Responses to Referees:

Response to Anonymous Referee # 1 (npgd-1-c37-2014)
We would like to thank Anonymous Referee # 1 for the excellent comments. The primary point raised by this reviewer, "… an inclusion of a broader range of other groups working in the same area could greatly improve the impact of this perspective" is very well taken. The reviewer suggests specific papers, all of which will be cited in the revised version. In addition, we will cite a few other papers in related areas (e.g., predictability). We are grateful to this reviewer for not only suggesting the citations but also including a short discussion about these papers in the review itself. These and other specific comments (e.g., labels in sub-figures etc.) will all be addressed in the revised version.

Response to Anonymous Referee # 2 (npgd-1-c46-2014)
We appreciate the short yet meaningful comments by Anonymous Referee # 2. Our responses to the specific points raised by this referee are as follows:

- "This paper appears to be a review paper… ": This manuscript is intended as a perspectives paper. The first goal is to motivate an emerging and urgent area of research while providing an assessment of the literature in that or related areas. The second goal is a call to action for researchers and practitioners. Specifically, two interdisciplinary communities are targeted. The first is that sub-community of geoscientists, earth and environmental engineers, and experts in climate impacts, adaptation and vulnerability, who would want to adopt or apply innovative data science methods. This group includes, for example, climate or impacts researchers engaged in statistical downscaling, statistical evaluation of models, or analysis of observed and modeled trends and patterns. In addition, this paper attempts to target the primary readership of the journal Nonlinear Processes in Geophysics, specifically, who are interested in innovative data science methods and their applications in geophysics. We note that data science in this context includes computer science (e.g., data mining and machine learning), nonlinear dynamics (including network science), signal processing, and statistics. The second target group is that sub-community of data scientists who would be interested to bring their methods and tools to address societal challenges in climate science and impacts or adaptation. This group includes researchers in the emerging area of climate informatics.
- "…because it does not contain new previously unpublished information": This primary intent of the perspectives paper is not to overwhelm with previously unpublished materials. However, there is significant unpublished information in each of the examples, which are all designed to motivate the next steps in this important research topic. Thus, Figures 4 and 5 are unpublished, Figure 1 is an (unpublished) summary of the state of knowledge which also serves as a motivation, while Figures 2, 3 and 6 are new summaries of our prior publications designed to motivate and/or exemplify the research topics.
- "However, as a review paper, it lacks sufficient context to allow a non-expert to read and get a sense of the field": This perspectives paper is designed to have enough depth and context for the target audience as described above. The primary audience is either expected to have background in the data science aspects of climate and impacts, or data scientists with an interest in contributing to climate science and impacts. While this perspectives paper is targeted towards specific (albeit broad and interdisciplinary) research communities, we hope that sections of the paper will also appeal to other readers of this journal who may not be experts in the specific areas mentioned above.

Detailed Manuscript Changes

1. In Section 1 – Introduction - After "The differing insights are summarized by Trenberth et al. (2014) in a perspectives article in Nature Climate Change.", added the following paragraph: *"Similar opposing insights have been reported for temperature extremes, which are generally relatively better simulated by climate models. Hansen et al. (2012) reported that seasonal temperature anomalies have significantly increased while Huntingtonford et al. (2013) reported significantly more uncertainty and did not find an increasing trend. Apparent insights can depend on metrics of choice and data analysis procedures (Alexander and Perkins, 2013; Huntingford et al., 2013), adding complexity to the analytic process and interpretation of findings."*

   Added the following references for this addition:

   a. Alexander, L. and S. Perkins.: Debate heating up over changes in climate variability. Environ. Res. Lett. 8, doi:10.1088/1748-9326/8/4/041001, 2013.
   b. Hansen, J., Sato, M., and Ruedy, R.: Perception of climate change, Proc. Natl. Acad. Sci. USA, 109, E2415-E2423, doi:10.1073/pnas.1205276109, 2012.
   c. Huntingford, C., Jones, P. D., Livina, V. N., Lenton, T. M., and Cox, P. M.: No increase in global temperature variability despite changing regional patterns, Nature, 500, 327-330, doi:10.1038/nature12310, 2013.
2. In section 4 - Societal Urgency and State of the Science, changed "Even for the relatively better understood temperature extremes, such as heat waves and cold snaps, large uncertainties remain, especially at regional scales (Ganguly et al., 2009b)." to: "*Even for the relatively better understood temperature extremes, such as heat waves and cold snaps, large uncertainties remain, especially at regional scales (Ganguly et al., 2009b). Recent studies (Fischer et al., 2013; Fischer and Knutti, 2014) suggest that these large uncertainties will likely persist even if climate models improve rapidly (Maslin and Austin, 2012; Kumar et al., 2014).*"
   Added the following references for this addition:
   a. Kumar, D., Kodra, E., and Ganguly, A. R.: Regional and seasonal intercomparison of CMIP3 and CMIP5 climate model ensembles for temperature and precipitation, Clim. Dynam., doi:10.1007/s00382-014-2070-3, 2014.
   b. Fischer, E., Beyerle, U., and Knutti, R.: Robust spatially aggregated projections of climate extremes, Nat. Clim. Change, 3, 1033-1038, doi:10.1038/nclimate2051, 2013.
   c. Fischer, E., and Knutti, R.: Detection of spatially aggregated changes in temperature and precipitation extremes, Geophys. Res. Lett., 41(2), 547-554, doi:10.1002/2013GL058499, 2014.

3. In section 5 – Characterization of climate extremes, after: "While phenomena like heat waves under climate change are better understood than most other climate-related extremes (Coumou and Rahmstorf, 2012; Field et al., 2012), their very definitions may depend on the impact sector of interest (Ebi and Meehl, 2007).", added: *"Quantitative research relating climate extremes and anomalies to impacts, for example terrestrial ecology (Reichstein et al., 2013; Zscheischler et al., 2013; Zscheischler et al., 2014) and agricultural production (Lobell et al., 2006; Lobell et al., 2012), often examine climate indices derived from extremes with disciplinary specificity."*
   Added the following references for this addition:

    a. Lobell, D. B., Field, C. B., Cahill, K. H., and Bonfils, C.: Impacts of future climate change on California perennial crop yields: Model projections with climate and crop uncertainties, Ag. Forest Meteorol., 141(2-4), 208-218, doi:10.1016/j.agrformet.2006.10.006, 2006.
    b. Lobell, D. B., Sibley, A., and Ortiz-Monasterio, J. I.: Extreme heat effects on wheat senescence in India, Nat. Clim. Change, 2, 186-189, doi:10.1038/nclimate1356, 2012.
    c. Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D. C., Papale, D., Rammig, A., Smith, P., Thonicke, K., van der Velde, M., Vicca, S., Walz, A., and Wattenbach, M.: Climate extremes and the carbon cycle, Nature, 500, 287-295, doi:10.1038/nature12350, 2013.
    d. Zscheischler, J., Mahecha, M. D., Harmeling, S., and Reichstein, M.: Detection and attribution of large spatiotemporal extreme events in Earth observation data, Ecol. Inform., 15, 66-73, doi:10.1016/j.ecoinf.2013.03.004, 2013.
    e. Zscheischler, J., Mahecha, M. D., Harmeling, S., Rammig, A., Tomelleri, E., and Reichstein, M.: Extreme events in gross primary production: a characterization across continents, Biogeosciences Discuss., 11, 1869-1907, doi:10.5194/bgd-11-1869-2014, 2014.
4. In section 5 – Characterization of climate extremes, after: "The value-addition of the MRF-based approach, beyond proof-of-concept detection of known droughts, would be demonstrated when the methods are generalized for multiple variables, and subsequently used for the evaluation of historical multi-model ensembles as well as for the generation of future projections with uncertainty from model projections in forecast mode." Added: *"Computationally scalable and flexible detection approaches based on spatio-temporal similarity between drought events (Lloyd-Hughes et al., 2012) have also recently been developed."*
   Added the following reference for this addition:
    a. Lloyd-Hughes, B.: A spatio-temporal structure-based approach to drought characterisation, Int. J. Climatol., 32, 406-418, doi:10.1002/joc.2280, 2012.
5. In section 8 - Enhanced Understanding and Predictions, after "In the same manner, temperature and updraft velocity profiles have been used to constrain or enhance

multimodel projections of precipitation extremes (Knutson et al., 2010; Wilhite and Glantz, 1985).", added: *"Additionally, ensembles have been found to simulate robust statistics of severe thunderstorm environments and imply increased risk in possible convective hazards under global warming (Diffenbaugh et al., 2013)."*
Added the following reference for this addition:

   a. Diffenbaugh, N. S., Scherer, M., and Trapp, R.J.: Robust increases in severe thunderstorm environments in response to greenhouse forcing, Proc. Natl. Acad. Sci. USA, 110, 16361-16366, doi: 10.1073/pnas.1307758110, 2013.

6. In section 8 - Enhanced Understanding and Predictions, we added a Runge et al., 2014 reference to the following sentence: *"Network-based graphical models have been used to discover causality among different modes of climate variability (Ebert-Uphoff and Deng, 2012; Runge et al., 2014)."*
   Added the following reference for this addition:
   a. Runge, J., Petoukhov, V., and Kurths, J.: Quantifying the Strength and Delay of Climatic Interactions: The Ambiguities of Cross Correlation and a Novel Measure Based on Graphical Models, J. Clim., 27, 720-739, doi:10.1175/JCLI-D-13-00159.1, 2009.
7. In section 9 – Summary, changed the first sentence to *"One of the largest scientific gaps in climate change studies is the inability to develop credible projections of extremes with the degree of precision required for adaptation decisions and policy (Fischer et al., 2013).*" The addition of Fischer et al., 2013 reference was the only change made. We added this to the bibliography already.
8. In section 9 – Summary, after: "As climate-related data approaches the scale of hundreds of petabytes (Overpeck et al., 2011), and climate data mining research continues to improve (Smyth et al., 1999; Robertson et al., 2004, 2006; Khan et al., 2006; Camargo et al., 2007a, b; Gaffney et al., 2007), new opportunities willemerge (e.g., Monteleoni et al., 2013; Ganguly et al., 2013)." , add: *"The 2014 Climate Data Initiative (Lehmann 2014) launched by the White House (United States President's Office) points to Big Data as a solution for climate adaptation, and lends further urgency of the theme discussed in this manuscript. However, despite the promise, pitfalls in pure data mining methods have been pointed out in the context of climate. Thus, Caldwell et al. (2014) shows how naive applications of data mining may yield spurious relationships in climate. This paper emphasizes the need to intelligently combine physics understanding with data mining, not just to avoid the risk of generating misleading insights, but also to produce novel results that may not have been possible otherwise. Data-driven methods may be complementary to physics and may need to be constrained by physics (e.g., see Majda and Yuan, 2012; Majda and Hardin, 2013). When mining climate model simulations, data mining is conditioned on the embedded physics in the models, and aspires to extract relations that may further inform and augment our current physical understanding. However, to be successful, data mining methods need to be aware of the complexity of climate processes*

and data." Removed the following sentence: *"Indeed, data-driven methods are complementary to, and indeed conditioned on, physics-based models or constrained by physics (e.g., see Majda and Yuan, 2012; Majda and Hardin, 2013); however, they need to be tailored to the complexity of climate data and processes."*

Added the following reference for this addition and change:

a. Caldwell, P. M., Bretherton, C. S., Zelinka, M. D., Klein, S. A., Santer, B. D., and Sanderson, B. M.: Statistical significance of climate sensitivity predictors obtained by data mining, Geophys. Res. Lett., 41, doi:10.1002/2014GL059205, 2014.
b. Lehmann, E: Can Big Data Help U.S. Cities Adapt to Climate Change? White House data splurge meant to "change the game" on climate, Scientific American, 2014.
c. Majda, A. J., and Yuan: Fundamental Limitations of Ad Hoc Linear and Quadratic Multi-Level Regression Models for Physical Systems, Discrete Contin. Dyn. S., 17(4), 1333-1363, doi:10.3934/dcdsb.2012.17.1333, 2012.
d. Majda, A. J., and Harlin, J.: Physics Constrained Nonlinear Regression Models for Time Series, Nonlinearity, 26(1), 201-217, doi:10.1088/0951-7715/26/1/201, 2013.

9. In section 8 – Enhanced Understanding and Predictions, added the reference Kinney and Atwal, 2014 to modify the following sentence: *"Developments in correlative analysis (Khan et al., 2007; Reshef et al., 2011; Kinney and Atwal, 2014), extended to handle correlated data at multiple spatial and temporal scales, may help quantify conceptual understanding and possibly even discover new dependencies (Khan et al., 2006)."*
   Added the following reference for this addition and change:
   a. Kinney, J. B. and Atwal, G. S.: Equitability, mutual information, and the maximal information coefficient, Proc. Natl. Acad. Sci. USA, 111(9) 3354-3359, doi:10.1073/pnas.1309933111, 2014.
10. Added two authors to the author list: W. Liao and A. Agrawal.
11. In section 8 – Enhanced Understanding and Predictions, changed the sentence that formerly read: "New methods in nonlinear data sciences, from complex networks (Steinhaeuser et al., 2011a) to multifractals (García-Marín et al., 2013; Muzy et al., 2006), have demonstrated initial promise for better description and predictive insights on climate-related extremes, such as extreme monsoonal rainfall over South Asia (Malik et al., 2011)." to *"Applications of methods in nonlinear data sciences, from complex networks (Steinhaeuser et al., 2011a) to multifractals (Garc\'{\i}a-Mar\'{\i}n et al., 2013; Muzy et al., 2006), and dynamic Bayesian networks (Troy et al., 2013) have demonstrated initial promise for better description and predictive insights on climate-related extremes, such as extreme monsoonal rainfall over South Asia (Malik et al., 2011)."*
    Added the following reference for this addition and change:

    a. Troy, T. J., Devineni, N., Lima, C., and Lall, U.: Moving towards a new paradigm for global flood risk estimation, European Geosciences Union General Assembly, 7-12, 2013.

12. In section 5 – Characterization of Climate Extremes, change: "Fig. 3 (bottom) presents fully automated and computationally efficient spatio-temporal characterization of long-term droughts using a Markov random field-based approach (Fu et al., 2012)." To: *"Figure 3 (b-c) presents fully automated and computationally efficient spatio-temporal characterization of long-term droughts using a Markov random field (MRF)-based approach (Fu et al., 2012); this type of MRF approach has been validated by automatically detecting the intertropical convergence zone from instantaneous satellite data (Bain et al., 2011)."*
    Add the following reference for this change:
    a. Bain, C. L., De Paz, J., Kramer, J., Magnusdottir, G., Smyth, P., Stern, H., and Wang, C.: Detecting the ITCZ in Instantaneous Satellite Data using Spatiotemporal Statistical Modeling: ITCZ Climatology in the East Pacific, J. Clim., 24, 216–230, doi:10.1175/2010JCLI3716.1, 2011.
13. In section 9 – Summary, the following sentence was changed: "Improving regional projections (e.g., through variable selection or statistical downscaling) and characterizing natural variability (e.g., irreducible uncertainty at decadal scales: Deser et al., 2012) are necessary for informing adaptation at stakeholder-relevant scales and planning horizons." Via the addition of several new references: "*Improving regional projections (e.g., through variable selection or statistical downscaling) and characterizing natural variability (e.g., irreducible uncertainty at decadal scales: Hawkins and Sutton, 2009, 2011; Branstator and Teng, 2012; Deser et al., 2012a-b; Fischer et al., 2013; Hu and Deser, 2013; Rosner et al., 2014) are necessary for informing adaptation at stakeholder-relevant scales and planning horizons."*
    As a result, the following references were added:
    a. Branstator, G., and Teng, H.: Potential impact of initialization on decadal predictions as assessed for CMIP5 models, Geophys. Res. Lett., 39(12), doi:10.1029/2012GL051974, 2012.
    b. Deser, C., Knutti, R., Solomon, S., and Phillips, A. S.: Communication of the role of natural variability in future North American climate, Nat. Clim. Change, 2(11), 775–779, doi:10.1038/nclimate1562, 2012b.
    c. Hawkins, E., and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bull. Am. Meteorol. Soc., 90(8), 1095–1107, doi:10.1175/2009BAMS2607.1, 2009.
    d. Hawkins, E., and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, Clim. Dyn., 37(1-2), 407–418, doi:10.1007/s00382-010-0810-6, 2011.

   e. Hu, A., and Deser, C.: Uncertainty in future regional sea level rise due to internal climate variability, Geophys. Res. Lett., 40(11), 2768–2772, doi:10.1002/grl.50531, 2013.
   f. Rosner, A., Vogel, R. M., and Kirshen, P. H.: A risk-based approach to flood management decisions in a nonstationary world, Water Resour. Res., 50(3), 1928-1942, doi:10.1002/2013WR014561, 2014.

14. In section 8 – Enhanced Understanding and Predictions, after "However, this method can also be applied for improved understanding of the complex dependence structure between climate variables, especially in a high-dimensional setting (Chatterjee et al., 2012; Das et al., 2012, 2013)." Added the following: "*Dimensionality reduction techniques that utilize manifold, atomic, and topological structures derived directly from physical laws (Kpotufe, 2009; Kpotufe, 2011; Balakrishnan et al., 2013a-b; Kpotufe and Garg, 2013; Lum et al., 2013; Wang et al., 2014) at once could make the prediction problem both more computationally tractable and physically sensible.*"
    As a result, the following references were added:
    a. Balakrishnan, S., Rinaldo, A., Singh, A., Wasserman, L., Tight Lower Bounds for Homology Inference, arXiv, 2013a.
    b. Balakrishnan, S., Narayanan, S., Rinaldo, A., Singh, A., Wasserman, L., Cluster Trees on Manifolds, arXiv, 2013b.
    c. Kpotufe, S.: Fast, smooth and adaptive regression in metric spaces, Adv. Neur. In., 2009.
    d. Kpotufe, S.: k-NN Regression adapts to local intrinsic dimension, Adv. Neur. In., 2011.
    e. Kpotufe, S., Garg, V. K.: Adaptivity to Local Smoothness and Dimension in Kernel Regression. Adv. Neur. In., 2013.
    f. Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., and Carlsson, G.: Extracting insights from the shape of complex data using topology, Nature Scientific Reports, 3, 1236, doi:10.1038/srep01236, 2012.
    g. Wang, H., Fazayeli, F., Chatterjee, S., and Banerjee, A.: Gaussian Copula Precision Estimation with Missing Values, International Conference on Artificial Intelligence and Statistics, 2014.
15. We added clarity throughout the manuscript in terms of figure subpanels – all figures except figure 2 now have subpanels denoted by letters, i.e., (a). All references to figures, either in the main text or in the figure manuscripts, now refer to those subpanels as appropriate. This should further clarify discussions surrounding the figures.