

APPLICATION PAPER

A climatic random forest model of agricultural insurance loss for the Northwest United States

Erich Seamon^{1,*} , Paul E. Gessler², John T. Abatzoglou³, Philip W. Mote⁴ and Stephen S. Lee⁵

¹Institute for Modeling, Collaboration, and Innovation, University of Idaho, Moscow 83844, Idaho, USA

²Department of Forest, Rangeland, and Fire Sciences, University of Idaho, Moscow 83844, Idaho, USA

³Department of Management of Complex Systems, University of California-Merced, Merced 95343, California, USA

⁴The Graduate School, Oregon State University, Corvallis 97331, Oregon, USA

⁵Department of Mathematics and Statistical Science, University of Idaho, Moscow 83844, Idaho, USA

*Corresponding author. E-mail: erichs@uidaho.edu

Received: 18 October 2022; **Accepted:** 28 October 2022

Keywords: Agriculture; climate; drought; random forest; insurance

Abstract

We compared climatic relationships to insurance loss across the inland Pacific Northwest region of the United States, using a design matrix methodology, to identify optimum temporal windows for climate variables by county in relationship to wheat insurance loss due to drought. The results of our temporal window construction for water availability variables (precipitation, temperature, evapotranspiration, and the Palmer drought severity index [PDSI]) identified spatial patterns across the study area that aligned with regional climate patterns, particularly with regards to drought-prone counties of eastern Washington. Using these optimum time-lagged correlational relationships between insurance loss and individual climate variables, along with commodity pricing, we constructed a regression-based random forest model for insurance loss prediction and evaluation of climatic feature importance. Our cross-validated model results indicated that PDSI was the most important factor in predicting total seasonal wheat/drought insurance loss, with wheat pricing and potential evapotranspiration having noted contributions. Our overall regional model had a R^2 of 0.49, and a RMSE of \$30.8 million. Model performance typically underestimated annual losses, with moderate spatial variability in terms of performance between counties.

Impact Statement

This work examines the effects of drought on agricultural insurance loss.

1. Introduction

Climate change significantly adds to the challenges facing agriculture, such as ensuring food security and preserving the economic prosperity of a growing global population (Rosenzweig et al., 2001; Schlenker & Roberts, 2009; Manandhar et al., 2014; Fan et al., 2016; Deschênes and Greenstone, 2018). Of noted interest are the varied effects of climate on weather-related phenomena, such as drought, heat waves, floods, hurricanes, and extreme precipitation. Drought in particular affects a number of factors associated with cropping systems, including excessive temperatures, water availability, long-term soil moisture drawdown, and levels of evapotranspiration (Rosenzweig and Parry, 1994; Lobell et al., 2011).

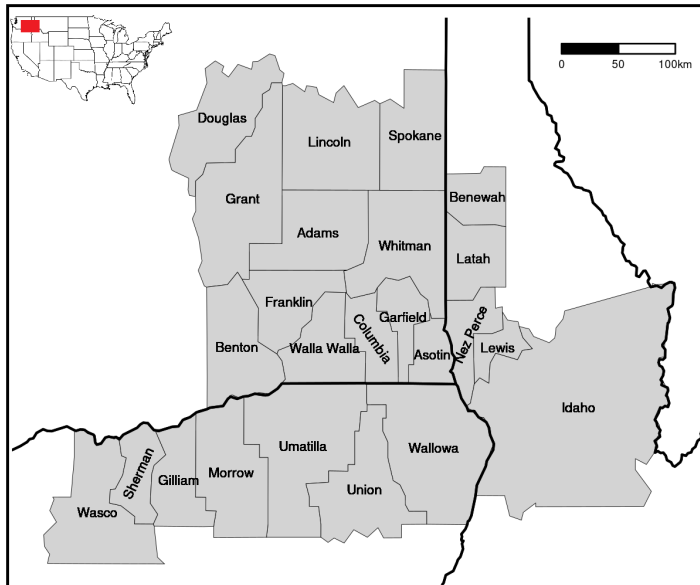


Figure 1. 24-county inland Pacific Northwest (iPNW) study area, which includes counties from Washington, Idaho, and Oregon.

This analysis focuses on the highly productive inland Pacific Northwest (iPNW) agricultural region of the United States (U.S.) (Figure 1), which relies heavily upon dryland farming for cereal production (Karimi et al., 2017; Yorgey and Kruger, 2017) and is considerably impacted by water availability. Our research objectives were to model how climatic effects for the iPNW are related to agricultural insurance loss, with a particular focus on drought-related claims for wheat. Two questions are addressed: What climate variables and temporal windows best relate to drought claims for wheat, and how do these relationships vary across the iPNW? Additionally, based on these optimum relationships, what climate variables have a greater influence on wheat insurance loss due to drought for the region, and can we utilize this framework for the prediction of insurance loss?

From a regional perspective, the iPNW produces around 17% of the U.S. wheat harvest (Karimi et al., 2017; Roesch-McNally, 2018). Given that cereal production is directly linked with variations in precipitation and temperature across much of the U.S. and the globe, iPNW wheat yields are significantly correlated with variability in plant available water during the growing season (Chi et al., 2017; Yorgey and Kruger, 2017). For example, 2015 iPNW wheat cropping outputs were negatively impacted by drought and extreme temperatures, evidenced in reduced crop yield outputs, increased agricultural insurance loss claim totals, and overall reduced wheat quality resulting in approximately \$200 million of insurance loss in Washington state alone (Sandison, 2015; Seamon et al., 2019). As one of the key agricultural production regions in the U.S., the iPNW supports a variety of cropping systems and management practices, which are dominated by dryland wheat farming (Chi et al., 2017). The region has cool-to-cold, wet winters and warm-to-hot, dry summers with considerable interannual variability based on regionalization across the Pacific Northwest (PNW) three state area (Abatzoglou et al., 2014; Yorgey and Kruger, 2017). Agricultural production for the area is typically limited by water rather than by growing season length (Stöckle et al., 2018), with annual precipitation increases from west to east, ranging from 200 mm to over 600 mm (Schillinger et al., 2010; Chi et al., 2017). Interannual climate variability in the region can lead to considerable variation in available moisture, temperature, and evaporative demand.

In terms of mitigating long-term and seasonal variability with regard to climatic impacts, crop insurance is a key mechanism that is used to reduce such risk (Miranda and Glauber, 1997; Seamon et al., 2019). Of particular note are long-term climatic variability impacts on short-term extreme weather

events, as well as shifts in seasonal/subseasonal weather outcomes that are exacerbated by a changing climate over an extended period of time (Backlund et al., 2008; USGCRP, 2017). In the PNW alone, from 2001 to 2015, over 35,000 insurance claims were filed, of which 20,600 were for wheat (Seamon et al., 2019). For the iPNW in particular, drought and heat insurance claims for all commodities resulted in approximately \$760 million in insurance losses from 2001 to 2015, which account for approximately 55% of all losses for this time period. Given this relationship between climate and agriculture, our research focus was to determine which climate variables and temporal windows best relate to drought claims for wheat, and what spatial variability exists with regard to this climate/agriculture association. Armed with this optimized information, we then developed a predictive model for estimating agricultural insurance loss based on climatic influences.

2. Data and Methods

Two datasets were used for this analysis. The first dataset was the U.S. Department of Agriculture (USDA) agricultural crop insurance claim archive from 2001 to 2015 (<http://rma.usda.gov>). When U.S. farmers experience economic losses for particular agricultural commodities, they typically file a crop insurance claim for that loss, as a part of the Federal Crop Insurance Corporation's program, which underwrites agricultural insurance policies in conjunction with private insurance organizations. The filing of a crop insurance claim is the result of a complex set of decision-making processes, where farmers may incorporate multiple factors, spanning biophysical, climatic, economic, and socio-demographic disciplines. Over time, these insurance claim records are systematically provided to the USDA, who administers the program via the USDA Risk Management Agency (RMA) and makes these data available as a public archive. For the purposes of this analysis, individual agricultural insurance claims for wheat due to drought were aggregated at the county level. We only included claims during January to September given our focus on winter wheat phenological cycles, and hereafter focused on annual claims for each county. The second dataset used was daily gridded climate data at 1/24th degree (4 km) spatial resolution for the four climate variables most associated with water availability (maximum temperature, precipitation, potential evapotranspiration, and the Palmer drought severity index [PDSI]) from Abatzoglou (2013). Daily values were summarized by county and to a monthly time scale, with totals aggregated for precipitation and potential evapotranspiration, and average values for maximum temperature and PDSI.

In order to explore the associations of climate to insurance loss, we constructed a set of time lagged climate variable design matrices, to search for the optimal seasonal temporal relationship between a climate variable (i.e., maximum temperature, precipitation, evapotranspiration, and PDSI) and a county's seasonal wheat insurance loss due to drought. We used this optimization approach to identify the most influential county-level time periods per season that were best correlated to wheat/drought insurance claims. Our initial step was to perform a general comparison of county-level wheat/drought-related claim acreage to water availability variables (precipitation and potential evapotranspiration), as well as aridity (precipitation/potential evapotranspiration), to examine general trends over time (2001–2015). The purpose of this initial analysis was to verify that expected patterns of wheat/drought insurance loss acreage versus water availability were seen (e.g., counties with relatively lower precipitation totals had overall increased wheat/drought insurance loss acreage). These annual, county-specific ratios were calculated by taking the total number of acres of wheat/drought insurance loss and dividing it by the total number of acres for all other wheat damage cause losses. As the ratio approaches 1, we identify county/year combinations where drought was the dominant factor in terms of total farmland acreage attributed to insurance loss. These analyses are provided as a part of our supplementary materials.

Second, we aggregated and associated climate data to wheat/drought crop loss seasonal totals by county, analyzing all outputs by using a time-lagged association by searching for the highest correlations between all combinations of monthly climate values (scaled and mean centered) and wheat insurance loss claims due to drought. We transformed annual loss amounts by a cube root function which resulted in a normal distribution of annual values. Following previous efforts to elucidate the optimum correlation (e.g., Du et al., 2013), we examined these climate/insurance loss correlations for each county, based on

independently and with the same distribution for all trees in the forest (Breiman, 2001). Random forest modeling reduces the potential for overfitting through the use of bootstrap aggregation, averaging across many trees, and provides a level of feature importance for assessing predictor power. As a part of our random forest construction, we utilized 10-fold cross validation, a model validation technique used to assess the generalizability of the model. Model construction for this analysis utilized the recursive partitioning and regression trees package (rpart) within R (Atkinson and Therneau, 2000; Breiman et al., 1984).

3. Results

The results of our initial county-level wheat/drought acreage ratios (2001–2015), as compared to precipitation, potential evapotranspiration, and aridity are shown in the supplementary materials. Each observation represents the mean annual totals of each individual variable, for each county, from 2001 to 2015. We see expected climate/insurance loss relationships, with county-level precipitation totals, as well as aridity, inversely proportional to increased drought acreage ratios. Similarly, potential evapotranspiration totals increased as ratios increased. These general patterns supported moving forward to explore more specific spatiotemporal relationships of climate to insurance loss using our climate-lagged correlation framework. Using these time-lagged matrix correlation outputs, we mapped the spatial and temporal variations for each climatic variable, at a county scale (Figure 3).

3.1. Potential evapotranspiration

An overview analysis of potential evapotranspiration (PET) from 2001 to 2015 for the study area indicates a climate gradient from east to west, with the annual maximum occurring in July. When examining optimum correlations of PET to cube root transformed wheat/drought insurance loss, the counties of Walla Walla, Columbia, and Whitman (Washington) had the highest values, along with Latah and

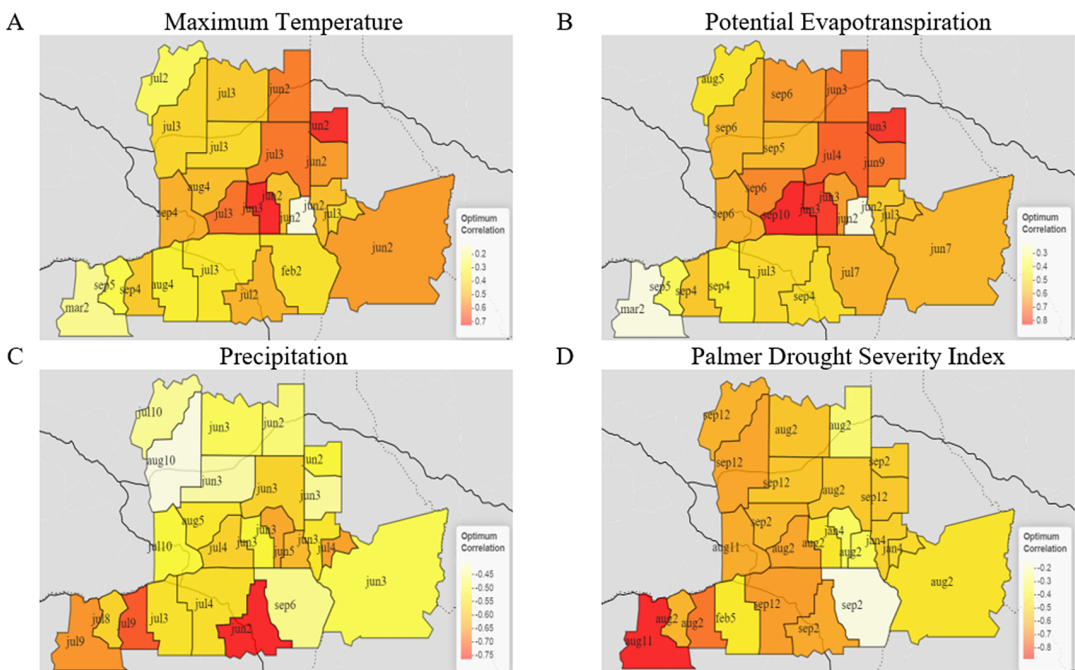


Figure 3. Spatial plot of correlation between insurance loss due to drought, for A) maximum temperature, B) potential evapotranspiration, C) precipitation and D) the Palmer drought severity index, across all counties within the study area, 2001–2015.

Benewah counties (Idaho). Southernmost counties (Oregon) tended to have lower correlations than northern study area counties.

3.2. Precipitation

Regional cycles of precipitation, averaged for the period from 2001 to 2015, show increasing values from northwest to southeast. In all counties, the wettest months are November through January. Negative correlations of precipitation with insurance loss were highest in the southernmost study area counties, which tended to border the Snake and Columbia rivers (Wasco, Sherman, and Union counties, OR, and Columbia and Whitman counties, WA), ranging from -0.61 to -0.81 . Counties further to the northwest (Franklin, Douglas, Lincoln, Spokane counties, WA) tended to have lower correlations, with April/May/June precipitation more highly correlated with drought loss for most counties.

3.3. Maximum temperature

Temperature tends to increase from southeast to northwest, which, as expected, is opposite to the pattern for precipitation. July was the peak month for high temperatures for every county in the iPNW. Maximum temperature was most highly correlated with drought in southeastern Washington counties, including Walla Walla, Whitman, Columbia, Garfield, and Spokane counties, WA, as well as Latah county, ID, ranging from 0.59 to 0.72. Temporally, most counties had an optimum relationship that fell between April and July, with Oregon counties and several western WA counties shifting that time window slightly later. The two counties with the highest correlations (Columbia and Garfield counties, WA), both had optimum time windows that included April.

3.4. Palmer drought severity index

PDSI spatial gradients had an increasing trend from southeast to northwest. PDSI optimum correlations with wheat/insurance loss appeared to have a spatial gradient that similarly increased to the northwest, with Douglas, Benton, and Grant counties, WA, as well as Wasco county, OR, having the highest correlation coefficients ($r = 0.84$). July/August/September was the most common window for optimum climate/insurance loss relationships, yet those counties with the highest correlations had window ranges that were much longer. As with precipitation, PDSI is negatively correlated with wheat/drought loss.

Overall results of our 24-county, 10-fold cross-validated random forest model using insurance dollar loss totals yielded a R^2 of 0.49 with a RMSE of approximately \$30,880,545 (Figure 4). Individualized county model results provided variable performance, with R^2 ranging from 0.10 to 0.94 (see the Supplementary Material for individualized county model outputs). Feature importance rankings indicated that PDSI was the most influential predictor, with wheat prices and potential evapotranspiration as the second and third most important. Precipitation was the lowest predictor in terms of feature importance. When evaluating model error across training size ($n = 348$) using a learning curve analysis, we see considerable variation between training and validation error, which indicates that the model may be overfitting, and would potentially benefit from a reduction of complexity, or an increase in observations to improve performance.

4. Conclusions

The results of our climate lagged correlation analysis provide an interesting view of the spatial and temporal relationships of climate with localized insurance loss, with a particular focus on the region's dryland wheat production. In particular, our results indicate the importance in understanding the varying temporal effects of drought-related climatic variables, as they vary spatially, which is also supported by the previous work of Semenov (2009) and Lobell et al. (2015).

The results of our random forest model efforts indicate the potential importance of economic factors not accounted for in the current model, particularly given the underprediction of years with extreme

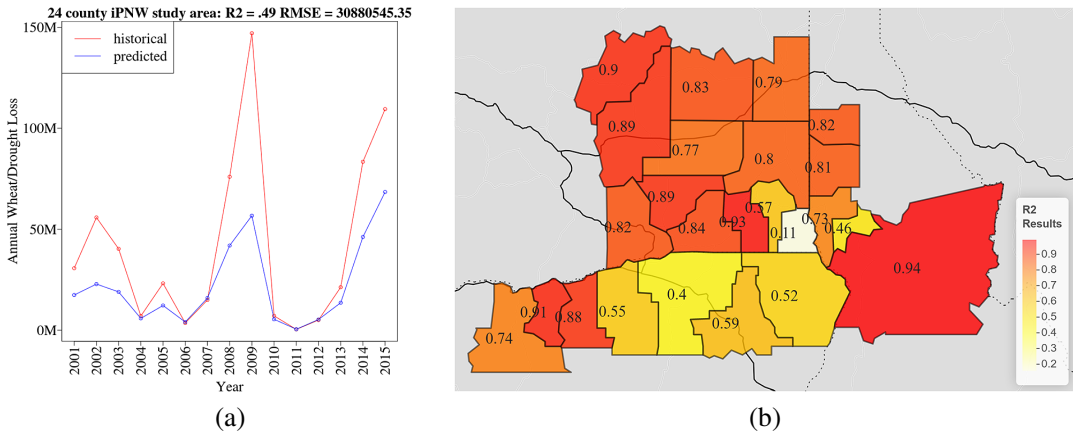


Figure 4. Historical versus predicted annual wheat insurance loss (\$) due to drought, constructed using a random forest model, for the 24-county iPNW study area. Input variables were precipitation, maximum temperature, and potential evapotranspiration, as well as annual wheat pricing, from 2001 to 2015. Climate variables were refined using the aforementioned time-lagged correlation methodology ($R^2 = 0.49$).

drought claim totals. Given the considerable declines in wheat prices from 2008 to 2009 during the U.S. great recession (Fan et al., 2016), the underprediction of wheat insurance loss totals in 2009 may indicate the difficulty in the model to factor in economic considerations given the possible inflation of drought loss due to economics rather than climatic outcomes. By contrast, in 2015, which was a year of noted severe drought in the iPNW (Mote et al., 2016), the model prediction was considerably closer to the observed loss. The comparisons of these two years may suggest a broader question around insurance loss and the interaction of climate and economics: are there economic thresholds that may produce climatic-associated loss claims, without clear justification in the climate data? While the developed model does attempt to broadly incorporate economic agricultural indicators (e.g., wheat pricing), the authors acknowledge that additional economic variables would potentially assist in addressing variance issues in loss prediction (consumer price indexing, unemployment rates, and housing/land values). Model performance may additionally be improved by using additional normalized climatic variables.

Given these associations of insurance loss with climate change, there are important considerations to contemplate. As future climatic conditions in the iPNW will likely lead to increased evapotranspiration rates, increased temperatures, and thus more extreme soil moisture deficits (Suyker and Verma, 2012), crop insurance programs will likely be negatively impacted. With limited long-term resilience, crop insurance efforts face added financial pressure under prolonged extreme weather conditions: subsidization mechanisms by the federal government to indemnify authorized crop insurance providers have typically only 1 year of coverage in cases of severe claim payouts (USDA RMA, 2016). As such, future extreme conditions for crop commodity systems, particularly around water availability, will likely increase the likelihood of financial stress with consecutive year drought events.

Author Contributions. Conceptualization: E.S., P.E.G., J.T.A.; Data curation: E.S.; Formal analysis: E.S.; Funding acquisition: P.W.M., P.E.G.; Investigation: E.S.; Methodology: E.S., P.E.G., J.T.A.; Project administration: E.S.; Resources: P.E.G., J.T.A.; Supervision: P.E.G., J.T.A., P.W.M., S.S.L.; Validation: E.S., J.T.A., S.S.L.; Visualization: E.S., J.T.A.; Writing—original draft: E.S.; Writing—review and editing: all authors.

Competing Interests. The authors declare no competing interests exist.

Data Availability Statement. A supplementary appendix as well as associated datasets can be found at <https://doi.org/10.5061/dryad.h9w0vt4kh>.

Ethics Statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Funding Statement. This study was supported by a grant from the National Oceanic and Atmospheric Administration under grant number NA15OAR4310145, as well as by the National Institute of General Medical Sciences of the National Institutes of Health under award number P20GM104420. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Oceanic and Atmospheric Administration or the National Institutes of Health.

Provenance. This article is part of the Climate Informatics 2022 proceedings and was accepted in *Environmental Data Science* on the basis of the Climate Informatics peer review process.

References

- Abatzoglou JT** (2013) Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology* 33(1), 121–131.
- Abatzoglou JT, Rupp DE and Mote PW** (2014) Seasonal climate variability and change in the Pacific Northwest of the United States. *Journal of Climate* 27(5), 2125–2142.
- Atkinson, E. J., & Therneau, T. M.** (2000). An Introduction to Recursive Partitioning Using the RPART Routines. *Mayo Foundation Technical Report* (Vol. 61, October).
- Backlund, P, Janetos, A, Schimel, D** (2008) *The effects of climate change on agriculture, land resources, water resources, and biodiversity in the United States*. Synthesis and Assessment Product 4.3. Washington, DC: U.S. Environmental Protection Agency, Climate Change Science Program. 240 p.
- Breiman L** (2001) Statistical modeling: The two cultures. *Statistical Science* 16(3), 199–231.
- Breiman L, Friedman J, Stone CJ and Olshen RA** (1984) *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor and Francis, Boca Raton, FL.
- Chi J, Waldo S, Pressley SN, Russell ES, O’Keeffe PT, Pan WL, Huggins DR, Stöckle CO, Brooks ES and Lamb BK** (2017) Effects of climatic conditions and management practices on agricultural carbon and water budgets in the inland Pacific Northwest USA. *Journal of Geophysical Research: Biogeosciences* 122(12), 3142–3160.
- Clark L and Pregibon D** (1992) Tree-based models. In Chambers JHT (ed), *Statistical Models in S*. Routledge/Taylor Francis, Boca Raton, FL.
- Deschênes BO and Greenstone M** (2018) American economic association the economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather: Reply author(s): Olivier Deschênes and Michael Greenstone source. *The American Economic Review* 102(7), 3761–3773.
- Du L, Tian Q, Yu T, Meng Q, Jancso T, Udvardy P and Huang Y** (2013) A comprehensive drought monitoring method integrating MODIS and TRMM data. *International Journal of Applied Earth Observation and Geoinformation* 23, 245–253.
- Fan M, Pena AA and Perloff JM** (2016) Effects of the great recession on the U.S. agricultural labor market. *American Journal of Agricultural Economics* 98(4), 1146–1157.
- Karimi T, Stöckle CO, Higgins SS, Nelson RL and Huggins D** (2017) Projected dryland cropping system shifts in the Pacific Northwest in response to climate change. *Frontiers in Ecology and Evolution* 5(April):1–9.
- Lobell DB, Hammer GL, Chenu K, Zheng B, Mclean G and Chapman SC** (2015) The shifting influence of drought and heat stress for crops in Northeast Australia. *Global Change Biology* 21(11), 4115–4127.
- Lobell DB, Schlenker W and Costa-Roberts J** (2011) Climate trends and global crop production since 1980. *Science* 333(6042), 616–620.
- Manandhar S, Pandey VP, Kazama F and Kazama S** (2014) Economics of climate change. In S Shrestha, S Babel and V Pandey (Eds). *Climate Change and Water Resources*. CRC Press, 153–182.
- Miranda MJ and Glauber JW** (1997) Systemic risk, reinsurance, and the failure of crop insurance markets. *American Journal of Agricultural Economics* 79(1), 206–215.
- Mote PW, Rupp DE, Li S, Sharp DJ, Otto F, Uhe PF, Xiao M, Lettenmaier DP, Cullen H and Allen MR** (2016). 2015 snowpack in the western United States. *Geophysical Research Letters* 43, 10,980–10,988.
- Prasad AM, Iverson LR and Liaw A** (2006) Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* 9(2), 181–199.
- Roesch-McNally G** (2018) U.S. inland Pacific Northwest wheat farmers’ perceived risks: Motivating intentions to adapt to climate change? *Environments* 5(4), 49.
- Rosenzweig C, Iglesias A, Yang X, Epstein PR and Chivian E** (2001) Climate change and extreme weather events: Implication for food production, plant diseases, and pests. *Global Change and Human Health* 2, 90–94.
- Rosenzweig C and Parry M** (1994) Potential impact of climate change on world food supply. *Nature* 367, 133–138.
- Sandison D** (2015) *2015 Drought and Agriculture*. Technical report, Washington State Department of Agriculture. Olympia, WA.
- Schillinger WF, Papendick RI and Meccool DK** (2010) Soil and water challenges for Pacific Northwest Agriculture. In T Zobeck and W Schillinger (Eds). *Soil and Water Conservation Advances in the United States*, SSSA, Madison, WI. pp. 47–80.
- Schlenker W and Roberts MJ** (2009) Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences* 106(37), 15594–15598.

- Seamon E, Gessler PE, Abatzoglou JT, Mote PW and Lee SS** (2019) *Agricultural Insurance Loss and Relationships to Climate across the Inland Pacific Northwest Region of the United States*. Dissertation, chapter 1, University of Idaho.
- Semenov MA** (2009) Impacts of climate change on wheat in England and Wales. *Journal of the Royal Society Interface* 6(33), 343–350.
- Stöckle CO, Higgins S, Nelson R, Abatzoglou J, Huggins D, Pan W, Karimi T, Antle J, Eigenbrode SD and Brooks E** (2018) Evaluating opportunities for an increased role of winter crops as adaptation to climate change in dryland cropping systems of the U.S. inland Pacific northwest. *Climatic Change* 146(1–2), 247–261.
- Suyker, A. E., & Verma, S. B.** (2012). Gross primary production and ecosystem respiration of irrigated and rainfed maize-soybean cropping systems over 8 years. *Agricultural and Forest Meteorology*, 165, 12–24. <https://doi.org/10.1016/j.agrformet.2012.05.021>
- United States Department of Agriculture** (2016) *Crop Insurance Program*. USDA Risk Management Agency. Washington D.C.
- USGCRP** (2017) *Climate Science Special Report: Fourth National Climate Assessment*, Volume I. Technical report, Washington, DC: U.S. Global Change Research Program.
- Verbyla D** (1987) *Classification trees: A new discrimination tool*. *Canadian Journal of Forest Resources*.
- Yorgey G and Kruger CE** (2017) *Advances in Dryland Farming in the Inland Pacific Northwest*. Extension mimeo. Washington State University Extension Publications. Pullman, WA.