Responses to referee comment [RC1]

Interactive comment on "Chilling accumulation in temperate fruit trees in Spain under climate change" by Alfredo Rodríguez et al.

We thank the reviewer for his thoughtful comments. Our answers are highlighted in green italics.

Rodriguez et al. present an assessment of past and future winter chill in Spain, using an ensemble of climate scenarios and four chill models. It seems to me that the climate data processing was very well done; the way scenarios were prepared seems very reasonable. The authors' expertise in this field is evident.

Thank you for this comment

Unfortunately, the study has some shortcomings regarding the estimation of winter chill, which will have to be addressed.

Major issues:

1) Similar work has been done before, for various countries and also at global scale. It remains somewhat unclear what the particular advantage of this new approach is. A (smaller) ensemble approach was already used 10 years ago (Luedeling et al., 2009a) for California and shortly afterwards at the global scale (Luedeling et al., 2011). In these studies, we used a weather generator rather than just climate model outputs, which (in my view) makes the methodology used then more robust than what is presented here. Admittedly, some other elements of these assessments were not as well done as what is described in the current manuscript, and it's good to see a study using RCPs rather than SRES scenarios (though we did this here: Benmoussa et al., 2018, but not as a spatial analysis), but the novelty of the current methodology isn't sufficiently described.

We will soften the language about the novelty of our study throughout the paper and we will acknowledge as previous works the studies pointed by the reviewer (Luedeling et al. 2009, 2011). Besides, we will further describe the methodology followed to design the climate ensemble, enhancing the description of the improvements and contributions of our study in the text: 1) Studies done in other countries would be of little help for Spanish farmers that previously could only find scarce information from studies performed in other regions, or worldwide with not enough resolution; 2) In recent studies working with multi-model ensembles formed by crop models, ensemble results tend to improve as the number of members of the ensemble increases, for instance, in Martre et al. (2015) the committed errors decreased as the ensemble members grow with little decrease beyond 10 members. This debate was analysed from the statistical point of view in Wallach et al. (2018). We consider that this is an improvement from the former studies using 3 climate models. From this point only, in our view this work

represents an improvement in terms of robustness, due to the ensemble design and composition.

We will clarify the text to stress that in this study we did not use the climate model outputs directly. Instead, a bias adjustment process was applied to the outputs prior to be applied to the models. The bias adjustment techniques are considered a valid alternative to apply on climate model outputs to crop models, especially suitable for handling the complex orography of the Iberian Peninsula (Maraun and Widmann, 2018). Of course, weather generators can also be a reasonable approach.

Luedeling, E., Zhang, M., and Girvetz, E. H.: Climatic Changes Lead to Declining Winter Chill for Fruit and Nut Trees in California during 1950–2099, PLOS ONE, 4, e6166, 10.1371/journal.pone.0006166, 2009.

Luedeling, E., Girvetz, E. H., Semenov, M. A., and Brown, P. H.: Climate Change Affects Winter Chill for Temperate Fruit and Nut Trees, PLOS ONE, 6, e20155, 10.1371/journal.pone.0020155, 2011.

Maraun, D., and Widmann, M.: Statistical Downscaling and Bias Correction for Climate Research, Cambridge University Press, Cambridge, 2018.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J. W., Rotter, R. P., Boote, K. J., Ruane, A. C., Thorburn, P. J., Cammarano, D., Hatfield, J. L., Rosenzweig, C., Aggarwal, P. K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A. J., Doltra, J., Gayler, S., Goldberg, R., Grant, R. F., Heng, L., Hooker, J., Hunt, L. A., Ingwersen, J., Izaurralde, R. C., Kersebaum, K. C., Muller, C., Kumar, S. N., Nendel, C., O'Leary, G., Olesen, J. E., Osborne, T. M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M. A., Shcherbak, I., Steduto, P., Stockle, C. O., Stratonovitch, P., Streck, T., Supit, I., Tao, F. L., Travasso, M., Waha, K., White, J. W., and Wolf, J.: Multimodel ensembles of wheat growth: many models are better than one, Glob. Change Biol., 21, 911-925, 10.1111/gcb.12768, 2015.

Wallach, D., Martre, P., Liu, B., Asseng, S., Ewert, F., Thorburn, P. J., Ittersum, M., Aggarwal, P. K., Ahmed, M., Basso, B., Biernath, C., Cammarano, D., Challinor, A. J., De Sanctis, G., Dumont, B., Eyshi Rezaei, E., Fereres, E., Fitzgerald, G. J., Gao, Y., Garcia-Vila, M., Gayler, S., Girousse, C., Hoogenboom, G., Horan, H., Izaurralde, R. C., Jones, C. D., Kassie, B. T., Kersebaum, K. C., Klein, C., Koehler, A.-K., Maiorano, A., Minoli, S., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G. J., Palosuo, T., Priesack, E., Ripoche, D., Rötter, R. P., Semenov, M. A., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Wolf, J., and Zhang, Z.: Multimodel ensembles improve predictions of crop–environment–management interactions, Glob. Change Biol. (in press), doi:10.1111/gcb.14411, 2018.

2) Another innovation the authors point out isn't really a feature but rather a bug in my view. As highlighted on page 9, ll. 1-2, this may well be the first study that projected climate change impacts for these four chill models. However, there are good reasons for there not being more studies, in particular no recent studies. The reason is simply that most of these models can't be trusted to accurately describe chill accumulation. There have been a number of model comparisons over the years that have consistently found the Dynamic Model to be superior to the others (e.g. Benmoussa et al., 2017; Luedeling et al., 2009b; Ruiz et al., 2007; Zhang and Taylor, 2011; there are quite a few more). Adding old, obsolete models to such a study would be like adding a flat-earth model to a GCM ensemble – it makes little sense to consider models that have been shown to be inadequate. The situation with chill models is not the same as with GCMs – we do have a clear idea of which models are better, and there is really no rationale in my view to go for an ensemble approach.

We admit that the methodology has probably not been adequately transmitted, as this is a key point of this work. The ensemble approach was only considered for climate models but results from the different chill models were considered, calculated and interpreted individually. The difference is that while chilling projections calculated with different climate projections and the same chilling model have been averaged, chilling projections from different chilling models were not. We will try to clarify this point to avoid any impression of comparison between chilling model results in the results and discussion sections.

We are willing to discuss the validity of the models used. We agree with the first reviewer that there are several studies concluding that the Dynamic Model (DM) exhibits a higher accuracy than the Richardson based models (RbM, as Utah, North Carolina and De Melo-Abreu models). However, the reported improvement in the papers quoted by the referee is very small (e.g. Ruiz et al., 2007). Those studies also report varieties and locations where RbM models perform better. Also, some of these papers and others claim there is not a significant difference between models, for instance:

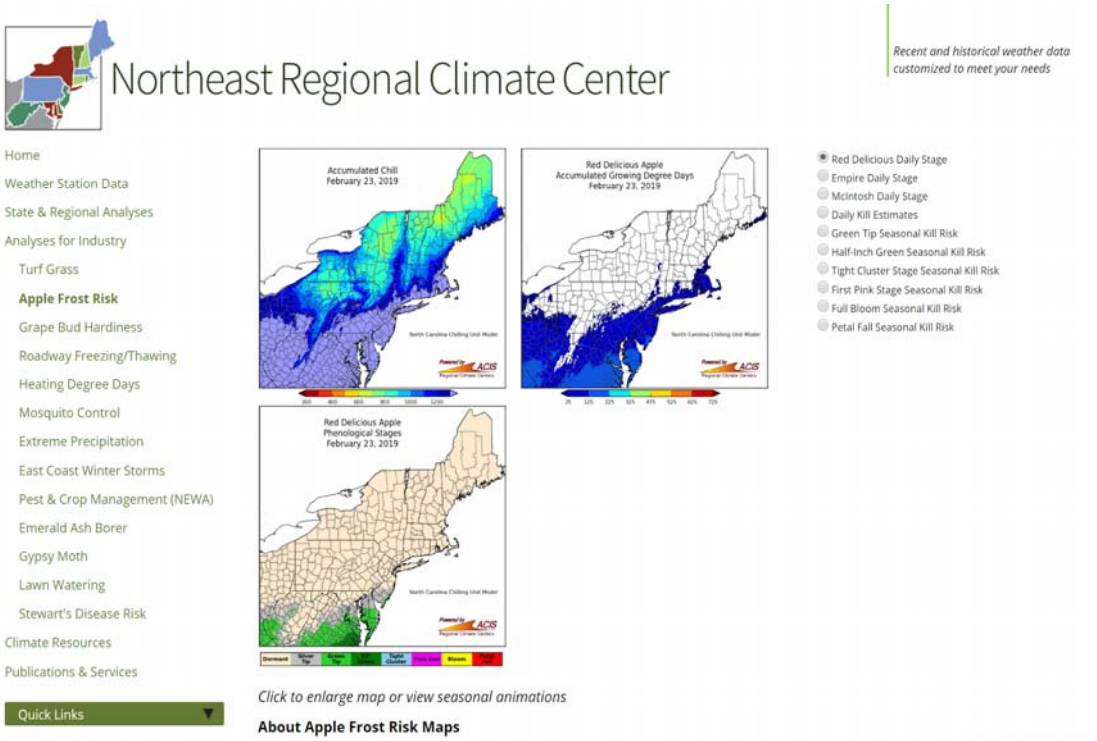
*"In this study, **differences [between DM and Utah model] were not found** between these two models when estimating the chilling requirements for seven sweet cherry cultivars in north-western Murcia", Alburquerque et al. (2008), for cherry trees in Spain*

*"We have obtained **very homogeneous results with the Utah and Dynamic models**[...] The chilling requirements of the evaluated cultivars in the 3 years studied were quite homogeneous, according to the Utah and Dynamic models. Besides, the relationship between the two models was very close (R= 0.99).", Ruiz et al. (2007), for apricots in Spain. This team is also using Utah model for prune tree in Spain (Ruiz et al., 2015).*

These results take part of the Luedeling et al. (2012) review; see Table 2, where for two studies in Spain DM appears with better performance, and for other two cases Utah model and DM appear with equivalent performance.

Therefore, in our view, although under for some cases DM has shown a better performance, we cannot conclude that DM is a better model for Spain in general terms.

This is also supported by other researchers using other models besides the DM in recent papers, as for instance (Darbyshare et al, 2013, made a study to evaluate the global warming on winter chilling in Australia using 0-7.2ºC, Modified Utah, and Dynamic Model; Miranda et al., 2013 compares Weinberger, Utah, North Carolina, Low Chilling and Dynamic for peach; Aybar et al., 2015, using a de Melo-Abreu model for analysing the suitability olives varieties in Argentina; Marra et al., 2017, using Richardson model and Chilling hours model, but no DM, for cherry in Italy; Sawamura et al., 2017, investigated the chilling requirements of peach in Japan using the Weinberger and Utah model). Also, the North Carolina model is currently being used by the Northeast Regional Climate Center from USA administration, implemented by the University of Cornell for apple tree (see below).



Source: *http://www.nrcc.cornell.edu/industry/apple/apple.html*

Therefore, we respectfully disagree with the referee's assessment of North Carolina model being obsolete as it is being currently used. Even the Chilling Hours Model, which is the oldest method to estimate winter chill accumulation (not considered in our paper), and considers all hours with temperatures ranging from 0 to 7.2 ºC equally effective, continues to be useful, as is still widely used in climate change impact and adaptation studies (see for example grapevine studies as Londo et al., 2014; Houston et al., 2018).

Additionally, if it was the case that DM is superior for our particular case, it would be important to notice that even with climate models, one could argue that some are non

adequate to reproduce specific (and also very important) climate aspects (e.g., the monsoon), but they are used anyway. We agree with the reviewer where he says (Luedeling et al, 2012): "All the models still leave a lot to be desired in terms of accuracy and some dormancy breaking behaviour at warm sites could not be explain at all".

Finally, we think that the main point is here is that all these models were developed, more than for specific locations, for specific tree species (peach for RbM). And the current practice is two or three models of chilling accumulation being used against phenological data of a specific species, generally with several varieties, and obviating that the model was fitted for a different crop (peach), assuming that there are not differences among species. In the few works where chilling accumulation models have been fitted for a different species than peach, differences respect to the fit for peach appeared. In our work, we have prioritized the adjustment parameters made to the RbM for different species (apple, which became North Carolina model; and olive, which became De Melo-Abreu model), under the hypothesis that the model would perform better if fitted to the behaviour of that species than if the model was is used with the parameters established for peach. In the case of the peach tree, the initial parameters of MbR and DM have been used.

Alburquerque, N., García-Montiel, F., Carrillo, A., and Burgos, L.: Chilling and heat requirements of sweet cherry cultivars and the relationship between altitude and the probability of satisfying the chill requirements, Environ. Exp. Bot., 64, 162-170, https://doi.org/10.1016/j.envexpbot.2008.01.003, 2008.

Aybar, V. E., Melo-Abreu, J. P., Searles, P. S., Matias, A. C., Del Rio, C., Caballero, J. M., and Rousseaux, M. C.: Evaluation of olive flowering at low latitude sites in Argentina using a chilling requirement model, Span. J. Agric. Res., 13, 10, 10.5424/sjar/2015131-6375, 2015.

Darbyshire, R., Webb, L., Goodwin, I., and Barlow, E. W. R.: Impact of future warming on winter chilling in Australia, International Journal of Biometeorology, 57, 355-366, 10.1007/s00484-012-0558-2, 2013.

Houston, L., Capalbo, S., Seavert, C., Dalton, M., Bryla, D., and Sagili, R.: Specialty fruit production in the Pacific Northwest: adaptation strategies for a changing climate, Clim. Change, 146, 159-171, 10.1007/s10584-017-1951-y, 2018.

Londo, J. P., and Johnson, L. M.: Variation in the chilling requirement and budburst rate of wild Vitis species, Environ. Exp. Bot., 106, 138-147, https://doi.org/10.1016/j.envexpbot.2013.12.012, 2014.

Luedeling, E.: Climate change impacts on winter chill for temperate fruit and nut production: A review, Scientia Horticulturae, 144, 218-229, https://doi.org/10.1016/j.scienta.2012.07.011, 2012.

Marra, F., Bassi, G., Gaeta, L., Giovannini, D., Palasciano, M., Sirri, S., and Caruso, T.: Use of phenoclimatic models to estimate the chill and heat requirements of four sweet cherry cultivars in Italy, Acta Hortic., 1162, 57-64, 10.17660/ActaHortic.2017.1162.10, 2017.

Miranda, C., Santesteban, L. G., and Royo, J. B.: Evaluation and fitting of models for determining peach phenological stages at a regional scale, Agric. For. Meteorol., 178-179, 129-139, https://doi.org/10.1016/j.agrformet.2013.04.016, 2013.

Ruiz, D., Campoy, J. A., and Egea, J.: Chilling and heat requirements of apricot cultivars for flowering, Environ. Exp. Bot., 61, 254-263, https://doi.org/10.1016/j.envexpbot.2007.06.008, 2007.

Ruiz, D., Egea, J., Salazar, J. A., and Campoy, J. A.: Necesidades de frío para la salida del letargo y necesidades de calor para florecer en variedades de ciruelo japonés (Prunus salicinia L.), XIV Congreso Nacional de Ciencias Hortícolas. SECH 2015. Retos de la Nueva Agricultura Mediterránea, Orihuela, Spain, 2015.

Sawamura, Y., Suesada, Y., Sugiura, T., and Yaegaki, H.: Chilling Requirements and Blooming Dates of Leading Peach Cultivars and a Promising Early Maturing Peach Selection, Momo Tsukuba 127, The Horticulture Journal, 86, 426-436, 10.2503/hortj.OKD-052, 2017.

3) Related to the previous points, we've done several studies to compare the response of various chill metrics to climate change. First, they differ greatly in their sensitivity to warming (Luedeling et al., 2009c). Second, they are not comparable, with the ratio between different chill metrics varying tremendously across the globe, especially along climate gradients (Luedeling and Brown, 2011). Especially at the warmest end of the climatic range for temperate fruit trees, most models fail (Balandier et al., 1993; Benmoussa et al., 2017a, 2017b; Linsley-Noakes and Allan, 1994). The Dynamic Model is the only model I know that has a chance of somewhat describing changes correctly across different climates. This is the reason why in our 2011 paper (Luedeling et al., 2011) we only report Chill Portions (we actually calculated other metrics too, if I remember correctly, but I consider the results meaningless). This reasoning is actually described in several places in this paper and elsewhere (e.g. Luedeling, 2012). Just as an illustration, in the literature we found the chilling requirement of 'Ohadi' pistachios quantified at 1000+ CH in Turkey, but they grow well at 100 CH in Tunisia. This difference is not trivial at all and illustrates how badly off we can be if we use the wrong model.

- *With regard the comparison of various chill metrics:*

We will introduce the reference Luedeling et al. (2009) as previous work on the comparison of response of various chill metrics to climate change. At the same time, we will stress that this is not the objective of this paper and review the manuscript removing explicit and implicit comparisons between models. In fact, we have not

averaged results from different chilling models, keeping results separately, as explained in Answer#2. We will stress it more in the paper, specifically in the Material and Methods section.

- *With regard the performance of different models:*

For a general answer, please see Answer#2. With regard models' performance in warm regions particularly, a worst performance is found not only for RbMs, but also for DM (Benmoussa et al., 2017 for pistachio in warm Sfax region in Tunisia):
https://www.sciencedirect.com/science/article/pii/S0098847217301119
***"highlight: The Dynamic Model does not work well under Tunisian climate conditions."** This supports our argument that the performance of these models is not so different. Due to the lack of knowledge and data (especially for chilling portions) for accurate model calibration, including warm regions, we believe that uncertainty is better handled if not just one model is considered, even if they are not directly comparable.*

Benmoussa, H., Luedeling, E., Ghrab, M., Ben Yahmed, J., and Ben Mimoun, M.: Performance of pistachio (Pistacia vera L.) in warming Mediterranean orchards, Environ. Exp. Bot., 140, 76-85, https://doi.org/10.1016/j.envexpbot.2017.05.007, 2017.

Luedeling, E., Zhang, M., Luedeling, V., and Girvetz, E. H.: Sensitivity of winter chill models for fruit and nut trees to climatic changes expected in California's Central Valley, Agric. Ecosyst. Environ., 133, 23-31, 10.1016/j.agee.2009.04.016, 2009.

4) One particular criticism of chill models has been that they are calibrated for a particular site and not necessarily generally valid. There is a reason why the North Carolina Model and the Utah Model are named after geographic areas, not after crops, and why researchers in various places saw the need to make adjustments. For example, in South Africa the Utah Model regularly produced negative chill totals at the end of the season. This was 'addressed' by removing the chill negation (resulting in the Positive Utah Model: Linsley-Noakes and Allan, 1994). The necessity of these 'empirical hacks' clearly indicates that these models can't be trusted across climatic gradients – which is critically important for a credible climate change assessment.

We agree with the referees 1 and 2 that, ideally, a site specific calibration would be desirable for any simulation exercise, as is the general practice in agronomic studies. As the second reviewer points out, indeed, the conditions of a calibrated model at one site do not completely coincide with those found in other locations. However, the state of the art of chill modelling is not yet there, and current practice is to apply these models elsewhere (see for instance many previous studies using these models in locations other than Utah, all of them without site-specific calibration, e.g. Alburquerque et al., (2008) for cherries in Spain; Razavi et al., (2011) for peach and Apricot in Iran; Sawamura et al., (2017) for peach in Japan). We think that in our case this is justified because in the model all the parameters that the researchers believe have relevance in the process are included. In our case, the main driver is temperature regime; and

actually, in the case of North Carolina model for apples, the main production area is Northern Spain, with climatological characteristics (temperature) more similar to North Carolina than the Spanish average. Accordingly, we will delimit more the concrete area of the apple tree production in the introduction section.

However, as we have discussed previously in Answer#2, we think that the main point here is that all these models were developed, more than for specific locations, for specific tree species (peach for RbM). And the current practice is two or three models of chilling accumulation being used against phenological data of a specific species, generally with several varieties, and obviating that the model was fitted for a different crop (peach), assuming that there are not differences among species. In the few works where chilling accumulation models have been fitted for a different species than peach, differences respect to the fit for peach appeared. In our work, we have prioritized the adjustment parameters made to the RbM for different species (apple, which became North Carolina model; and olive, which became De Melo-Abreu model), under the hypothesis that the model would perform better if fitted to the behaviour of that species than if the model was used with the parameters established for peach. In the case of the peach tree, the initial parameters of MbR and DM have been used.

This is a research gap indeed. As stated in Luedeling et al. (2011), estimates in Chill Portions (for the Dynamic model) are less widely available than estimates in other metrics, and although if the knowledge gap in that sense have been reduced nowadays, estimates for many crops and varieties are still not available. We agree that, ideally, more experimental data should be generated to improve the chilling simulation, not only because of the differences between locations, but mainly due to the huge uncertainty related to the species and variety requirements, that in our view, is much more important than that related to the models. We agree that is a scientifically relevant issue, so we will include a comment on this on the discussion to raise awareness on the referee's point.

Alburquerque, N., García-Montiel, F., Carrillo, A., and Burgos, L.: Chilling and heat requirements of sweet cherry cultivars and the relationship between altitude and the probability of satisfying the chill requirements, Environ. Exp. Bot., 64, 162-170, https://doi.org/10.1016/j.envexpbot.2008.01.003, 2008.

Luedeling, E., Girvetz, E. H., Semenov, M. A., and Brown, P. H.: Climate Change Affects Winter Chill for Temperate Fruit and Nut Trees, PLOS ONE, 6, e20155, 10.1371/journal.pone.0020155, 2011.

Razavi, F., Hajilou, J., Tabatabaei, S., and Dadpour, M.: Comparison of Chilling and Heat Requirement in Some Peach and Apricot Cultivars, Research in Plant Biology, 1, 40-47, -, 2011.

Sawamura, Y., Suesada, Y., Sugiura, T., and Yaegaki, H.: Chilling Requirements and Blooming Dates of Leading Peach Cultivars and a Promising Early

Maturing Peach Selection, Momo Tsukuba 127, The Horticulture Journal, 86, 426-436, 10.2503/hortj.OKD-052, 2017.

5) The presumably innovative outlook of possibly using estimates of the amount of chill that is exceeded 90% of the time (p. 10, l. 29) isn't so innovative after all. In fact, we already used this 'Safe Winter Chill' approach in several publications, dating back to 2009 (Luedeling et al., 2009a, 2011). It has also been picked up by others (though I don't currently remember who that was).

The novelty was referred to the EOA index application (see Rodríguez et al., 2019) to analyse the robustness of projections of having a safe winter chill. In other words, it refers to the robustness metric (the EOA index) application, not to the safe winter chill definition, which is used only as the hypothesis for the EOA index. We will reformulate the sentence in the further work paragraph to make it clearer. Also, we will add a quotation (Luedeling et al., 2009) wherever in the manuscript that reference to safe winter is done.

Luedeling, E., Zhang, M., and Girvetz, E. H.: Climatic Changes Lead to Declining Winter Chill for Fruit and Nut Trees in California during 1950–2099, PLOS ONE, 4, e6166, 10.1371/journal.pone.0006166, 2009.

Rodríguez, A., Ruiz-Ramos, M., Palosuo, T., Carter, T. R., Fronzek, S., Lorite, I. J., Ferrise, R., Pirttioja, N., Bindi, M., Baranowski, P., Buis, S., Cammarano, D., Chen, Y., Dumont, B., Ewert, F., Gaiser, T., Hlavinka, P., Hoffmann, H., Höhn, J. G., Jurecka, F., Kersebaum, K. C., Krzyszczak, J., Lana, M., Mechiche-Alami, A., Minet, J., Montesino, M., Nendel, C., Porter, J. R., Ruget, F., Semenov, M. A., Steinmetz, Z., Stratonovitch, P., Supit, I., Tao, F., Trnka, M., de Wit, A., and Rötter, R. P.: Implications of crop model ensemble size and composition for estimates of adaptation effects and agreement of recommendations, Agric. For. Meteorol., 264, 351-362, https://doi.org/10.1016/j.agrformet.2018.09.018, 2019.

6) Another alleged innovation is the variable duration of the chilling period, which is determined by the minimum and maximum chill accumulation. Sure, this is new, but is it correct? The authors don't present any evidence for this. I realize that some authors have claimed that something like this makes sense (e.g. Cesaraccio et al., 2004 for their own model, but others have also said this for the Utah Model I think), but is there really any evidence? Actually, I strongly doubt that trees can make use of chill accumulation over the entire cold period. We've done a number of studies where we tried to statistically determine the chill-responsive period (Guo et al., 2015; Luedeling and Gassner, 2012; Luedeling et al., 2013a, 2013b), and we've always found periods that are much shorter than the full winter season. Now this may mean various things, including that trees are pretty safe from chill shortfalls in many places, but I suspect that it would make sense to end the chilling period earlier than an automatic algorithm

would suggest (actually, if I could change one thing about our earlier studies, I would shorten the period we considered, which seems much too long now in hindsight).

The referee's discussion and the studies quoted (Guo et al., 2015; Luedeling and Gassner, 2012; Luedeling et al., 2013a, 2013b), in our view, reflect that there is a lot to learn about how trees work in relation to chilling accumulation. We agree that it is reasonable to question if trees can make use of the whole chilling accumulation period, and we will comment this fact about the possibility of an overestimation of the chilling accumulation in the discussion.

At the same time, we have decided not to choose a fixed period approach. On the one hand, a fixed starting date and duration for the chilling period for sure will introduce errors, as every year is different for every location and for every climate model. Some studies use self-regulating dates (we will quote them) for chill models because of the lack of reliable physiological markers and the inefficacy of fixed dates to account for the mentioned seasonal climate variability (Measham et al. 2017). For instance, Marra et al. (2017), where an approach to calculate the starting date, using a self-regulating algorithm similar than in the present study, found that the applied method allowed a significant improvement compared to other studies that fix the date at October 1st. Also, results in the Measham et al. (2017) study show a larger variability in the chilling portions accumulation using a fix dates approach than a self-regulating one, as some chilling portions were excluded due to a late initial date. On the other hand, we have decided not to select a fix final date, even when it could be very well defined, because it will become eventually meaningless in a climate change context. A fixed period would cause a lot of problems and inconsistencies when the cold period is clearly shifted along the year at the end of the century.

Other argument that supports the use of a self-regulating method is that changes in chilling projections become very much comparable among different methods and with the present, even when having in mind the possible overestimation mentioned by the referee.

Marra, F., Bassi, G., Gaeta, L., Giovannini, D., Palasciano, M., Sirri, S., and Caruso, T.: Use of phenoclimatic models to estimate the chill and heat requirements of four sweet cherry cultivars in Italy, Acta Hortic., 1162, 57-64, 10.17660/ActaHortic.2017.1162.10, 2017.

Measham, P. F., Darbyshire, R., Turpin, S. R., and Murphy-White, S.: Complexity in chill calculations: A case study in cherries, Scientia Horticulturae, 216, 134-140, https://doi.org/10.1016/j.scienta.2017.01.006, 2017.

7) The paper starts with a strange introduction about the classification of fruit trees, which I'm not sure I agree with and which is also not relevant here. This paper is only about temperate species, so no need for such a general take. The first two paragraphs should be deleted.

Our attempt was to take into account that this journal serves to a wide and diverse community of readers (as stated in the NHESS journal aims and scope) with this general introduction. However, we will reduce and focus it following referee's suggestion. First two paragraphs will be removed.

8) I strongly urge the authors to make their code public, either in a repository or as supplementary materials to this paper. This will make it much easier to understand what was done. For instance, the statement that the authors used the method by Fishman et al. (1987a, 1987b) is not sufficiently detailed – anyone who's seen these papers knows that this is not at all trivial to implement (and I wonder if this is really the authors' source of the algorithm). Ideally, a paper should be reproducible, meaning that the methods should be sufficiently detailed for readers to repeat an experiment. This is often not really achievable, but it is not difficult for a modeling study such as the one described here. Please share the code. The main reason for this is that the actual results of this paper are not particularly helpful – pretty much the same has been shown before. The innovation (for the chill modeling community) lies in the climate data processing, but if this isn't actually shared with readers, nobody can easily make use of this methodology. In my view, the offer that readers can contact the authors isn't sufficient.

All the algorithms used in this paper have been programmed, implemented and executed by the authors. In our team we have experts from different fields, being a computer engineer one of them. The implementation of the model was done by using the model constants commonly used in standard applications, following other studies like Luedeling et al., (2011). We will include the reference in that sense as it has been followed the same procedure.

We chose to share the code by the formula "under request and quotation", that means a simple e-mail message of request without further registration, as our institution recommends to do so, to keep track of the research groups and publications derived using it. This is the case of many software developments (e.g. DSSAT source code available upon request).

However, as both referees raised this point, we are willing to include the code as supplementary material. Specifically, we will include: chilling model codes, hourly temperature calculation and chilling computation period for the RbM models.

Luedeling, E., and Brown, P.: A global analysis of the comparability of winter chill models for fruit and nut trees, 411-421 pp., 2011.

9) Finally, I suggest that the authors compare their results (and maybe also their methods) with similar studies that have been done before. There have been quite a

few, as the authors will realize if they do a systematic search, not necessarily on Spain, but on various other regions.

We will compare our results with the references included in the Answers#1, 2, 3, 4 and 6, and with Luedeling et al. (2009a and 2009b) for California and Darbyshire et al. (2013) for Australia, which are particularly interesting for us because they were conducted in regions with Mediterranean climate. This will be done in the discussion section.

Darbyshire, R., Webb, L., Goodwin, I., and Barlow, E. W. R.: Impact of future warming on winter chilling in Australia, International Journal of Biometeorology, 57, 355-366, 10.1007/s00484-012-0558-2, 2013.

Luedeling, E., Zhang, M., and Girvetz, E. H.: Climatic Changes Lead to Declining Winter Chill for Fruit and Nut Trees in California during 1950–2099, PLOS ONE, 4, e6166, 10.1371/journal.pone.0006166, 2009a.

Luedeling, E., Zhang, M., Luedeling, V., and Girvetz, E. H.: Sensitivity of winter chill models for fruit and nut trees to climatic changes expected in California's Central Valley, Agric. Ecosyst. Environ., 133, 23-31, 10.1016/j.agee.2009.04.016, 2009b.

10) Even more finally, I suggest language editing. There is still some room for improvement in terms of language, and some statements are unclear.

The manuscript was edited by a professional language service previous to submission (the invoice will be privately sent to the editor due to data protection). The same service will be used on the revised manuscript.

Minor issues:

p. 1, l. 14: what are 'inner physiological factors'?

Lang et al., (1987) defined endodormancy as that which is regulated by physiological factors inside the affected structure. It is a definition widely used. We will include the definition instead the expression 'inner physiological factors'.

Lang, G. A., Early, J. D., Martin, G. C., and Darnell, R. L.: Endo-, para-, and ecodormancy: physiological terminology and classification for dormancy research, HortScience, 22, 371-377, 1987.

p. 1, l. 14: 'accumulating cool temperatures to finish dormancy is unclear (at least in terms of what dormancy this is – I'd most likely associate finishing dormancy with bloom of leaf out, but that also requires heat). No need for "be broken" in quotation

marks. This is commonly used and doesn't need to be identified as an odd term (or whatever the purpose of the quotation marks is).

Quotation marks will be removed, and the sentence will be reformulated as follows: "accumulating chilling temperatures to finish this sort of dormancy".

p. 1, l. 16: I don't think the chilling requirement is different for each variety (which means that no two varieties have the same requirement). They are crop and variety-specific, but not all different.

Yes, the referee is right. We will modify the sentence to avoid this possible misunderstanding, as follows:
"chilling accumulation required to break dormancy depends on specie and variety"

p. 1, l. 28 – p.2, l. 10: irrelevant – delete
We will delete the sentence.

p. 2, l. 12: income, not wealth
Yes, the referee is right. We will modify the sentence as suggested.

p. 2, several places: for simplicity and reader-friendliness, I recommend replacing 10^6 by 'million'
We will modify the sentence as suggested.

p. 2, ll. 18-19: FAOSTAT doesn't directly provide such values I believe, so it would be important to state how this was determined (also note that there are all kinds of issues with this database). It is also not obvious that this sentence refers to the global scale, since the previous sentence talks about Spain. Overall, this isn't a very relevant statement in a paper that's just on Spain.

In the FAOSTAT /Data/Crops webpage, it is possible to select a crop and gather worldwide data for a particular crop. According to those data, Spain is a major fruit producer in the world and, consequently, studies on Spain are relevant. We will briefly mention the process we followed to obtain the showed information from FAOSTAT service in the text.

p. 2, l. 24: I believe the thing trees are sensitive to is frost (not generally cold temperatures)

Yes, the referee is right. We will modify the sentence as suggested.

p. 3, l. 1: 'accumulation of cold periods' is an unfortunate choice of words. Sounds like trees need, say, 5 cold periods to break endodormancy.

Yes, the referee is right. We will reformulate it for making it clear as follows:

"the accumulation of time exposed to cold temperatures"

p. 3, l. 3: not all models are based on temperatures between certain thresholds. The Dynamic Model works differently, and even the Utah-type models don't really follow this simple structure.

Yes, the referee is right. We will modify the sentence by removing
"all based on the accumulated time with temperatures between certain thresholds".

p. 3, 12: I disagree that the chilling requirement corresponds to conditions where a tree is grown. It may rather correspond to conditions where it evolved/was bred

Yes, the referee is right. We will modify the sentence as follows:

"Each tree species and variety has specific chilling requirements for correct plant development, usually related to the environmental conditions where it evolved or was bred".

p. 3, ll. 13-17: not sure what information is conveyed here. The initial statement is about considering a range, but then the examples are precise values, not ranges. If this is supposed to illustrate intra-specific variation, then please make sure to use the appropriate terminology (not sure what 'crop tree' refers to).

Yes, we understand the referee's point. We will modify the text as follows:
"As a result, for a given species a range of estimates of chill accumulation encompassing all varieties has to be considered. For instance, for the apricot varieties considered in Campoy et al. (2012), the estimated accumulated chilling varies between 413 ('Palsteyn' variety) and 1172 ('Orange red' variety) chill hours (chilling hours method). This range is 613-777 when chilling units by Utah method are computed, and 37-64 chill portions when Dynamic method is applied."
Also, we will replace the expression "crop tree" by "fruit tree" throughout the paper.

p. 4, l. 9 (or elsewhere): Somewhere the authors need to mention the various chill assessments that have already been done by a number of people in a wide range of places.

Yes, the referee is right. We will mention the references listed in the answers to major issues (above in this document) in several parts of the text.

p. 4, l. 17: no, the models do not need hourly Tmin and Tmax. They just need hourly temperature, which can be derived (if no other information available) from daily Tmin and Tmax.

Yes, the referee is right. We will modify the sentence as follows:

"The climate variable required by the chilling models used in this study is hourly temperature, which can be derived, when no other information is available, from minimum (Tmin) and maximum (Tmax) temperatures."

p. 4, l. 22: not sure what 'freely distributed' means. Open-access?

We have used the exact term used by the creators of the dataset (http://www.meteo.unican.es/datasets/spain02)
It means that you can download the data with the condition of quoting two references provided. We will clarify it in the text specifying that free downloading is possible.

p. 4, l. 24: is this really an observational dataset?

Yes, it is. The methodology for generating these databases is robust and widely known on climate modelling studies: direct observations are interpolated in a physically-based way to a regular grid to be usable for climate models' comparison purposes. For instance, E-OBS (Haylock et al., 2008) and CRU (Harris et al., 2014) databases were built using this methodology.

Also, please see the link in the previous comment, where the database is described. Also, the quote Herrera et al., 2016 title reads:
Herrera et. al. (2016) Update of the Spain02 Gridded ***Observational*** *Dataset for Euro-CORDEX evaluation: Assessing the Effect of the Interpolation Methodology. International Journal of Climatology, 36:900–908. DOI: 10.1002/joc.4391.*

We will add the link (http://www.meteo.unican.es/datasets/spain02) in the text.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34, 623-642, 10.1002/joc.3711, 2014.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P., and New, M.: A European daily high-resolution gridded dataset of surface temperature and precipitation, D20119 pp., 2008.

Herrera et. al. (2016) Update of the Spain02 Gridded Observational Dataset for Euro-CORDEX evaluation: Assessing the Effect of the Interpolation Methodology. International Journal of Climatology, 36:900–908. DOI: 10.1002/joc.4391.

p. 5, l. 15: more details are needed on the temperature generation, especially since the source will be hard to find for most readers. What mathematical functions were used for constructing daily curves? The common method in horticultural studies such as this one is a methodology by Linvill (1990), which is based on a sine curve during the day

and logarithmic cooling at night (implemented in the chillR package; Luedeling, 2018). I'd be quite curious to learn how de Wit's method compares with this, but the authors provide insufficient information about their approach.

Yes, we used de Wit's method. MATLAB code will be made available in the supplementary material as requested by the referee.

p. 6, ll. 11-13: The authors compute a mean and then a median. Later in the paper they argue that one should calculate a 10% quantile. Why didn't they do this here?

The objective of this paper as stated in page 4 line 10 is to assess the impact of climate change on temperate fruit tree chilling accumulation Spain. This general objective is better achieved by an averaged indicator, as median and mean. The suggestion of using the 10th quantile was only introduced in page 10, starting from line 20, proposals for further work, consisting in using the EOA index for analysing chances of robust, high confidence, local adaptation. This EOA index needs a threshold definition, for which we propose the 10th quantile (so we do not need nor suggest using it for other purpose than that). This is a refinement of previous assessment of average impact, but we consider this further work out of the scope of the current study.

We will modify the text to make this point clearer.

p. 6, ll. 16-17: As stated above, I'd prefer to have the code made publically available, for full transparency and usefulness.

Please see the Answer#8. Codes will be provided as supplementary material.

p. 6, l. 23: Is the full name of MAPE really 'mean percentage absolute error'? That would seem to lead to the acronym 'MPAE'

Yes, you are right, this is a typo. That line will be changed by "mean absolute percentage error". In other parts of the document (i.e. page 19, line 6) the order is correct.

p. 7, l. 19: 'similarly simulated' is awkward wording

It will be changed to "simulated in a similar way".

p. 7, ll. 23-27: All these models use different units, so they can't be compared (the fact that they're probably all called chill units doesn't make them equivalent). While it's obvious that the Dynamic Model can't be compared to the others (because values are much smaller than for the other metrics), the others are also not comparable!

Yes, we understand that the reader could interpret that the models with the same units could be comparable. We will modify the text to clarify these aspects.

p. 8, l. 27: scenarios were averaged in this study, but we also provided information for determining the impact of climate model and emissions scenario.

We will modify this sentence as follows:

"The chilling portion results are in agreement with the projections from Luedeling et al. (2011) in the Mediterranean region for different periods, where emission scenarios and global climate models were averaged (see Fig. 6 in Luedeling et al., 2011; information for determining the impact of climate model and emissions scenario was provided in that study)."

p. 9, ll. 1-2: As stated above, I don't consider it an asset to include outdated models in a study…

Please see Answer#2 to major issues above in this document.

p. 9, l. 22: not sure what 'discrete nature' means. And I also think that this may be an indication that these models are too sensitive for warm places.

We meant discrete as opposite as continuous. We think that the high values of CV are related to the low values of the chilling in absolute terms, which actually is in agreement with referee's suggestion: this might makes these models too sensitive for warm places. We will include this explanation in the discussion.

p. 9, ll. 26: this study didn't 'find' this, it just reported on it. Luedeling et al. (2011) sort of found this.

We will modify this as suggested, using the verb "report".

p. 10, ll. 4: Yes, it would be great to have more datasets, but we actually already have a lot. Rather than call for collecting more data, I'd call for better use of such data for model development and validation.

Probably the referee is right and it is more about data availability and access and less about data existence. At least in the case of Spain, although it is true that there are several works on the subject, there are species/varieties with little data availability and the models developed up to now have important shortcomings. We will specify that the scarcity mentioned in the paper is referred to the available data in Spain, as we rely on referee's knowledge about elsewhere.

p. 10, ll. 11-12: 'crop varieties depending on the RCP' is unfortunate wording. First, crop varieties don't depend on RCP. Second, RCPs are theoretical pathways that will not be followed precisely. Better to say something like 'depending on how rapidly GHG emissions can be reduced' or something like that.

Yes, we understand how the sentence could be misunderstood. We will modify it according referee's suggestion.

p. 10, l. 23: not sure what 'low-limit chill requirements' are

We meant the variety with the lower chilling requirement within a given species. We will use that expression to make the sentence more understandable.

p. 10, l. 29: as mentioned above, this is exactly what the Safe Winter Chill metric achieves.

Yes, we are referring to that, we will introduce a quote here (Luedeling et al., 2009)

p. 11, 2-4: It's obvious that RCP8.5 causes greater change, similar to the end vs. middle of century. Doesn't need to be mentioned or should clearly be marked as expected.

Text we will be modified as suggested.

p. 11, l. 6: why especially in warm regions? The impact depends not only on chill loss, but also on what is grown there and how much chill it needs.

The text we will be modified as follows:

"A winter chill reduction may threaten the viability of some crops and varieties, especially in some areas that already have a low number of chilling units and are cultivated with chilling demanding species, where their reduction may jeopardise the cultivation of some tree crops within the near future."

p. 11, ll. 17-18: confusing sentence.

The text will be modified as follows:

"Such an adaptation would benefit from mitigation, as adaptation is assumed to be more feasible for moderate warming scenarios."

Reference list: It would be so much easier to look through this, if all but the first row of a reference were indented.

The section we will be modified as suggested.

Maps: maps should have a coordinate system, north arrow, scale bar etc.

We will include the suggested information in the corresponding figures.

Fig. 1: I doubt that all the olive data are right. If so, some parts of Spain would be almost exclusively olives.

We have checked the data and they are correct. Source is the Spanish Ministry related to agriculture and official statistics. Jaen province (Andalusia) is the largest area of olive trees in the world. When travelling through it (simply from the highway) you can only see olive trees for kilometres (please see image below).

Source:https://www.google.es/maps/@37.6076977,-4.0473674,3a,60y,283h,73.24t/data=!3m6!1e1!3m4!1sTVFJSEzMRW_Jco1F645SpA!2e0!7i13312!8i6656)

Maps 3-7: very hard to compare changes, which is really the most important part of this paper, if the maps are scattered across various places.

We will rearrange the figures to bring map of change together.

Fig. 5: is the scale used for the change useful.

We will adapt the scale to the new figures, and we will try to make it useful.

In summary, I think this contribution has potential, since the way the climate data were processed is very robust. But the team should consider adding some chill modelling capacity to the study to make this more convincing. While chill seems like an easy application of a climate change projection framework (it's assumed to just depend on temperature after all), things are actually quite complicated due to the invisibility of chill induced changes, which has precluded the development of convincing models so far. In consequence, there are many models, and most of them are not suitable for studies across climates. If the authors manage to adequately consider this, this manuscript may become publishable.

Thank you for your thoughtful revision. We have addressed the issues summarized by the referee in the answers above. We are convinced that our arguments are correct and sound, but if the editor and both referees ask us to remove some of the chilling models considered, we would be willing to do so.

References:

Balandier, P., Bonhomme, M., Rageau, R., Capitan, F., and Parisot, E. (1993). Leaf bud endodormancy release in peach trees - evaluation of temperature models in temperate and tropical climates. Agricultural and Forest Meteorology 67, 95–113.

Benmoussa, H., Ghrab, M., Ben Mimoun, M., and Luedeling, E. (2017a). Chilling and heat requirements for local and foreign almond (Prunus dulcis Mill.) cultivars in a warm Mediterranean location based on 30 years of phenology records. Agricultural and Forest Meteorology 239, 34–46.

Benmoussa, H., Luedeling, E., Ghrab, M., Ben Yahmed, J., and Ben Mimoun, M. (2017b). Performance of pistachio (Pistacia vera L.) in warming Mediterranean orchards. Environmental and Experimental Botany 140, 76–85.

Benmoussa, H., Ben Mimoun, M., Ghrab, M., and Luedeling, E. (2018). Climate change threatens central Tunisian nut orchards. International Journal of Biometeorology 62, 2245–2255.

Cesaraccio, C., Spano, D., Snyder, R.L., and Duce, P. (2004). Chilling and forcing model to predict bud-burst of crop and forest species. Agricultural and Forest Meteorology 126, 1–13.

Fishman, S., Erez, A., and Couvillon, G.A. (1987a). The temperature dependence of dormancy breaking in plants – computer simulation of processes studied under controlled temperatures. Journal of Theoretical Biology 126, 309–321.

Fishman, S., Erez, A., and Couvillon, G.A. (1987b). The temperature dependence of dormancy breaking in plants - mathematical analysis of a two-step model involving a cooperative transition. Journal of Theoretical Biology 124, 473–483.

Guo, L., Dai, J., Wang, M., Xu, J., and Luedeling, E. (2015). Responses of spring phenology in temperate zone trees to climate warming: A case study of apricot flowering in China. Agricultural and Forest Meteorology 201, 1–7. Linsley-Noakes,

G.C., and Allan, P. (1994). Comparison of 2 models for the prediction of rest completion in peaches. Scientia Horticulturae 59, 107–113. L

invill, D.E. (1990). Calculating chilling hours and chill units from daily maximum and minimum temperature observations. HortScience 25, 14–16.

Luedeling, E. (2012). Climate change impacts on winter chill for temperate fruit and nut production: A review. Scientia Horticulturae 144, 218–229.

Luedeling, E. (2018). chillR: Statistical methods for phenology analysis in temperate fruit trees (R package).

Luedeling, E., and Brown, P.H. (2011). A global analysis of the comparability of winter chill models for fruit and nut trees. International Journal of Biometeorology 55, 411–421.

Luedeling, E., and Gassner, A. (2012). Partial Least Squares Regression for analyzing walnut phenology in California. Agricultural and Forest Meteorology 158–159, 43–52.

Luedeling, E., Zhang, M., and Girvetz, E.H. (2009a). Climatic Changes Lead to Declining Winter Chill for Fruit and Nut Trees in California during 1950–2099. PLoS ONE 4, e6166.

Luedeling, E., Zhang, M., McGranahan, G., and Leslie, C. (2009b). Validation of winter chill models using historic records of walnut phenology. Agricultural and Forest Meteorology 149, 1854–1864.

Luedeling, E., Zhang, M., Luedeling, V., and Girvetz, E.H. (2009c). Sensitivity of winter chill models for fruit and nut trees to climate change. Agriculture, Ecosystems & Environment 133, 23–31.

Luedeling, E., Girvetz, E.H., Semenov, M.A., and Brown, P.H. (2011). Climate Change Affects Winter Chill for Temperate Fruit and Nut Trees. PLoS ONE 6, e20155.

Luedeling, E., Guo, L., Dai, J., Leslie, C., and Blanke, M.M. (2013a). Differential responses of trees to temperature variation during the chilling and forcing phases. Agricultural and Forest Meteorology 181, 33–42.

Luedeling, E., Kunz, A., and Blanke, M.M. (2013b). Identification of chilling and heat requirements of cherry treesâˇAˇTa statistical approach. International Journal of Biometeorology 57, 679–689.

Ruiz, D., Campoy, J.A., and Egea, J. (2007). Chilling and heat requirements of apricot cultivars for flowering. Environmental and Experimental Botany 61, 254–263. Zhang, J., and Taylor, C. (2011). The Dynamic Model provides the best description of the chill process on “Sirora” pistachio trees in Australia. HortScience 46, 420–425.