Authors' response to the Referees

For clarifying our answers to the reviewers' comments, the following scheme is used: Comments of the reviewers are denoted in blue while the authors response is given in black. The changes made to the manuscript in response to the reviewers comments have been appended.

It should also be noted that the title of the paper has been changed to *"A method to encapsulate model structural uncertainty in ensemble projections of future climate: EPIC v1.0"* in reference to Short Comment #1.

Anonymous Referee #1

We would like to take this opportunity to thank the reviewers for their thorough review of this paper. We recognize the value of their time and appreciate the improvements in our paper that this review has led to.

Received and published: 8 May 2017

In this paper the authors have developed a method to encapsulate structural uncertainty in ensemble projections of future climate by combining regional climate model output with that from a simple climate model. The aim is produce a large ensemble of climate variables representing that which would be produced from a global climate model if it weren't made impossible by computational demands. A small ensemble of climate variables (T_min and T_max) and annual mean global temperature is produced from the regional climate model. The relationships between the climate variables and temperature are used to produce a large ensemble of the climate variables from the more readily available annual mean global mean temperature (simulated in the simple climate model). Observations are used to provide climatology, weather variability and maintain spatial coherence to the predictions. I think the underlying method that has been developed is interesting and would benefit the scientific community if published, my concern about publishing in GMD is that this paper is about a method rather than a model.

We agree that the paper is more about a method than a model but it is a method that extends the utility of regional climate models and so we felt that it would still be of considerable interest to the modelling community. We therefore felt that GMD would still be a suitable vehicle for communicating this method to the target audience.

If the method has been developed such that it is a model in itself then the GMDD paper needs to be focussed on this and it requires more discussion about how users will run the model beyond what is presented in the paper, particularly with the new climate variables that will be included, how a new region is considered, how a new RCP is used and what will happen to the model when developments to the simple climate model and global/regional climate model are made. It may perhaps be better placed in a journal such as 'Advances in Statistical Climatology, Meteorology and Oceanography (ASCMO)'.

We believe that it would be best to defer to the editor regarding whether this paper is best suited for publication in GMD or elsewhere.

I will continue the review based on the method development.

General comment:

As a method I think this is a neat way to produce a large ensemble of climate variables at a regional level that would otherwise be unavailable and attempting to encapsulate the structural uncertainty and weather variability. In some places I find the method hard to follow and the paper needs clarity. In some places I think it would benefit from further equations to back up the text. I would also like to see more discussion on how parametric uncertainty could be included or why it is not deemed necessary. I would also like to see how additional RCPs are included in the method – are they included in the final PDFs or do you envisage separate PDFs for each scenario?

Particular points:

1. Introduction:
There is not enough of a literature review here. In particular, the UK Met Office have developed methods to produce probabilistic climate projections with a similar aim to this Paper.

We have followed the reviewers suggestions and look more widely for existing papers that support, or align with, our paper, and have cited an additional 7 papers that relate to our work.

Please discuss how this work achieves the goal differently. It would be good to see the work of Tebaldi and Knutti referenced here too discussing the difficulties in producing probabilistic information from multiple models.
In particular, the following papers should be referenced and discussed.

- Murphy, J. M. et al. A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. Phil. Trans. R. Soc. A 365, 1993–2028 (2007).
- Sexton, D., Murphy, J., Collins, M. & Webb, M.Multivariate predictions using imperfect climate models: Part 1 outline of methodology. Clim. Dynam. http://dx.doi.org/10.1007/s00382-011-1208-9 (2011).
- Harris, G. R., et al. "Probabilistic projections for 21st century European climate." Natural Hazards and Earth System Sciences 10.9 (2010): 20092020.

We have cited and discussed the pertinent features of the papers mentioned by the reviewer. However, we decided not to cite the review paper by Tebaldi and Knutti (2007) as their study focuses on multi-model ensembles, which is a different approach to generating probabilistic projections of climate from the methodology presented in our paper. With citing the other studies, we have included an overview of other methods that have been used in the past to generate climate projection.

Line 2: ‘will correctly simulate that trajectory’. I think the inclusion of ‘correctly’ is necessary since all climate models are attempting to simulate this trajectory.
Corrected.

2.1 Regional Climate Model
How dependent on the RCM are the relationships that are established? Would the relationships be expected to change with a new RCM?
The relationship between the predictors and the predictants is linear, and the regression model fit coefficients should, ideally, be robust properties of the climate system and should not depend on the RCM being used. However, to accommodate the spread of sensitivities of climate variables to changes in global mean temperature due to different model parameterisations, we combine the obtained relationships from different RCM simulations as described in the paper.

Page 3, Line 10: What does ‘adequate’ mean?
The resolution of the RCM is sufficient to describe the large scale processes New Zealand faces.

Page 3, Line 21: ‘The model simulates all atmospheric and land surface processes’. There are missing processes and those that are included are subject to uncertainties – is the method robust to this?
In this study we used simulations from a single RCM which was forced by a number of different AOGCM realizations (different AOGCM boundary conditions for each simulation). Therefore, if the processes are not included in the RCM then these will not impact the structural uncertainties of the ensemble. The missing processes and uncertainties in AOGCM models are what play a role in determining the structural uncertainty in the ensemble.

The effect of the different processes and model parameterisations of RCMs could be assessed by applying this method using a training set which included a number of different RCMs. The methodology would still hold.

Page 3, Line 26: What is the implication of the remaining bias?
The RCM data are used to model how the variable of interest and its variability change over time, while the baseline climatology is obtained from observations. The static and time varying components of the ensemble members are generated separately. Any remaining bias in the RCM data is removed when anomalies are calculated therefore it does not influence the ensemble members.

2.2 Simple climate model
Is MAGICC the only such model? How is it known to simulate annual mean global mean temperature adequately? Are the 19 AOGCMs and 10 carbon cycle models defined by MAGICC or could a user change what is included?

MAGICC is an open access model, publicly available from (incl. website) which has been widely used in the scientific community, but is not the only simple climate model that is able to produce annual mean global mean temperature. The tuning files are provided with MAGICC, but if the user needs additional tunings (other AOGCMs or carbon cycle combinations), additional model tunings can be created.

Page 4, Line 2: It is you that considers the 190 simulations as equally probable. 'We consider the resultant 190 different 'tunings' for MAGICC to be equally probable.' Could a user make different decisions?
Yes, a user could make different decisions if they had additional insight into which simulations could be more probable than others. The method we have developed can easily accommodate this more sophisticated approach. For the purposes of demonstrating our methodology, we have simply followed the methodology of Reisinger, A.; Meinshausen, M.; Manning, M. and Bodeker, G., Uncertainties of global warming metrics: CO2 and CH4, Geophysical Research Letters, 37, L14707, doi:14710.11029/12010GL043803, 2010. We do not have additional insight from the MAGICC team as to which tuning files, both for the AOGCM emulation, and for the carbon cycle emulation, would be more probable than others and so could not encompass such additional information in our demonstration of the methodology. But there is no reason why that would not be possible if that information was available.

Page 4, Line 3: What does 'some' mean? Since the title of the paper states that the structural uncertainty is encapsulated this needs explaining.
The annual mean global mean temperatures from MAGICC do not include inter-annual variability. An additional sentence has been added to the manuscript clarifying this limitation.

Page 4, Line 7: This is the first mention of 'local' -I think this should be included earlier as it's an important point. Earlier, X is used. It would add clarity to define X better earlier on and use X here and in future.
We have made this addition.

2.3 Virtual Climate Station Network
Is this the best observational dataset beyond New Zealand?
No. The VCSN data set is for New Zealand only - we have now clarified that in the paper. People wanting to deploy this method for other countries would need to source their own version of such a data set.

3. Methodology
This section would benefit from some equations to clarify the text – especially right at the beginning to go with the bullet points.
We feel that adding additional equations at this point in the manuscript would likely confuse readers. The following sentence states that more detailed descriptions are given below and the purpose of this section is to provide a high level overview of the methodology.

Why is the period 1960-2100 used?

Once the model has been trained on RCM data, which span 1971-2100, ensemble members can be generated over any reasonable timeframe for which SCM output can be produced. This assumes that the climate system continues to respond linearly, which may not be true if applied under conditions well outside those for which the model was trained on. The period 1960-2100 was chosen arbitrarily to suit our later analysis.

Page 4, Line 27: 'one or more RCM simulations' – I don't quite understand how the RCM simulations are being used here? Later it says you use five – what is the benefit of five and how did you choose these five? How could a user choose a different number?
The methodology is agnostic to the number of RCM simulations available. The 5 simulations that we used were produced using the same model and were all that were available for this study at the time of writing. The uncertainty in the response of the RCM model can be explored even further if more RCM simulations were available.

Page 4, Line 28: You haven't defined alpha yet so this is confusing. Can you state what the sampling of these alpha values is representing instead?
We have added a clarifying sentence.

Page 5, Line 2: 'valid for any GHG emissions scenarios' – is it valid if the regression isn't robust? How can a user validate this if they change the GHG scenario?
No, the method should only be performed if a robust regression fit is obtained. This caveat has been added to the manuscript.

Page 5, Line 2: Why is the anomaly calculated to the 2000-2010 baseline? You later say it is rather short so please say why you have chosen it.
The baseline period was chosen with respect to the application for which we used the data. We were interested in assessing how maximum and minimum surface temperatures change over the 21st century. A sentence describing this choice of the baseline period was added to the manuscript.

3.1 Climatology
Page 5, Line 9: Are you following the method cited or have you expanded it? Can you say why you have chosen this particular method to account for seasonality?
We are following the method presented in Kremser et al. (2014) and expand the fit-coefficients in Fouriers to account for seasonality. Another method would be to fit the regression model completely independently for each month, i.e. first fit the model to only the January data, then fit the model to only the February data etc.. You will then end up with 12 regression coefficients, one for each month, that capture the seasonal dependence of the data on the basis functions. The disadvantage of this approach is that the number of fit coefficients increases by a factor of 12. This significantly increases the uncertainty on the fit coefficients. Using Fourier expansions to account for seasonality reduces the number of fit-coefficients to 5 and therefore reduces the uncertainty.

Figure 2:
I think magenta and red are mixed up.
Corrected.

3.2.1 Training phase

I find this section particularly hard to follow and think it would benefit from more precise equations. Equation 2 doesn't help to explain what you actually did in terms of using time series from 5 RCMs.

We acknowledge the reviewers concern that this section is hard to follow and have reordered the section and provided more background information about the process that is being undertaken. Specifically, we discuss how each RCM simulation is handled independently and have included an additional equation (Eq. 3) with the fit coefficient term expanded.

How have you built a model to explain the relationship between variables of a different scale (daily and annual)? How robust is this relationship with only 5 RCMs? Can you explain the statistical models that you build and how they are validated to produce robust relationships that can be used with the simple climate model output?

A majority of the uncertainty in the ensemble arises from the the range of T'_{global}'s produced by MAGICC. The regression from the 5 RCMs agree relatively well for a given GHG emissions scenario. More information regarding the regression model used is provided in Bodeker and Kremser (2015) and Moore et al. (2003). This shortcoming has been addressed in the revised manuscript.

Page 7, Line 6: The relationship itself is calculated over the period 1960-2100?

No, during the training phase, the relationship is calculated using the RCM data sets which span the 1971-2100 period.

Page 7, Line 7: 'an annual mean global mean ..temp.. series' – do these match the period of X'? Are there five series from the five models?

Correct. This is discussed in the paragraphs that follow.

Page 7, Line 7:

\alpha(d) is the fit coefficient – are there 365 of these values?

No there are not 365 alpha values. As explained in the manuscript, there are 5 fit coefficients values, α_0 to α_4 as given in Eq. 3. We have now provided an additional equation and text to clarify this.

Page 7, Line 13: GCM 'and' RCM.

Done.

Page 7, Line 17: I don't see how the Fourier pairs are embedded in equation 2 to find five fit coefficients from 365?

An additional equation has been added to clarify the expansion into five fit coefficients.

Page 7, Line 18: 'recall, that this is applied at..' – I don't see where it told me that is is applied for each RCM simulation – are they not all used to find the relationship? How do you choose a specific RCM simulation? You should also remind the reader at the start of this section that this is done at a grid box level.

This has been addressed in the revised manuscript by reordering of section 3.2.1.

Figure 3: Could you zoom in to show what the red line looks like? It's quite hard to see. Equation 2 must be more complicated than it looks to produce this line.

The interannual variability arises from the T_{global} time series. The α values do not change from year to year, but simply scale the T_{global} values. A sentence has been added to this effect.

Page 8, Line 5: 'The unitless \alpha^'– a is five values obtained from 365 (maybe more for the five RCMs) so I can't picture what it really is or how it represents sensitivity to temperature. I'm sorry I'm quite lost with respect to alpha. Maybe more equations would help.

For a given RCM, the regression model fit-coefficient α represents the relationship between the variable of interest X' and T'_{global}: X'(t) = α x T'global(y)

To capture the seasonal dependence of X' and T'_{global}, the fit-coefficient α is expanded in Fourier series (as described in Eq. 1). We have now include another equation into the revised manuscript to clarify this expansion and to clarify the resulting 5 fit-coefficients:

$X'(t) = (\alpha_0 + \alpha_1 \times \sin(2\pi\, d/365) + \alpha_2 \times \cos(2\pi\, d/365) + \alpha_3 \times \sin(4\pi\, d/365) + \alpha_4 \times \cos(4\pi\, d/365)) \times T'_{global}(y) + R(t)$

Page 9, Line 6: alpha depends on the RCM. How have you used the RCMs here? Have you chosen one – if so, why start with five? What happens to the data from the other 4? Have you found alpha with all five and then randomly selected from all of them to produce the MC sample? Perhaps reordering the writing here might clear things up with regards to how you are using the RCMs.

The next sentences describe how the alpha value is chosen. We generate 5 alpha values, one for each RCM simulation available. Each ensemble member randomly chooses one of the 5 alpha values.

3.3 Indirect response.

Page 10, Line 26: What does series mean here? Please be more precise.

We have removed "a series of daily maps of" as we agree that it confuses the explanation of EOF analysis. This paragraph has been reworded for clarity.

Page 11, Line 4: I don't understand this point. Can you reword it to be more precise.

We have followed the suggestion by the review and reworded the sentence to:

"That there will be patterns of variability (weather) whose amplitude and variability will respond to climate change as well as others which do not change with increases in T'_{global}."

3.3.1 Identifying the modes.

Page 11, Line 6: Are you referring to the residuals in Figure 3? Your language appears to have changed here and it sounds like you are doing something new. If these are the residuals please use the same language and link it better to the previous sections.

Figure 3 shows the anomalies of the daily surface temperature. The residuals are what remains unexplained by the fit of Eq. 2. This sentence has been reworded to expressly mention where these residuals come from.

Page 11, Line 7: ‘Where the patterns of variability obtained from EOF projections of VCSN’
This sentence has been corrected.

Figure 5: Can you interpret these EOFs? Why is the period 1972-2013 used here? Also, this is not the RCM discussed in the introduction? Why have you changed the RCM?
The first EOF means that the mode common mode of variability in New Zealand is that the entire country is warmer or colder than average on a given day. The corresponding principal component time series shows a strong correlation with T_{global} (not shown). EOF 2, 3 and 4 represent different large scale weather patterns typically seen in New Zealand with the Southern Alps (Middle of the South Island) causing differences in East-West and North-South. For example, EOF2 shows that areas of the North Island and East coast of the South Island are often anomalously warm (or cold if the PC is negative) on the same day.

Figure 5 presents an example of the EOF output for one particular RCM simulation which has been forced by the sea surface temperatures from the NorESM1-M mode. The same analysis has been performed for the 4 other RCM simulations using the boundary conditions from other AOGCMs as detailed in the manuscript (not shown). All examples use the simulation that was forced by prescribed RCP 8.5 NorESM1-M sea surface temperatures for consistency.

The period 1972 to 2013 is used for the VCSN data only (because the observations are only available over that period), while for the RCM simulations output till 2100 can be used.

Page 13, Line 12: Is it just New Zealand or is it likely to be more representative everywhere?
This would hold everywhere, assuming a suitable observational dataset was obtained for the location of interest.

Page 13, Line 14: Equations would be help clarity here. I’m still struggling to understand how you are correlating data on different scales.
This is the same process as described in the training phase (Sec 3.2.1), but in this case the X’ is replaced with PC and α is replaced with β.

$$PC\_pseudo(t) = \beta \times T'_{global}(y) + R(t)$$

Page 13, Line 23: How did you remove the autocorrelation?
By transforming the α basis function using a first order autoregressive model as described in Bodeker and Kremser (2015)

Page 13, Line 25: Why did you not remove autocorrelation at larger time lags? Can you talk about the implications? If it would matter that the interannual variability is too small why not calculate and remove it? If it’s not ‘too small’ then justify not doing it.

We do not have a suitable proxy for modelling how this interannual variability changes with time and annual mean global mean surface temperature.

Page 13, Line 28: How did you create pdfs from time series? Is the data from the whole time series in the pdf or are these time slices? What is the implication of having time series data in these pdfs? Have you got a sample of PCs to create the pdf?
As mentioned in the caption for Figure 6, these PDFs are generated from time slices (2000-2010 and 2090-2100) for each PC time series. Choosing relatively short time slices ensures that any temporal trends in the data do not skew the PDFs.

3.3.3 Modelling higher order modes.
Page 15, Line 2: Can you better explain where \sigma comes from? Perhaps use an equation.
This equation has been restructured and references the equation describing fourier pairs in an earlier section

Page 15, Line 9: Why is n=50?
The first 50 EOFs explain approximately 98% of the variability in the weather noise which was deemed sufficient to capture the patterns of variability. The explained variability in each subsequent EOF rapidly decreases.

Figure 7: Can you show a close up of one of the blue lines perhaps – it's difficult to see how the smaller scale temporal patterns are captured and discussed later.
The blue line shows the median of 1900 daily temperature values with small but growing annual cycle over the period of interest. The importance of the blue line is to show the long term trend over the 21st century, rather than showing a close up of the annual cycle. A close up of the blue line does not add significant information to the paper

Page 17, Line 5: I don't understand the line starting 'The apparent annual cycle..'. Could you elaborate? Is this what you'd expect and is it adequate?
To clarify, the annual cycle seen in Figure 7, is not an annual cycle in the temperature anomalies (because the annual cycle has been removed as stated in the figure caption), but rather represents that the variability is changing over time.

Yes, this is an expected result due to plotting daily data as percentiles for a long time span.

Page 17, Line 9: Are there any implications to the interannual variability being smaller? Is improving this a future direction?
The inter-annual variability limits the utility of analysing each ensemble member in isolation. We would like to address this limitation in future research.

Page 17, Line 23: Can you say where 19000 comes from? Previously, you mentioned 1900 simulations so I'm unsure what the extra simulations are taking into account.
The Monte Carlo analysis and modelling of the weather noise are stochastic processes. Each model run produces a different set of ensemble members. Therefore, a large number of ensemble members can be generated drawing from the same statistical relationships

established in the training phase. In this case, 10 ensemble members were generated for each T_{global}. Clarifying sentences have been added to the manuscript.

Technical points: Abstract: Page 1, Line 11: change the direction of the first quote mark. This happens in other parts of the paper too.
Corrected.

Introduction: Page 2, Line 7: Do you mean ‘uncertain’?
Removed the word.

Page 4, Line 28: Make all of global subscript – use _{global}.
Fixed.

Anonymous Referee #2

We would like to take this opportunity to also thank this reviewer for their thorough review of this paper. Their suggested changes have certainly improved the quality of the paper.

Received and published: 9 May 2017

Jared and colleagues present a method to compute distributions of local weather variables, and provide an example of how it can be applied to the case of New Zealand.
For each geographic location, the method generates combinations of local climatology, long-term forced changes, and stochastic weather. By combining long-term temperature projections consistent with a large set of AOGCMs and carbon-cycle models, the paper claims to encapsulate model structural uncertainty in ensemble projections of future climate, and climate change.

The method proposed by the paper is interesting. However, the paper neglects references to earlier literature that have proposed related approaches. Furthermore, the method as it is currently described shows some serious shortcomings, particularly in the assessment and inclusion of structural uncertainty in the Probability Density Functions (PDFs) suggested. Two issues stand out:

1) Rather than encapsulating structural model uncertainty in a sensible and robust way, the current method basically multiplies and preserves model sampling bias. Just like the proposed method explores stochastic weather variations with an EOF analysis to understand dominant modes of variability, the same should be carried out for the 19 AOGMs and carbon-cycle models.
While what the review states is, in principle, possible, this is not the purpose of this analysis. Nobody has run all 190 combination of 19 different AOGCM and 10 different carbon cycle models to generate the fields that would be required to conduct such an EOF analysis. Few modelling groups would have the computing power or personnel to achieve that. In an ideal world this would be the optimal way to achieve an ensemble of projections that encapsulates structural model uncertainty in a sensible and robust way - the method we have published provides a practical means of achieving this in a world that is not ideal. It is clear from the

description of our method that it does not solve **every** aspect of the model structural uncertainty problem, but we are not aware of any method that does.

The implicit assumption that each AOGCM realization is statistically or structurally independent is not supported.
Our method does not assume that each AOGCM realization is statistically or structurally independent - noting that we have no AOGCM realizations. It is therefore not immediately obvious to us what the reviewer's criticism is. To be clear: our method relies on the existence of 19 different AOGCM tuning files for the MAGICC simple climate model and 10 different carbon cycle model tuning files for MAGICC. We acknowledge that neither set of tuning files spans the entire potential tuning space and we acknowledge that the set may not reflect the true distribution of the most probably tunings. We make no pretence that we achieve these goals and we are not aware of how would could achieve them; one would need to know, for example, what the exact PDF for climate sensitivity looks like. We have now added the paragraph.

"The EPIC method does not attempt to faithfully represent the full, true PDF of potential tuning parameters both for the AOGCM tunings and the carbon cycle model tunings i.e. were MAGICC tuned to a different set of AOGCMs (e.g. the CMIP5 set rather than the CMIP3 set), we would obtain a different set of tuning files which could lead to a somewhat different spread in our generated ensembles. The purpose of this paper is not to generate perfect ensembles that encapsulate structural model uncertainty in a completely accurate way but rather to describe a method that provides a better representation of that uncertainty than can be achieved with only a limited set of RCM simulations. The robustness of the EPIC method depends on the set of AOGCM and carbon cycle model tunings available and as more comprehensive sets (that better reflect the likelihood of some tunings over others) become available, we expect that the large ensembles generated by EPIC to better reflect the true underlying uncertainties."

This would benefit strongly from appreciating the findings from Masson & Knutti (2011) or Knutti et al (2013) to determine the structural independence of AOGCMs, and apply the methods of model weighting as described in Knutti et al (2017) in order to address structural model uncertainty in a meaningful way.
We have cited and discussed the pertinent features of the papers cited by the reviewer. While we see the value in the weighting of projections in the generation of multi-model ensembles as described in Knutti et al (2017), we feel that this method does not apply. The annual mean global means obtained from MAGICC do not provide enough information to develop a robust model quality metric. Not all combinations of AOGCM and carbon cycle models are available in CMIP3, therefore, a model quality metric cannot be established from AOGCM output for every MAGICC T'_{global}.

2) The proposed method hinges on the assumption that fields of variable X are independent of the structural model uncertainty in AOGCMs.
This is not true. The proposed method hinges on the assumption that the model structural uncertainty (both for AOGCM and carbon cycle model) is reflected in the 190 annual mean global mean surface temperature time series that we use as predictors to generate fields of

X. Quite the opposite from what the review states is true, our reconstructions of fields of X are completely dependent on the structural uncertainties of the AOGCMs and carbon cycle models whose combinations of tunings were used to create the 190 member set of T'_{global}. Our assumption is that the distribution better reflects that PDF of the resultant ensemble projections of X than would a limited set of output from RCM simulations.

This assumption is not supported by any evidence.
We agree that if we had made this assumption, that it would not be supported by the evidence.

Not all patterns have to be equally probable to occur with a certain T_global response to a specific forcing path.
We agree and have captured this by having the principal component time series correlate with T'_{global} where such a correlation is found to be statistically robust. This allows for the probability of the patterns to change with changing T'_{global} in a way that is consistent with RCM projections.

What is required here is evidence based on a re-analysis of AOGCM data which shows that local patterns (or patterns of boundary conditions for the RCM) are either equally probably across high and low-response AOGCMs, or differ across these responses pointing towards the limitations of the proposed method.
Our analysis does not use AOGCM data. It uses RCM data because we are focussing on an RCM domain. And we do indeed analyse the RCM data for their long term change in the probability of different patterns occurring.

The claims about the applicability and usefulness of the method would be unsupported if both these points are not addressed in a significantly revised manuscript.
We have now added material and clarification to address these points, where appropriate, in the paper.

Smaller editorial comments:

P4L28: T_global is formatted incorrectly
Corrected.

P5L2: Please edit this sentence for spelling and grammar. The authors need to provide evidence to make the claim that the methodology is valid for any chosen GHG emissions scenario.
Corrected.

P6Fig2: Color descriptions do not match the figure.
Corrected.

References:
Masson, D., and R. Knutti (2011), Climate model genealogy, Geophys. Res. Lett., 38, L08703, doi:10.1029/2011GL046864.

Knutti, R., D. Masson, and A. Gettelman (2013), Climate model genealogy: Generation CMIP5 and how we got there, Geophys. Res. Lett., 40, 1194–1199, doi:10.1002/grl.50256.
Knutti, R., J. Sedlá.cek, B. M. Sanderson, R. Lorenz, E. M. Fischer, and V. Eyring (2017), A climate model projection weighting scheme accounting for performance and interdependence, Geophys. Res. Lett., 44, 1909–1918, doi:10.1002/2016GL072012.

A method to encapsulate model structural uncertainty in ensemble projections of future climate: EPIC v1.0

~~J.~~ Jared Lewis[1], ~~G.~~Greg E. Bodeker[1], ~~S.~~ Stefanie Kremser[1], and Andrew Tait[2]

[1]Bodeker Scientific, 42 Russell Street, Alexandra, 9320, New Zealand
[2]National Institute of Water and Atmospheric Research, Wellington, New Zealand

Correspondence to: Jared Lewis
(jared@bodekerscientific.com)

Abstract. A method, based on climate pattern-scaling, has been developed to expand a small number of projections of fields of a selected climate variable (X) into an ensemble that encapsulates a wide range of model structural uncertainties. The method described in this paper is referred to as the Ensemble Projections Incorporating Climate model uncertainty (EPIC) method. Each ensemble member is constructed by adding contributions from (1) a climatology derived from observations that represents the time invariant part of the signal, (2) a contribution from forced changes in X where those changes can be statistically related to changes in global mean surface temperature (T_{global}), and (3) a contribution from unforced variability that is generated by a stochastic weather generator. The patterns of unforced variability are also allowed to respond to changes in T_{global}. The statistical relationships between changes in X (and its patterns of variability) with T_{global} are obtained in a ~~’~~‘training’ phase. Then, in an ~~’~~‘implementation’ phase, 190 simulations of T_{global} are generated using a simple climate model tuned to emulate 19 different Global Climate Models (GCMs) and 10 different carbon cycle models. Using the generated T_{global} time series and the correlation between the forced changes in ~~X~~ X and T_{global}, obtained in the ~~’~~‘training’ phase, the forced change in the X field can be generated many times using Monte Carlo analysis. A stochastic weather generator ~~model~~ is used to generate realistic representations of weather which include spatial coherence. Because GCMs and Regional Climate Models (RCMs) are less likely to correctly represent unforced variability compared to observations, the stochastic weather generator takes as input measures of variability derived from observations, but also responds to forced changes in climate in a way that is consistent with the RCM projections. This approach to generating a large ensemble of projections is many orders of magnitude more computationally efficient than running multiple GCM or RCM simulations. Such a large ensemble of projections permits a description of a Probability Density Function (PDF) of future climate states rather than a small number of individual story lines within that PDF which may not be representative of the PDF as a whole; the EPIC method largely corrects for such potential sampling biases. The method is useful for providing projections of changes in climate to users wishing to investigate the impacts and implications of climate change in a probabilistic way. A web-based tool, using the EPIC method to provide probabilistic projections of changes in daily maximum and minimum temperatures for New Zealand, has been developed and is described in this paper.

1 Introduction

While future changes in climate will follow a single trajectory, it is highly unlikely that any single climate model projection will correctly simulate that trajectory. The use of a single model projection is therefore insufficient for assessing the potential future state of the climate. Rather, what is required is a large (e.g. 10,000 member) ensemble of projections that provides a probabilistic portrayal of how the climate is expected to evolve. Clustering of trajectories within that probabilistic envelope then shows where any single trajectory has a higher likelihood of occurring. Probabilistic simulations of future climate, presented as Probability Density Functions (PDFs), give decision-makers a much clearer picture of likelihoods of certain future climate states compared to a single projection, or a small set of projections (Watterson et al., 2008). That said, if decision-makers are presented with PDFs obtained from the same family of models, these may be biased by the assumptions and limitations inherent in a single family of models that do not explore the possible trajectories seen in other model families. PDFs of future climate that consider all sources of uncertainty, including uncertainty resulting from structural differences in the underlying models, provide the information needed for quantitative risk assessments, since the likelihood of any particular trajectory can be estimated.

Exploring expected changes in extreme weather events also requires probabilistic simulations of future climate. While climate change may result in a small shift in the mean and/or standard deviation of a PDF of a selected climate variable, the tails of the distribution, which represent extreme weather events, can exhibit fractionally much larger changes (see Figure 1.8 in Solomon et al. (2007)). It is especially important that extreme events, which by their nature are unusual, are captured in an ensemble of projections.

Resolving changes in the frequency of regional-scale extreme weather events requires large ensembles of projections of high spatial and temporal resolution. Generating such ensembles using models which simulate all important physical processes, such as Global Climate Models (GCMs) or Regional Climate Models (RCMs), is currently computationally prohibitive. The ideas underlying climate pattern-scaling suggest a means of overcoming this hurdle and form the basis for the newly developed Ensemble Projections Incorporating Climate model uncertainty (EPIC) method described here. First, a robust statistical relationship is derived between the local climate variable of interest (X) and some associated readily generated predictor. In climate pattern-scaling, this predictor is typically the global mean surface temperature (T_{global}). If observations are being used to establish this relationship, then observed values of X and T_{global} would be used. If GCM or RCM output is used to establish the relationship, then X and T_{global} should come from the same model simulation.

Once the relationship between X and T_{global} has been determined, then, given multiple versions of T_{global}, multiple time series of X can be generated based on that relationship. This methodology assumes that many versions of T_{global} can be simulated in such a way as to capture the inherent variability resulting from structural uncertainties in GCMs and carbon cycle models in a computationally efficient way, e.g., through the use of a simple climate model (SCM). If the large ensemble of T_{global} time series spans the range of model structural uncertainties, then the resultant ensemble of generated X time series will reflect that spread in uncertainties, e.g., as done in Reisinger et al. (2010).

A number of previous studies (e.g. Murphy et al. (2007), Sexton (2012) and Harris et al. (2010)) used a method that was designed by the UK Met Office (Murphy et al., 2009) to provide probabilistic projections of future climate for Europe. This method combines information from a perturbed physics ensemble (PPE), multi-model ensembles to capture model structural uncertainties, and observations. Since GCMs have been shown to not be structurally independant (Masson and Knutti (2011), Knutti et al. (2013)), multi-model ensembles benefit from model weighting to improve the ensemble performance (Knutti et al., 2017). The limitations of these methods are that large computer resources are required to run the large ensembles of simulations required which limit the ability to apply this method across a large number of different greenhouse gas emissions scenarios.

2 Models and Data Sources

2.1 Regional Climate Model

An RCM simulation, or a number of RCM simulations, are used to provide the time series used to train EPIC i.e. to quantitatively establish the relationship between the change in annual mean global mean surface temperature and the change in the climate variable of interest and its variability. The RCM simulations used in this study were performed using the Hadley Centre regional climate model HadRM3-PRECIS (Jones et al. , 2004) that has been modified to be used for New Zealand (Bhaskaran et al., 1999, 2002; Drost et al., 2007). The RCM domain spans 32°S to 52°S and 160°E to 193°E (167°W) on a regular rotated grid with a horizontal resolution of 0.27° and with the North Pole at 48°N and 176°E. Such a rotated grid, with the equator running through the New Zealand domain, ensures a quasi-uniform grid box spacing. The 0.27° resolution results in a domain of 75×75 grid points, reduces computation time for long simulations, and has been shown to be adequte in previous studies (Drost et al., 2007). The spatial resolution necessitates a computational time step of 3 minutes. The model orography and vegetation data sets were updated from those used by Drost et al. (2007) to the high resolution surface orography data set used in NIWA's operational forecast model (Ackerley et al., 2012); differences in the vegetation fields are small. The first year of model simulation (the spin-up) is excluded from the analysis as this is used to achieve quasi-equilibrium conditions of the land surface and the overlying atmosphere.

The RCM lateral boundary conditions can be sourced either from meteorological reanalyses (these are typically used for hindcast simulations) or from Atmosphere-Ocean Global Climate Model (AOGCM) output. The AOGCM used in this study was HadAM3p developed by the Hadley Centre in the UK and forced by Coupled Model Intercomparison Project Phase 3 (CMIP3) sea surface temperatures at the air-sea interface for past and future climate simulations. HadAM3p is a slightly improved version of the atmospheric component of HadCM3 with 19 vertical levels and a horizontal resolution of 1.875° longitude by 1.25° latitude. HadAM3/HadAM3p simulate all atmospheric and land surface processes relevant to climate (Pope et al., 2000). Processes related to clouds, radiation, the boundary layer, diffusion, gravity wave drag, advection, precipitation and the sulfur cycle are all parameterized in HadAM3p. Additional details regarding HadAM3p are available in Gordon et al. (2000), Pope and Stratton (2002), Pope et al. (2000), and Gregory et al. (1994).

Most GCM and RCM simulations display biases when compared to observations. The RCM simulations used in this study were partially bias-corrected by bias-correcting the sea surface temperatures that are used as lower boundary conditions for the

HadAM3p simulations, which then provided the lateral boundary conditions for the RCM simulations (personal communication A. Sood, 2016).

2.2 Simple Climate Model

In this study, MAGICC (Model for Assessment of Greenhouse-gas Induced Climate Change; Meinshausen et al. (2011); Meinshausen et al. (2011)) is the simple climate model (SCM) used to generate an ensemble of T_{global} time series. MAGICC is a reduced complexity climate model with an upwelling diffusive ocean and is coupled to a simple carbon cycle model that includes carbon dioxide (CO_2) fertilization and temperature feedback parameterisations of the terrestrial biosphere and oceanic uptake. MAGICC can be tuned to emulate the behaviour of 19 different AOGCMs (Meehl et al., 2007) and 10 carbon cycle models (Friedlingstein et al., 2006). The resultant 190 different ''tunings' for MAGICC can be used to generate 190 equally probable T_{global} time series that provide some representation an indication of the spread in T_{global} resulting from structural uncertainties in AOGCMs and the carbon cycle models used in C4MIP (Coupled Carbon Cycle Climate Model Intercomparison Project). When used as predictors for changes in local climate variables, and using the prior established quantitative relationship between T_{global} and the local climate variable X, these 190 T_{global} time series can be used to generate 190 equally probable time series of the local climate variable time series emulating X.

The EPIC method does not attempt to faithfully represent the full, true PDF of potential tuning parameters both for the AOGCM tunings and the carbon cycle model tunings i.e. were MAGICC tuned to a different set of AOGCMs (e.g. the CMIP5 set rather than the CMIP3 set), we would obtain a different set of tuning files which could lead to a somewhat different spread in our generated ensembles. The purpose of this paper is not to generate perfect ensembles that encapsulate structural model uncertainty in a completely accurate way, but rather to describe a method that provides a better representation of that uncertainty than can be achieved with only a limited set of RCM simulations. The robustness of the EPIC method depends on the set of AOGCM and carbon cycle model tunings available and as more comprehensive sets (that better reflect the likelihood of some tunings over others) become available, we expect that the large ensembles generated by EPIC to better reflect the true underlying uncertainties.

2.3 Virtual Climate Station Network

While the RCM simulations have been partially bias corrected, we recognise that some biases may remain. Therefore, we build our projections off an observational data set, so that, in the absence of any forced changes in climate, the projections default to observations (this is described in greater detail below). Observationally-based time series are obtained from the so-called Virtual Climate Station Network (VCSN). The VCSN data set for the New Zealand land surface is constructed on a regular $0.05° \times 0.05°$ grid from spatially inhomogeneous and temporally discontinuous quality controlled weather station data (Tait et al., 2005). The values estimated on the $0.05° \times 0.05°$ grid are based on thin plate smoothing spline interpolation using a spatial interpolation model as described in Tait, (2008).

3 Methodology

For a given geographic location, each ensemble member, covering the period 1960 to 2100, is constructed from contributions including:

1. a climatology derived from observations that represents the time invariant part of the signal,
2. a contribution from long-term forced changes in the magnitude of the variable of interest where those changes scale with changes in anomalies in global mean surface temperature (T'_{global}), and
3. a contribution from weather, generated by a stochastic weather generator that incorporates both forced and unforced variability.

The construction of each of these signals is described in greater detail below with a high level overview of how these contributions are related shown in Figure 1. The methodology described below pertains to a selected single greenhouse gas (GHG) emissions scenario and the daily maps of the climate variable of interest (X; here daily maximum (T_{max}) and daily minimum (T_{min}) surface temperatures) are obtained from one or more RCM simulations. The results presented in this study were obtained from a 1900 member ensemble (10 ~~Monte Carlo resampled alpha values were choosen for each of the 190 T_{global} time series from MAGICC~~ensemble members were generated for each T_{global}), that was generated over the period 1960 to 2100. The T_{max} and T_{min} fields were obtained from five RCM simulations driven by the Representative Concentration Pathway (RCP) 8.5 GHG emissions scenario for the period 1971-2100. RCP 8.5 was choosen ~~to be presented due to the~~ as it displays a high climate signal to noise ratio~~which results in more~~ , resulting in the most robust regression results, but the methodology is valid for any choosen GHG emission scenarios assuming a robust regression fit is obtained during the training phase. The assumption, which has been verified (not shown here), is that the dependence of X on T_{global} is independant of the GHG emmisions scenario used for the training. All anomalies were calculated with respect to the period 2000 to 2010. This anomaly period was choosen as the change in X over the 21^{st} century was of interest.

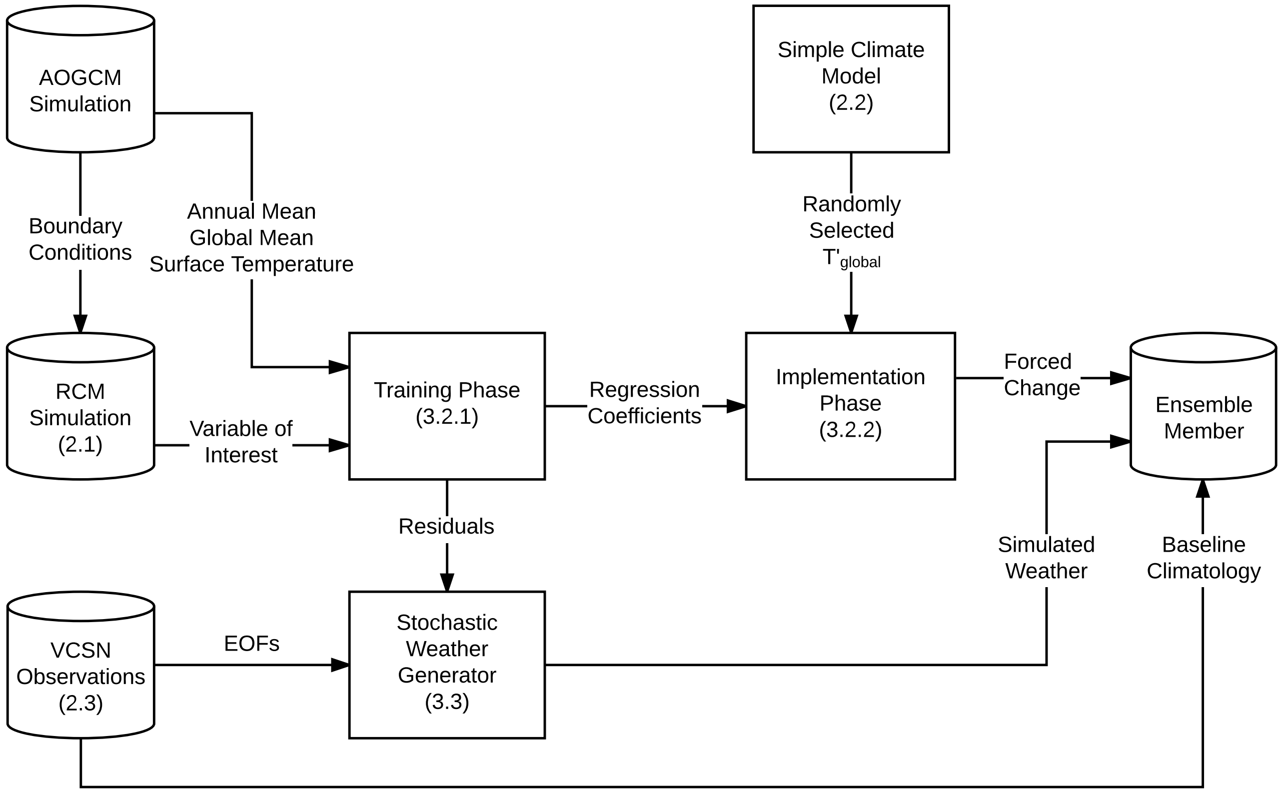

Figure 1. Flow chart illustrating the processes involved in generating a single EPIC ensemble member from a selected RCM simulation. Numbers in brackets refer to the section where more details are provided.

3.1 The climatology

At each 0.05° by 0.05° (approximately 5km) grid point, a mean annual cycle is calculated from daily observational data from 2001 to 2010. For this study, these observational data were obtained from VCSN (Section 2.3). Since the 10 year baseline period is rather short, a climatology derived by calculating calendar day means would still contain some weather-induced noise. Therefore, a regression model which includes an offset basis function~~expanded using the first~~ , expanded using two Fourier series expansions (~~see Section 3.2.1 and Section~~ Fourier pairs) to account for seasonality (see Section 2.4 of Kremser et al. (2014)), is fitted to the daily observational data to obtain the mean annual cycle. The first 2 Fourier series expansions are given in Eq. 1.

$$\beta(d) = \beta_0 + \beta_1 \times \sin(2\pi d/365) + \beta_2 \times \cos(2\pi d/365) + \beta_3 \times \sin(4\pi d/365) + \beta_4 \times \cos(4\pi d/365) \quad (1)$$

where d is the day number of the year and β is a the regression coefficient being expanded. By using an offset basis function expanded in Fourier seriespairs, the resultant mean annual cycle is smooth. Examples of the mean annual cycle are shown in Figure 2 for four selected locations around New Zealand.

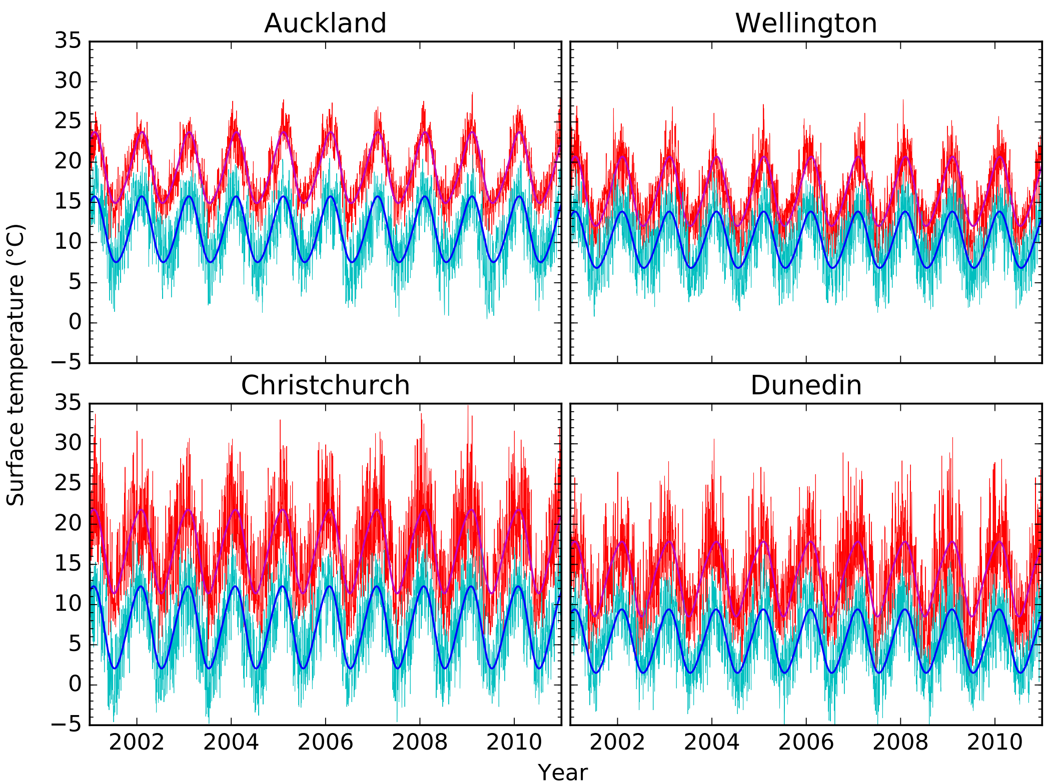

Figure 2. Observations of daily maximum surface temperature (magentared) and daily minimum surface temperature (cyan) from VCSN together with the mean annual cycle obtained from the regression model fit to the daily maximum surface temperature (redmagenta) and the daily minimum surface temperature (blue) time series for four selected locations in New Zealand, over the period 2001 to 2010.

This repeating mean annual cycle then provides the stationary baseline for the entire period of interest e.g. 1960 to 2100.

3.2 Direct response to T'_{global}

3.2.1 Training Phase

To determine In the training phase, the first-order long-term change in the magnitude of X, a forced change in X is established using the correlation between X' and T'_{global}. This relationship is expected to be dependent on the RCM simulation from which the variable of interest is obtained. There are two ways in which this can be managed, viz.:

1. A statistical relationship is quantified for each RCM simulations providing data for the training phase of EPIC. Then, in the ‘implementation phase’ of EPIC (see below), for each ensemble member, a single relationship is randomly selected to be applied.

2. A single statistical relationship is quantified using a concatenated time series obtained from all RCM simulations providing data for the training phase of EPIC. In the ‘implementation phase’, this relationship is used.

For the purposes of this study, method (1) is used as method (2) will tend to underestimate the true uncertainty of the relationship between X' and T'_{global}.

A simple linear correlation between X' and T'_{global} is calculated for each of the 5 RCM simulations and each grid point independently, viz.:

$$X'(d,yt) = \alpha(d) \times T'_{global}(y) + R(t) \quad (2)$$

where $X'(d,y)$ $X'(t)$ are the daily anomalies with respect to the 2001-2010 mean annual cycle of X (see Section 3.1), $T'_{global}(y)$ are the anomalies of an annual mean global mean surface temperature time series, and $\alpha(d)$ α is the fit coefficient . X', rather than X, is used in equation 2 as the change in the seasonal cycle is of interest. The residuals from this fit are then and R is the residual which is the part of the signal that cannot be explained by the statistical model. In this case, the residuals are used by the stochastic weather generator (see Section 3.3) to model higher order changes in the variability in X which are not captured by Eq. (2).

Because GCM or The mean annual cycle of X, which is used to calculate X', is generated using the same method and time period used to calculate the mean annual cycle of the observational set. X', rather than X, is used in Eq. 2 as the change in the seasonal cycle is of interest. Removing the mean annual cycle removed the need to add additional terms to describe the baseline seasonal cycle.

Because GCM and RCM output provide a much longer time series than observations and extend into a period of greater changes in X, GCM or RCM output are preferentially used in this training phase. When using projections from RCMs as input to the training phase of EPIC, T'_{global} is sourced from the GCM that provided the boundary conditions for the RCM simulation.

Because the fit coefficient, α, is expected to depend on season, it is expanded in two Fourier pairs to account for its seasonality i. e.

$$\alpha(d) = \alpha_0 + \alpha_1 \times \sin(2\pi d/365) + \alpha_2 \times \cos(2\pi d/365) + \alpha_3 \times \sin(4\pi d/365) + \alpha_4 \times \cos(4\pi d/365)$$

where d is the day number of the year(Eq. 1). The resulting α has a smooth seasonal cycle which would not be the case if each month was fitted independently. When embedded in Eq. (2), this expansion results in the resulting Eq. (3), has five fit coefficients for α(α_0 to α_4).

$$X'(t) = (\alpha_0 + \alpha_1 \times \sin(2\pi d/365) + \alpha_2 \times \cos(2\pi d/365) + \alpha_3 \times \sin(4\pi d/365) + \alpha_4 \times \cos(4\pi d/365)) \times T'_{global}(y) + R(t) \tag{3}$$

The statistical model is solved using a multivariate least squares regression approach (Moore and McCabe , 2003) to obtain the fit coefficients. We refer to each such set of five fit coefficients as a tuple; recall that this fit is applied at each grid point and for each available RCM simulation.

An example of a fit of Eq. (23) to daily maximum surface temperature anomalies is shown in Figure 3 for a location in the South Island of New Zealand.

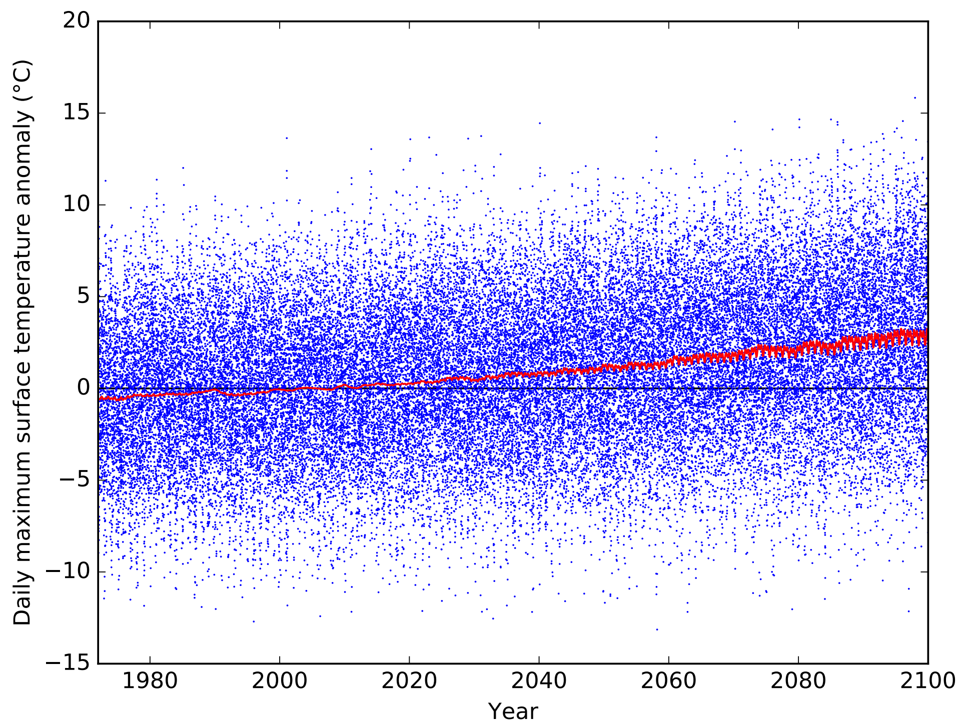

Figure 3. An example of the fit of Eq. (23) (red line) to daily maximum surface temperature anomalies (blue) obtained from the NorESM1-M RCM simulation under the RCP8.5 GHG emissions scenario at Alexandra, New Zealand (45.249°S, 169.396°E). Solid line represents the zero line (no change).

The small annual cycle in the fit, with growing amplitude, results from summer-time and winter-time daily maximum surface temperatures exhibiting different correlations against T'_{global}. The inter-annual variation arises from changes in T_{global} as α

does not change year to year. In addition to the long-term forced change, there is significant day-to-day variability. The use of the residuals from such fits in the stochastic weather generator model is described in Section 3.3.

The unitless α coefficient describes a locations sensitivity to changes in annual mean global mean surface temperature. The magnitude of α indicates whether T_{max} or T_{min} are changing faster (α>1) or slower (α<1) than the global mean surface temperature. Example maps of the magnitude of the α coefficient for four selected days in New Zealand are shown in Figure 4. Maps of α coefficients (unitless) which represent the sensitivity of changes in annual mean global mean maximum surface temperature for locations throughout New Zealand. The α coefficients were derived from fits of Eq. (2) to daily time series of daily maximum temperatures at each grid point of the NorESM1-M RCM simulation. The annual mean global mean surface temperature anomalies were taken from the AOGCM simulation that provided the boundary conditions for this particular RCM simulation. Black line indicates α values of 1.0. This analysis shows that daily maximum surface temperatures over most of New Zealand are warming slower than T_{global}. However, high altitude regions, such as the Southern Alps, indicate T_{max} increasing faster than T_{global} for southern hemispheric spring, summer and autumn.

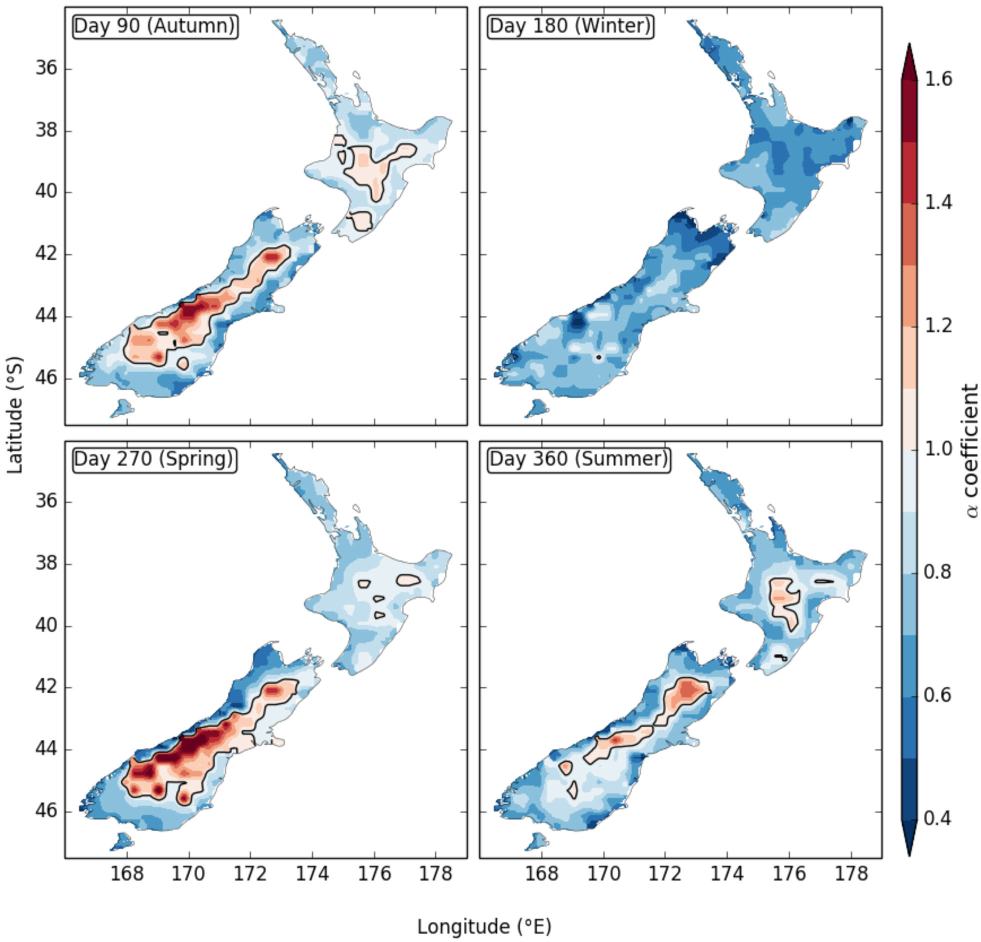

Figure 4. Maps of α coefficients (unitless) which represent the sensitivity of changes in annual mean global mean maximum surface temperature for locations throughout New Zealand. The α coefficients were derived from fits of Eq. (3) to daily time series of daily maximum temperatures at each grid point of the NorESM1-M RCM simulation. The annual mean global mean surface temperature anomalies were taken from the AOGCM simulation that provided the boundary conditions for this particular RCM simulation. Black line indicates α values of 1.0.

There is, of course, some uncertainty in α. To account for that uncertainty, a large set of α tuples is derived through a Monte Carlo bootstrapping approach (Efron and Tibshirani , 1994), whereby residuals from the Eq. (23) fit are randomly sampled and added to the regression model fit to generate multiple statistically equivalent time series which are then refitted to obtain equally probable α fit coefficients (Bodeker and Kremser, 2015). This approach allows for the incorporation of the uncertainty in the fit of Eq. (23) into the final ensemble of projections.

It is also expected that α will depend on the RCM simulation used for the training. There are two ways in which this can be managed, viz.:

1. ~~Tuples of α values are obtained for each RCM simulation providing data for the training phase of EPIC. Then, in the 'implementation phase' of EPIC (see below), for each ensemble member, a tuple of α values is selected from the Monte Carlo-derived set of α tuples from a RCM data set that is randomly selected.~~

2. ~~Tuples of α values are obtained simultaneously for all RCMs by fitting Eq. (2) to concatenated time series of X' and $\mathrm{T}'_{global}(y)$ obtained from all RCM simulations providing data for the training phase of EPIC.~~

~~For the purposes of this study, method (1) is used as method (2) will tend to underestimate the true uncertainty seen in the values of α.~~

3.2.2 Implementation Phase

Once the Monte Carlo derived sets (just one set if method (2) is used) of α tuples have been obtained, they are used in the implementation phase of EPIC. As described in Section 2.2, 190 simulations of T'_{global} can be generated using a SCM. A randomly selected T'_{global} time series from the 190 member set is used together with a randomly selected tuple of α values to generate a series of maps of X'_{forced} using Eq. (~~2~~3), where the forced subscript denotes these are changes which correlate with T'_{global}

There might be some concern that the random selection of an α tuple from the available set of tuples for a location could cause the spatial coherence in the forced signal across New Zealand to be lost, as at a nearby location a different tuple could be randomly selected. This was tested for and was found not to be the case as the multiple instances of tuples (multiple instances of Figure 4) are very similar and consistent (not shown here).

3.3 Indirect response to T'_{global} and weather noise

In addition to the change in X that correlates directly with T'_{global}, higher order components of variability, as well as realistic weather noise, must be present in the projections comprising the ensemble. One potential use of the ensemble of projections generated by EPIC is assessment of the impacts and implications of climate change on a regional scale. These impacts seldom happen at a single site, i.e. the impact is felt over a large area. For this reason it is important that any specific member of an ensemble is appropriately spatially coherent over multiple sites. This is not achieved if the method considers each site in isolation since any purely stochastically determined weather noise added to a site would not be spatially coherent at neighbouring sites. For this reason, an empirical orthogonal function (EOF) approach, described by Lorenz (1956), is used to ~~reveal the~~ describe the spatial weather patterns and how they change over time. EOF analysis is a statistical method which reveals the spatial patterns, or modes of variability ~~inherent in a series of daily maps of X' after dependence on T'_{global} has been removed, and their~~ in a data set, and how these patterns over time as given by the resulting principal component (PC) time series. Hereafter we refer to these modes of variability as ~~'~~'weather modes'. The EOF analysis is applied to X' after the dependence on T'_{global} has been removed. These weather modes, and ~~stochastically generated~~ PC time series~~(PC_{syn})~~, are then used to construct a ~~stochastic~~ weather generator which ~~generates~~ produces realistic weather noise by stochastically generating PC time series (PC_{syn}). The following is recognised in the construction of the stochastic weather generator model:

1. That VCSN data will provide the most realistic representation of weather noise.
2. That RCM simulations will simulate how that weather noise is likely to evolve in response to climate change (represented by T'_{global}).
3. That the RCM simulations will be imperfect in simulating the patterns of variability derived from the VCSN data.
4. That there will be ~~weather modes~~ patterns of variability (weather) whose amplitude and variability will respond to climate change as well as others which ~~will not~~ do not change with increases in T'_{global}.

3.3.1 Identifying the modes of variability responding to climate change

We begin by conducting an EOF analysis on VCSN data that have been detrended by removing the first order trend and on residuals from the ~~training phase in Section 3.2.1~~fit of Eq. 3 to RCM data in the training phase. Where the patterns of variability obtained from from EOF projections of VCSN and RCM diverge is considered to be the cut-off point for where the RCM simulation can be taken to have any integrity with regard to simulating forced changes in weather noise. Visual inspection of the EOF maps derived from VCSN and RCM data suggested that the first four modes of variability were well represented by the RCM simulations (see Figure 5).

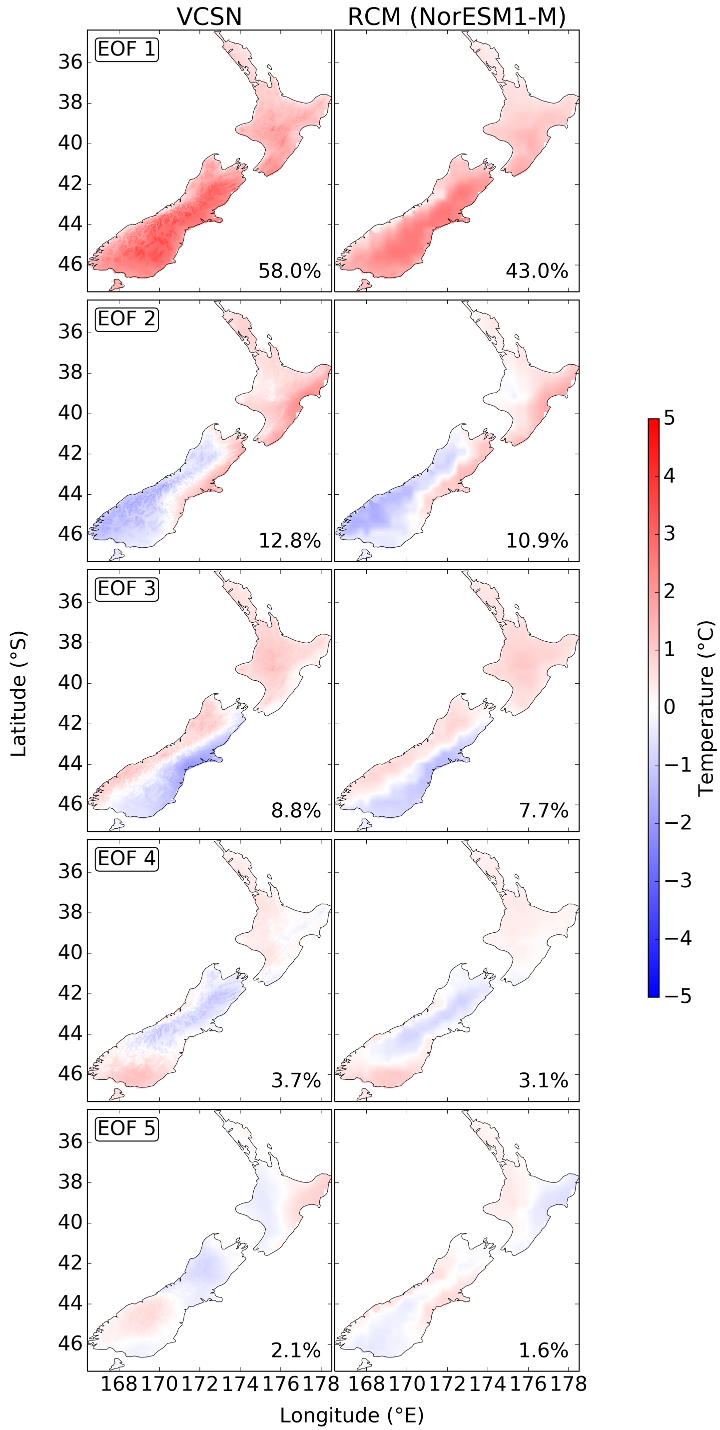

Figure 5. The first five EOF patterns of weather noise in daily maximum surface temperatures obtained from VCSN data from 1972 to 2013 (left column) and obtained from RCP8.5 NorESM1-M RCM output from 1972 to 2100. The colour bar shows the amplitude of the pattern in °C. The percentage values in each panel show the fraction of the total variability explained by each mode.

It is clear from Figure 5 that the RCM EOFs exhibit the same modes of weather variability as seen in the VCSN data up until EOF pattern 4. These first four patterns of variability together explain 83.3% of the total weather variability in the VCSN data and 64.7% of the variability in the RCM data. It is these four modes of weather variability that evolve with T'_{global} in our stochastic weather generator.

3.3.2 Modelling forced changes in the amplitude and variability of weather modes

To compare statistics from the PC time series calculated from VCSN and RCM data, they must share the same underlying weather modes. This is done by projecting the VCSN weather modes (EOF_{VCSN}) onto the RCM data to calculate a pseudo-PC time series. A pseudo-PC time series is calculated in the same way that a standard PC time series is calculated, except that the weather modes are prescribed instead of being calculated from the data. A pseudo-PC time series describes the magnitude of a particular pattern of variability from VCSN, which is present in the RCM data. The VCSN weather modes, rather than the RCM weather modes, were prescribed because the observational data set is more likely representative of patterns of variability seen in New Zealand. The pseudo-PC and $VCSN_{PC}$ time series can be compared as they both describe the same patterns of variability.

In the stochastic weather generator, we consider changes in:

1. The amplitude of the weather mode: this is quantified by correlating the associated pseudo-PC time series with T'_{global} and then using that correlation coefficient (β) to drive a trend in the PC time series obtained from the VCSN-based EOF analysis.

2. The variability of the weather mode: this is quantified by correlating the variability in the associated pseudo-PC time series with T'_{global} and then using that correlation coefficient (β_{var}) to drive a trend in the variability of the PC time series obtained from the VCSN-based EOF analysis. The mean variability of the weather mode is obtained from the VCSN PC time series rather than the pseudo-PC time series, so that the weather mode emulates the magnitude of variability seen in the VCSN data.

We also recognize that the PC time series will exhibit temporal auto-correlation and therefore, that correlation is quantified and removed before correlating the PC signal, and its variability, against T'_{global}. The resulting time series (PC_{syn_n}) captures both long-term shifts and/or changes in spread of the n^{th} weather mode. We note, however, that by considering only lag-one autocorrelation in these PC time series, we neglect longer term auto-correlation, e.g. that resulting from El Niño and La Niña events. As a result, our ensemble time series exhibit smaller ~~interannual~~ inter-annual variability than is observed in VCSN time series.

The ability of the method to generate a set of PDFs of the PC_{syn_1} to PC_{syn_4} time series is demonstrated in Figure 6.

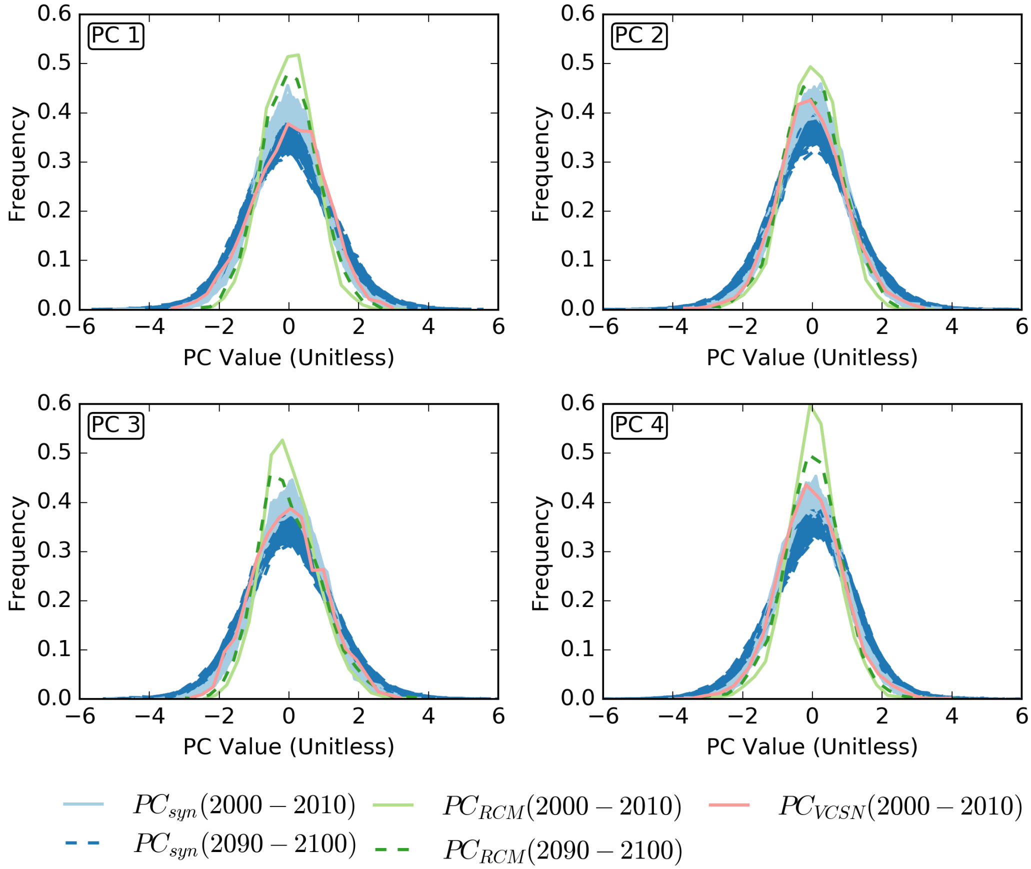

Figure 6. PDFs of the first four synthetic (PC$_{syn}$) and RCM (PC$_{RCM}$) PC time series for the first decade of the 21st century are shown in solid lines and PDFs for the last decade of the 21st century are shown in dashed lines. PC$_{RCM}$ and PC$_{syn}$ were both derived from the NorESM1-M RCM output as an example. The PDF from the PC time series (2000-2010) obtained from VCSN is also shown (PC$_{VCSN}$). The disagreement between the PC$_{RCM}$ and PC$_{VCSN}$ validates the use of VCSN weather noise as the basis for our stochastic weather generator model and the good agreement between the ensemble of PC$_{syn}$ and PC$_{VCSN}$ demonstrates that the EPIC method generates synthetic PC time series with a degree of variability that matches reality.

The EPIC method corrects for any shortcomings in the ability of the RCM to correctly simulate expected magnitudes of weather variability for these four primary modes and then accommodates these corrections when generating PC time series that evolve into the future.

3.3.3 Modelling higher order modes of variability in weather

The stochastic weather generator model includes the effects of EOF patterns five and higher but assumes that these modes show no dependence on T$'_{global}$ as the RCM simulations do not accurately simulate these higher modes of weather variability. The variability of the PC time series often has a strong seasonal cycle. Therefore, for EOF pattern five and higher, synthetic PC time series (PC_{syn}) are generated using a standard Monte Carlo approach, i.e. randomly selecting values from $N(0,\sigma(d))$, that is, a normal distribution with a mean of 0 and a standard deviation which depends on the day of the year which is being modelled. $\sigma(d)$ is determined by a zeroth-order fit expanded in linear fit of two Fourier pairs (Eq. 1) to the VCSN PC time series. The Fourier pairs model the seasonal cycle in the PC time series. This approach allows selection of extreme PC values that are outside of the range of PC values experienced in the 1972-2013 period, but noting that the PDFs of these PCs do not evolve with time. As with the forced changes in the amplitude and variability of weather modes, the auto-correlation in the PC time series is also quantified and captured in the statistically modelled PC time series.

For a given ensemble member, once synthetic PC time series at daily resolution have been generated, they are used to produce a reconstructed weather field, W, according to:

$$W(i,j,t) = \sum_{n=1}^{50} EOF_{VCSN\_n}(i,j)PC_{syn\_n}(t)$$

where i, j and t represent the latitude, longitude and time dimensions respectively and n is the nth weather mode.

Since W has been constructed from a linear combination of spatial patterns of variability, each of which is spatially coherent, it retains the property of spatial coherence. The variability evolves as expected under changes in T'_{global} for the first four modes of variability, as simulated by the RCM, and where extreme conditions, outside the range of the training period, do occur with a statistically reasonable frequency due to the stochasticity in the construction of the pseudo PC time series.

T_{min} is modelled identically to T_{max} with one small change: days with anomalously low T_{max} would be more likely to have anomalously low T_{min}. Not accounting for this correlation could result in stochastically modelled T_{min} values being higher than the modelled T_{max} value for that day. To avoid that, and to capture the correlation between T_{max} and T_{min} on any given day, the same set of random numbers used to generate the values in the synthetic PC_n time series for T_{max} for a given day is used to generate the values in the synthetic PC_n time series for T_{min}. This forces the selection of PC_{syn} values from the same region of the PDF for both T_{max} and T_{min}.

4 Results

Examples of the T_{max} and T_{min} time series generated by the EPIC method are shown in Figure 7 for four population centres in New Zealand together with the associated VCSN time series.

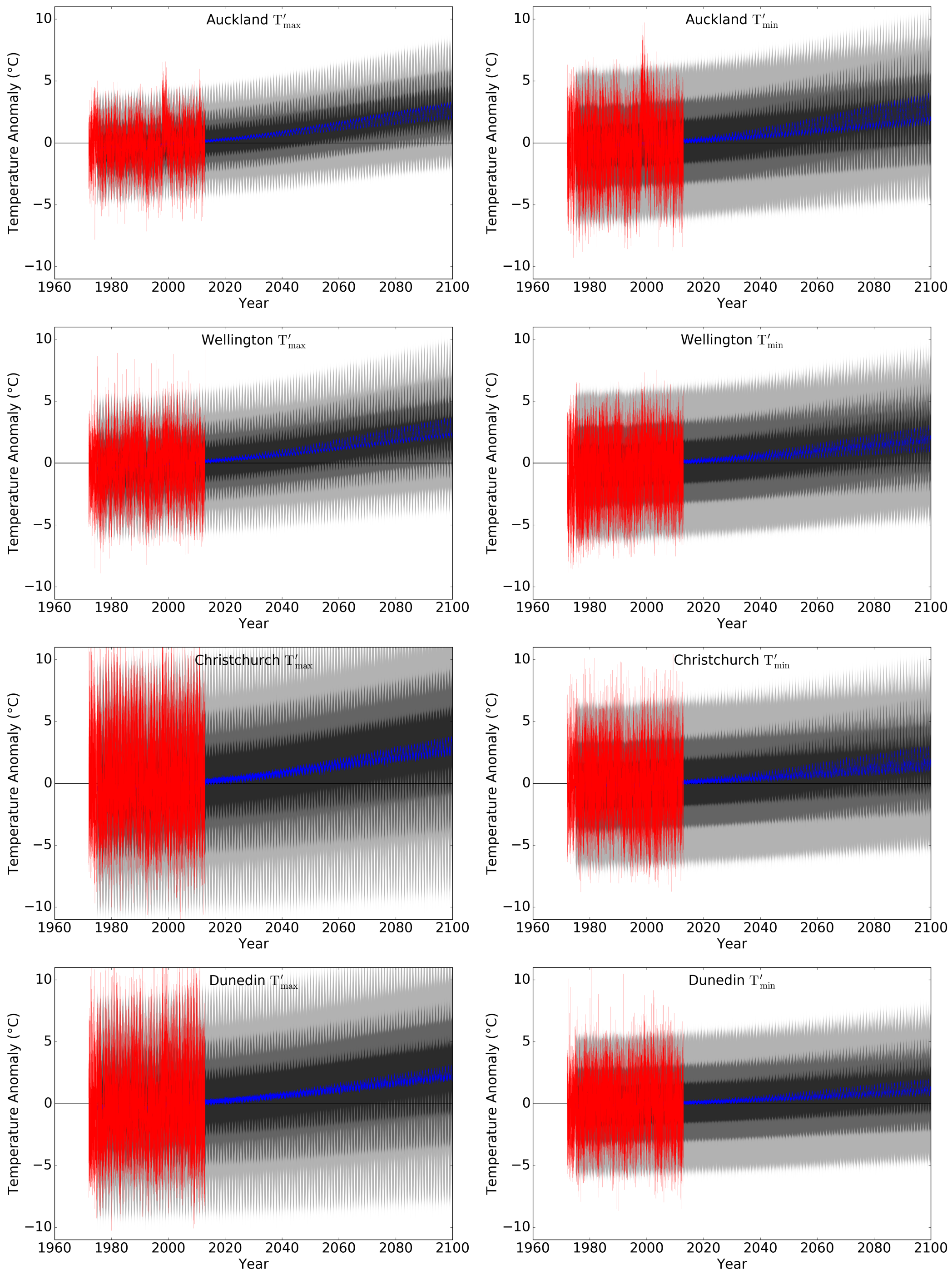

Figure 7. Example output from 1900 EPIC-generated time series for Auckland, Wellington, Christchurch and Dunedin from 1960 to 2100 under the RCP8.5 GHG emissions scenario. Grey shaded areas show the 1, 10, 25, 75, 90 and 99 percentiles while the blue line shows the median value on each day. T'_{max} (left column) and T'_{min} (right column) anomalies with respect to the 2000-2010 mean annual cycle. VCSN time series are overlaid in each panel (red lines).

Actual EPIC ensemble time series add these anomaly time series to the 2001-2010 VCSN-derived annual cycle climatology and therefore show no systematic bias with respect to the VCSN data. The EPIC-generated time series also show a long-term evolution consistent with expectations from RCM simulations, including the effects of the spread in those simulations. While it cannot be directly seen from the time series plotted in Figure 7, the EPIC-generated time series also exhibit changes in weather variability consistent with RCM projections of expected changes in the first four modes of weather variability. The apparent annual cycle in the anomaly time series reflects the annual cycle in the variance and not an annual cycle in the anomalies; towards the end of the period there is a true annual cycle in the anomalies from differential seasonal changes in T_{max} and T_{min}. The ~~interannual~~ inter-annual variability of the EPIC ensemble members is lower than that of the observational data set. This is due to EPIC not including any terms which describe patterns of variability which occur at time scales of longer than 1 year.

5 Discussion and Conclusions

The EPIC (Ensemble Projections Incorporating Climate model uncertainty) method, is able to generate large ensembles of daily time series of daily maximum and minimum temperatures that exhibit the following characteristics:

– No bias with respect to VCSN data.

– Long-term evolution consistent with projections from a suite of RCM simulations, incorporating the uncertainties inherent in those simulations as well as additional structural uncertainties that may arise from the use of a wider suite of RCMs as captured by the use of projections of T'_{global}. T'_{global} time series were generated by a SCM tuned to 19 different AOGCMS and 10 different carbon cycle models and used as a predictor for the long-term change in T_{max} and T_{min}.

– Weather variability with extremes that extend beyond that observed in the VCSN record and that evolve in a way consistent with RCM projections of changes in the four primary modes of weather variability.

– Spatial coherence in weather variability in any single ensemble member is preserved.

As such, EPIC-generated projections are suitable for generating robust PDFs of projections of T_{max} and T_{min}.

The number of members in each ensemble is essentially limited only by the computing resources available. The stochasticity introduced by the Monte Carlo analysis and modeling of the weather noise allows for many ensemble members to be generated for a given T_{global}. For calculating the PDFs that are delivered to users, we currently generate 19,000 ~~members in each ensemble (one ensemble for each RCP scenario)~~ member ensemble (10 ensemble members for each T_{global}) for a given RCP scenario at each 0.05° × 0.05° grid point across New Zealand.

A web-based tool has been developed to deliver PDFs of T_{max} and T_{min} for the period 2001-2010 and 2091-2100 to users along with statistics regarding the change in frequency of extreme events, i.e. days per year with T_{max} above 25°C and 30°C and T_{min} below 0°C and 2°C. The tool is available at http://futureextremes.ccii.org.nz/.

The next steps for the development of EPIC include extending the range of climate variables to daily surface broadband radiation, surface humidity and precipitation, and incorporating longer term sources of variability, e.g. those generated by

El Niño and La Niña events, into the stochastic weather model. The implementation of a model weighting scheme, such as Knutti et al. (2017), for the training data could increase the applicability of the model.

6 Code and data availability

The source code and data used is available upon request to the corresponding author. The VCSN data set employed is available from NIWA (https://www.niwa.co.nz/climate/our-services/virtual-climate-stations).

Acknowledgements. This research was funded by the Ministry of Business, Innovation and Employment as part of the Climate Changes, Impacts and Implications (contract C01X1225). We would also like to thank Malte Meinshausen for providing the tuning files used to emulate various AOGCMs using MAGICC, and Abha Sood for assistance with the RCM data.

References

Ackerley, D., Dean, S., Sood, A., and Mullan, A.B.: Regional climate modeling in NZ: Comparison to gridded and satellite observations, Wea. Clim., 32, 3-22, 2012.

Bhaskaran, B., Mullan, A.B., and Renwick, J.: Modelling of atmospheric variation at NIWA, Wea. Clim., 19, 23-36, 1999.

Bhaskaran, B., Renwick, J., and Mullan, A.B.: On application of the Unified Model to produce finer scale climate information, Wea. Clim., 22, 19-27, 2002.

Bodeker, G.E., and Kremser, S.: Techniques for analyses of trends in GRUAN data, Atmos. Meas. Tech., 8, 1673–1684, 2015

Drost, F., Renwick, J., Bhaskaran, B., Oliver, H., and MacGregor, J.L.: Simulation of New Zealand's climate using ~~high resolution~~ a high-resolution nested regional climate model, Intl. J. Climatol., 27, 1153-1169, 2007.

Efron, B. and Tibshirani, R.J.: An Introduction to the Bootstrap, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, ISBN 9780412042317, 1994.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H.D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A.J., Yoshikawa, C., and Zeng, N.: Climate–Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison, Journal of Climate, 19, 3337-3353, 2006.

Gordon, C., Cooper, C., Senior, C.A., Banks, H., Gregory, J.M., Johns, T.C, Mitchell, J.F.B., and Woods, R.A.: The simulation of SST, sea ice extents and ocean heat transports in a version of Hadley Centre coupled model without flux adjustments, Clim. Dyn., 16, 559, 147-168, 2000.

Gregory, D., Smith, R.N.B., and Cox, P.M.: Canopy, surface and soil hydrology, version 3. Unified model documentation Paper 25, UK Met Office, 1994.

Harris, G.R., Collins, M., Sexton, D.M.H., Murphy, J.M. and Booth, B.B.B.: Probabilistic projections for 21st century European climate, Natural Hazards and Earth System Sciences, 10, 2009-2020, doi:10.5194/nhess-10-2009-2010, 2010.

Jones, R., Noguer, M., Hassell, D.C., Hudson, D., Wilson, S.S., Jenkins, G.J., and Mitchell, J.F.B.: Generating high resolution climate change scenarios using PRECIS, Tech. rep., Met Oce Hadley Centre, UK, 40 pp., 2004.

Knutti, R.; Masson, D. and Gettelman, A., Climate model genealogy: Generation CMIP5 and how we got there, Geophysical Research Letters, 40, 1194-1199, doi:10.1002/grl.50256, 2013.

Knutti, R.; Sedláèek, J.; Sanderson, B.M.; Lorenz, R.; Fischer, E.M. and Eyring, V., A climate model projection weighting scheme accounting for performance and interdependence, Geophysical Research Letters, 44, 1909-1918, doi:10.1002/2016GL072012, 2017.

Kremser, S., Bodeker, G.E., and Lewis, J.: Methodological aspects of a pattern-scaling approach to produce global fields of monthly means of daily maximum and minimum temperature, Geosci. Model Dev., 7, 249-266, 2014

Lorenz, E.N.: Empirical orthogonal functions and statistical weather prediction, Scientific Report No. 1, Statistical Forecasting Project, Massachusetts Institute of Technology, Department of Meteorology, pp. 52, 1956.

Masson, D. and Knutti, R., Climate model genealogy, Geophysical Research Letters, 38, L08703, doi:08710.01029/02011GL046864, 2011.

Meehl, G.A., Stocker, T.F., Collins, W.D., Friedlingstein, P., Gaye, A.T., Gregory, J.M., Kitoh, A., Knutti, R., Murphy, J.M., Noda, A., Raper, S.C.B., Watterson, I.G., Weaver, A.J., and Zhao, Z.-C.: Global Climate Projections. In: , Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2007

Meinshausen, M., Raper, S.C.B., and Wigley, T.M.L.: Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration, Atmos. Chem. Phys., 11, 1417–1456, 2011.

Meinshausen, M., Wigley, T.M.L. and Raper, S.C.B.: Emulating atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 - Part 2: Applications, Atmos. Chem. Phys., 11, 1457–1471, 2011.

Moore, D.S., and McCabe, G.P.: Introduction to the practice of Statistics, W.H. Freeman and Company, New York, pp. 828, 2003.

Murphy, J.M., Booth, B.B.B., Collins, M., Harris, G.R., Sexton, D.M.H. and Webb, M.J.: A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles, Philosophical Transactions of the Royal Society A, 365, 1993-2028, doi:10.1098/rsta.2007.2077, 2007.

Murphy, J. M., Sexton, D. M. H., Jenkins, G. J., Boorman, P. M., Booth, B. B. B., Brown, C. C., Clark, R. T., Collins, M., Harris, G. R., Kendon, E. J., Betts, R. A., Brown, S. J., Howard T. P., Humphrey, K. A., McCarthy, M. P., McDonald, R. E., Stephens, A., Wallace, C., Warren, R., Wilby, R., and Wood, R. A.: UK Climate Projections Science Report: Climate change projections. Met Office Hadley Centre, Exeter, UK, available at: http:// ukclimateprojections.defra.gov.uk/content/view/824/517/, 2009.

Pope, V. D., Gallani, M. L., and Rowntree, P. R.: The impact of new physical parametrization in the hadley centre climate model: HadAM3. Clim. Dyn., 16, 123-146, 2000

Pope, V.D. and Stratton, R.A.: The process governing horizontal resolution sensitivity in a climate model. Clim. Dyn., 19, 211-236, doi:10.1007/s00382-001-0222-8, 2002.

Reisinger, A., Meinshausen, M., Manning, M., and Bodeker, G.: Uncertainties of global warming metrics: CO_2 and CH_4, Geophys. Res. Lett., 37, L14707, doi:10.1029/2010GL043803, 2010

Sexton, D.M.H., Murphy, J.M., Collins, M. and Webb, M.J.: Multivariate probabilistic projections using imperfect climate models part I: outline of methodology, Climate Dynamics, 38, 2513-, doi:10.1007/s00382-011-1208-9, 2012.

Sood, A.: Improved Bias Corrected and Downscaled Regional Climate Model Data for Climate Impact Studies: Part 1 Validation and Assessment for New Zealand, J. Clim, submitted 2016.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., and Miller, H.L. (eds.) Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Tait, A. and Turner, R.: Generating Multiyear Gridded Daily Rainfall over New Zealand, J. App. Meteor., 44, 1315-1323, 2005

Tait, A.B., Future projections of growing degree days and frost in New Zealand and some implications for grape growing, Weather and Climate, 28, 17-36, 2008.

Watterson, I.G.: Calculation of probability density functions for temperature and precipitation change under global warming, J. Geophys. Res., 113, 2008