi.org/10.1111/j.1365-2486.2010.02262.x

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. *Geoscientific Model Development*, *15*(5), 2133–2155. https://doi.org/10.5194/gmd-15-2133-2022

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. *Journal of Applied Meteorology and Climatology*, *61*(7), 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. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305

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. *Climate Dynamics*, *55*(3–4), 631–650. https://doi.org/10.1007/s00382-020-05288-1

Chinta, S., Gao, X., & Zhu, Q. (2023). 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. (2024). Machine learning driven sensitivity analysis of E3SM land model parameters for wetland methane emissions. *Journal of Advances in Modeling Earth Systems*, *16*(7), e2023MS004115. https://doi.org/10.1029/2023MS004115

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*(6), e2020EA001471. https://doi.org/10.1029/2020EA001471

Di, Z., Duan, Q., Gong, W., Wang, C., Gan, Y., Quan, J., et al. (2015). Assessing WRF model parameter sensitivity: A case study with 5 day summer precipitation forecasting in the greater Beijing area. *Geophysical Research Letters*, *42*(2), 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. *Science China Earth Sciences*, *60*(5), 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. *Climate Dynamics*, *50*(5–6), 1927–1948. https://doi.org/10.1007/s00382-017-3729-3

Di Virgilio, G., Evans, J. P., Di Luca, A., Olson, R., Argüeso, D., Kala, J., et al. (2019). Evaluating reanalysis-driven CORDEX regional climate models over Australia: Model performance and errors. *Climate Dynamics*, *53*(5–6), 2985–3005. https://doi.org/10.1007/s00382-019-04672-w

Duan, Q., Di, Z., Quan, J., Wang, C., Gong, W., Gan, Y., et al. (2017). Automatic model calibration: A new way to improve numerical weather forecasting. *Bulletin America Meteorology Social*, *98*(5), 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., et al. (2006). Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *Journal of Hydrology (Amsterdam)*, *320*(1–2), 3–17. https://doi.org/10.1016/j.jhydrol.2005.07.031

Dudhia, J. (1989). Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, *46*(20), 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. *Climate Dynamics*, *39*(6), 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. *Geoscientific Model Development*, *7*(2), 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*(722), 2202–2220. https://doi.org/10.1002/qj.3553

Gong, W., & Duan, Q. (2017). An adaptive surrogate modeling-based sampling strategy for parameter optimization and distribution estimation (ASMO-PODE). *Environmental Modelling and Software*, *95*, 61–75. https://doi.org/10.1016/j.envsoft.2017.05.005

Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Ye, A., et al. (2016). Multiobjective adaptive surrogate modeling-based optimization for parameter estimation of large, complex geophysical models. *Water Resources Research*, *52*(3), 1984–2008. https://doi.org/10.1002/2015wr018230

Goodman, J., & Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, *5*(1), 65–80. https://doi.org/10.2140/camcos.2010.5.65

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751–763. https://doi.org/10.1029/97wr03495

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109. https://doi.org/10.1093/biomet/57.1.97

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/QJ.3803

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*(9), 2318–2341. https://doi.org/10.1175/MWR3199.1

Hoversten, G. M., Cassassuce, F., Gasperikova, E., Newman, G. A., Chen, J., Rubin, Y., et al. (2006). Direct reservoir parameter estimation using joint inversion of marine seismic AVA and CSEM data. *Geophysics*, *71*(3), C1–C13. https://doi.org/10.1190/1.2194510

Issan, O., Riley, P., Camporeale, E., & Kramer, B. (2023). Bayesian inference and global sensitivity analysis for ambient solar wind prediction. *Space Weather*, *21*(9), e2023SW003555. https://doi.org/10.1029/2023SW003555

Ji, D., Dong, W., Hong, T., Dai, T., Zheng, Z., Yang, S., & Zhu, X. (2018). Assessing parameter importance of the weather research and forecasting model based on global sensitivity analysis methods. *Journal of Geophysical Research: Atmospheres*, *123*(9), 4443–4460. https://doi.org/10.1002/2017JD027348

Ji, F., Nishant, N., Evans, J. P., Di Virgilio, G., Cheung, K. K. W., Tam, E., et al. (2022). Introducing NARCliM1.5: Evaluation and projection of climate extremes for southeast Australia. *Weather and Climate Extremes*, *38*, 100526. https://doi.org/10.1016/j.wace.2022.100526

Kala, J., Evans, J. P., & Pitman, A. J. (2015). Influence of antecedent soil moisture conditions on the synoptic meteorology of the Black Saturday bushfire event in southeast Australia. *Quarterly Journal of the Royal Meteorological Society*, *141*(693), 3118–3129. https://doi.org/10.1002/qj.2596

Lindstrom, S. J., Nagalingam, V., & Newnham, H. H. (2013). Impact of the 2009 M elbourne heatwave on a major public hospital. *Internal Medicine Journal*, *43*(11), 1246–1250. https://doi.org/10.1111/imj.12275

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

Mackey, D. F., Hogg, D. W., Lang, D., & Goodman, J. (2013). emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific*, *125*, 306.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., et al., (2021). Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change 2.

Matear, R. J. (1995). Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at station P. *Journal of Marine Research*, *53*(4), 571–607. https://doi.org/10.1357/0022240953213098

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, *21*(6), 1087–1092. https://doi.org/10.1063/1.1699114

Montornès Torrecillas, A., Codina, B., & Zack, J. W. (2015). A discussion about the role of the shortwave schemes on real WRF-ARW simulations. Two case studies: Cloudless and cloudy sky. *Tethys: Journal of Mediterranean Meteorology & Climatology*, *2015*(12), 13–31.

Oke, T. R. (2002). *Boundary layer climates* (2nd ed.). Routledge.

Perkins-Kirkpatrick, S. E., & Lewis, S. C. (2020). Increasing trends in regional heatwaves. *Nature Communications*, *11*(1), 3357. https://doi.org/10.1038/s41467-020-16970-7

Quan, J., Di, Z., Duan, Q., Gong, W., Wang, C., Gan, Y., et al. (2016). An evaluation of parametric sensitivities of different meteorological variables simulated by the WRF model. *Quarterly Journal of the Royal Meteorological Society*, *142*(700), 2925–2934. https://doi.org/10.1002/qj.2885

Reddy, P. J., Chinta, S., Baki, H., Matear, R., & Taylor, J. (2024). Data for publication of "Gaussian process regression-based Bayesian optimisation (G-BO) of model parameters - A WRF model case study of southeast Australia heat extremes. [Dataset]. https://doi.org/10.5281/zenodo.12511781. *Zenodo*

Reddy, P. J., Chinta, S., Matear, R., Taylor, J., Baki, H., Thatcher, M., et al. (2024). Machine learning based parameter sensitivity of regional climate models—A case study of the WRF model for heat extremes over southeast Australia. *Environmental Research Letters*, *19*(1), 14010. https://doi.org/10.1088/1748-9326/ad0eb0

Reddy, P. J., Perkins-Kirkpatrick, S. E., & Sharples, J. J. (2021). Intensifying Australian heatwave trends and their sensitivity to observational data. *Earth's Future*, *9*(4), e2020EF001924. https://doi.org/10.1029/2020EF001924

Reddy, P. J., Sharples, J. J., Lewis, S. C., & Perkins-Kirkpatrick, S. E. (2021). Modulating influence of drought on the synergy between heatwaves and dead fine fuel moisture content of bushfire fuels in the Southeast Australian region. *Weather and Climate Extremes*, *31*, 100300. https://doi.org/10.1016/j.wace.2020.100300

Reiker, T., Golumbeanu, M., Shattock, A., Burgert, L., Smith, T. A., Filippi, S., et al. (2021). Emulator-based Bayesian optimization for efficient multi-objective calibration of an individual-based model of malaria. *Nature Communications*, *12*(1), 7212. https://doi.org/10.1038/s41467-021-27486-z

Roberts, G. O., & Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, *1*(none). https://doi.org/10.1214/154957804100000024

Skamarock, W. C., Klemp, J. B., Dudhia, J. B., Gill, D. O., Barker, D. M., Duda, M. G., et al. (2021). *A description of the advanced research WRF model version 4.3* (pp. 1–165). NCAR Technical Note TN–556+STR.

Smith, R. C. (2013). *Uncertainty quantification: Theory, implementation, and applications*. Siam.

Su, C.-H., Rennie, S., Dharssi, I., Torrance, J., Smith, A., Le, T., et al., (2022). BARRA2: Development of the next-generation Australian regional atmospheric reanalysis.

Temimi, M., Fonseca, R., Nelli, N., Weston, M., Thota, M., Valappil, V., et al. (2020). Assessing the impact of changes in land surface conditions on WRF predictions in arid regions. *Journal of Hydrometeorology*, *21*(12), 2829–2853. https://doi.org/10.1175/JHM-D-20-0083.1

Van Straten, G. T., & Keesman, K. J. (1991). Uncertainty propagation and speculation in projective forecasts of environmental change: A lake-eutrophication example. *Journal of Forecasting*, *10*(1–2), 163–190. https://doi.org/10.1002/for.3980100110

Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., & Miao, C. (2014). An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. *Environmental Modelling and Software*, *60*, 167–179. https://doi.org/10.1016/j.envsoft.2014.05.026

Wang, C., Qian, Y., Duan, Q., Huang, M., Berg, L. K., Shin, H. H., et al. (2020). Assessing the sensitivity of land-atmosphere coupling strength to boundary and surface layer parameters in the WRF model over Amazon. *Atmospheric Research*, *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. *Geophysical Research Letters*, *50*(23), e2023GL104330. https://doi.org/10.1029/2023gl104330

Williams, C., & Rasmussen, C. (1995). Gaussian processes for regression. *Advances in Neural Information Processing Systems*, *8*.

Williams, C. K. I., & Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. MIT press.

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. *Geoscientific Model Development*, *15*(12), 5021–5043. https://doi.org/10.5194/gmd-15-5021-2022

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. *Atmospheric Chemistry and Physics*, *12*(5), 2409–2427. https://doi.org/10.5194/acp-12-2409-2012

References From the Supporting Information

Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*, *129*(4), 569–585. https://doi.org/10.1175/1520-0493(2001)129<0569:caalsh>2.0.co;2

Dowdy, A. J., Fromm, M. D., & McCarthy, N. (2017). Pyrocumulonimbus lightning and fire ignition on Black Saturday in southeast Australia. *Journal of Geophysical Research: Atmospheres*, *122*(14), 7342–7354. https://doi.org/10.1002/2017jd026577

Engel, C. B., Lane, T. P., Reeder, M. J., & Rezny, M. (2013). The meteorology of black Saturday. *Quarterly Journal of the Royal Meteorological Society*, *139*(672), 585–599. https://doi.org/10.1002/qj.1986

Fiddes, S. L., Pezza, A. B., & Renwick, J. (2016). Significant extra-tropical anomalies in the lead up to the Black Saturday fires. *International Journal of Climatology*, *36*(2), 1011–1018. https://doi.org/10.1002/joc.4387

Hong, S. Y., & Lim, J. O. J. (2006). The WRF single-moment 6-class microphysics scheme (WSM6). *Asia-Pacific Journal of Atmospheric Sciences*, *42*(2), 129–151.

Kala, J., Valmassoi, A., & Hirsch, A. L. (2023). Assessing the maximum potential cooling benefits of irrigation in Australia during the "Angry Summer" of 2012/2013. *Weather and Climate Extremes*, *39*, 100538. https://doi.org/10.1016/j.wace.2022.100538

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M. A., Mitchell, K., et al. (2004). Implementation and verification of the unified NOAH land surface model in the WRF model. *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction*, 2165–2170.