Supplementary Material

APPENDIX 1: SINGLE FACTOR CLIMATE EXPERIMENTS (cSFEs)

Within the overall C-FEWS effort, cSFEs are analyzed over both historical and future time domains. The explanations given here apply specifically to the initial analysis reported in this paper describing results for the historical period (1980-2019). However, the same three Approaches are applied to future forecast periods and will be reported in forthcoming work. We use a series of specific examples below to demonstrate how our methodology works. These examples are drawn from a much larger matrix of interactions, are hypothetical, and used for demonstration purposes only.

APPROACH A:  Single Year-Event Studies

EXAMPLE: DROUGHT DURING EARLY PERIOD

The procedures for computing FEWS sensitivity to a specific drought year are outlined below, during the initial period of the historical record for a single region. These steps are repeated for other climate event categories (extreme precipitation, cold and heat waves), periods of time (early, middle, late), and regions of interest. These generic steps are applied to each assessment model output variable.
**A1.A.1: Compute all necessary Baseline assessment model output variables under the early period drought event.** The key assessment model variables (Table 4) are recorded and matched to the early period drought year identified within the historical Baseline time series (1980-2019) and its +/- 2-year surrounding window (Table 2) (e.g., for the Midwest the event year is 1988 with temporal range from 1986 to 1990).

**A1.A.2: Compute a DELTA (sensitivity) metric for each assessment model output variable to assess the impact of the climate event on key FEWS elements.** This step estimates the impact of the extreme year relative to the non-extreme years spanning the event (2 years before and 2 years after). Thus,

\[
\Delta_d = \frac{(Y_{cd} - (Y_{b2} + Y_{a2})/2)}{(Y_{cd} + (Y_{b2} + Y_{a2})/2)},
\]

where \(Y\) is any assessment model output variable, \(cd\) is the climate drought year, \(b2\) and \(a2\) are the 2 years before and after). This yields the climate extreme effect.

Since we also perform single-factor non-climate experiments (ncSFEs) over the same Baseline period, we can simultaneously explore drought impacts as they interact with the four themes. To do this, we compute for each non-climate target output variable, depicted as \(ncd\), a second sensitivity metric of the same general form as Eq. A.1:

\[
\Delta_d = \frac{(Y_{ncd} - (Y_{b2} + Y_{a2})/2)}{(Y_{ncd} + (Y_{b2} + Y_{a2})/2)}.
\]

**A1.A.3: Compute the economic impacts.** Here we determine drought-related impacts in $ terms. For each FEWS service, we use VM to estimate the value of the change in conditions; methodological details are in Chang et al. (this issue). This allows us to show the value of each service change, sum them, and report the aggregate difference.

**A1.A.4: Construct summary statistics and graphics depicting positive and negative cSFE (and any joint ncSFE) effects on TEI and NBI infrastructures.** The statistics are drawn from the 5-year time domain and derived from the \(\Delta\) values (Eqs. A1.A.1, A1.A.2) from which means, aggregate values, ranges and standard deviations of the sensitivities are computed.

**A1.A.5: Repeat A1.A.1-A1.A.4 for the remaining time periods (i.e., middle and late).**

**A1.A.6: Determine aggregate impacts over all three extreme periods (i.e., over full 40 years).** This enables statements to be made regarding the degree to which nature-based or engineered infrastructure variables were differentially sensitive to early period droughts, versus those in the middle and later period, uncovering potential technology, land use or management/ regulatory adaptation (or maladaptation) over the 40 year period. Here we can summarize overall results by averaging or documenting ranges over all three time periods.
APPROACH B: Intensified Extremes

EXAMPLE FOR AMPLIFYING THE HEAT WAVE EFFECT

The procedure here computes system sensitivity to the accumulated impact of heat wave extremes over the last decade of the 40-year historical baseline period across a single region. These steps are repeated for the remaining three climate extremes and other regions of interest.


A1.B.2: Compute a DELTA (sensitivity) metric for each assessment model output variable produced under the intensified climate extreme time series to quantify its impact on key FEWS elements. The cSFE here depicted is for a portion of the historical time series (2010-2019) with the increased frequency of extreme heat wave years tripled. This requires assembling the Baseline results recorded from A.1 for each of the key output variables, using the mean DELTA metric shown below, computed over T2 in the diagram representing the period 2010-19:

\[ \Delta = \frac{(\text{cSFE}_i - \text{Baseline}_i)}{(\text{cSFE}_i + \text{Baseline}_i)} \]  

(A1.B.1)

To minimize the idiosyncratic nature of a particular year, T2 represents the last full decade of the time series, with any output variables computed as a mean over this decade. The \( \Delta \) calculation will show both negative and positive values over the decade. As necessary, we express absolute values for overall impact, but keeping the “+” and “−” sense of the metrics to convey the direction of drought impacts for each assessment model output of interest.
A1.B.3: *Compute the economic impacts.* Here we determine heat wave-related impacts in $$ terms. For each element of the portfolio of services in the economic valuation model we estimate the value of the change in conditions from the baseline run to the cSFE scenario depicting elevated frequency of droughts. We then report the aggregate value of the changes.

A1.B.4: *Construct summary statistics and graphics depicting positive and negative cSFE effects on TEI and NBI infrastructures.* The statistics are drawn from results over the period 2010-19 and are evaluated using the Δ values (Eq. A1.C.1) from which means, aggregate values, ranges and standard deviations are computed.
APPRAOCH C: De-extremed Time Series

The procedure here computes system sensitivity to the accumulated impact of net reduction in extreme precipitation over the last decade (2010-2019) for a single region. These steps are repeated for the remaining three climate event categories and other regions of interest.

A1.C.1: Compute all Baseline assessment model output variables as a 40-year time series (identical to A1.A.1 but for the full time series).

A1.C.2: Compute a DELTA (sensitivity) metric for each assessment model output variable produced under the de-extremed climate time series to quantify its impact on key FEWS elements. The cSFE here depicted operates as over the historical time series from 2010-19 with the extreme precipitation extremes removed during this last decade. This requires assembling the Baseline results recorded from A.1 for each of the key output variables, using the DELTA metric shown below:

\[
\Delta = \frac{\sum_{i=T_1}^{T_2} (cSFE_i - Baseline_i)}{\sum_{i=T_1}^{T_2} (cSFE_i + Baseline_i)}
\] (A1.C.1)

To minimize the idiosyncratic nature of a particular year, T2 represents the last full decade of the time series, with any output variables computed as a mean over this decade. The \( \Delta \) calculation will show both negative and positive values over the decade. As necessary, we express absolute values for overall impact, but keeping the “+” and “-” sense of the metrics to convey the direction of high precipitation impacts for each assessment model output of interest.
A1.C.3: Compute the economic impacts. Here we determine extreme precipitation-related impacts in $ terms. For each element of the portfolio of services in the economic valuation model we estimate the value of the change in conditions from the baseline run to the cSFE scenario depicting the absence of significant extreme precipitation. We then report the aggregate value of the changes.

A1.C.4: Construct summary statistics and graphics depicting positive and negative cSFE effects on TEI and NBI infrastructures. The statistics are drawn from results over the period 2010-19 and are evaluated using the Δ values (Eq. A1.C.1) from which means, aggregate values, ranges and standard deviations are computed.
APPENDIX 2: NON-CLIMATE SINGLE FACTOR EXPERIMENTS (ncSFEs)

As for the cSFEs, ncSFEs can be analyzed over decade and multi-decadal time horizons. The explanations given below apply specifically to our initial analysis for the historical (1980-2019) period. The methodology described below is also applied to future forecast periods and will be reported elsewhere. We use one specific example below to demonstrate how our methodology works. It is drawn from a much larger matrix of interactions, is hypothetical, and used for demonstration purposes only.

EXAMPLE FOR A SINGLE TECHNOLOGY-RELATED VARIABLE

The methodology computes long-term system response to different variables representing technology applied across the time domain of the historical past. In these experiments, we inactivate a specific, individual variable (defined as technology-related), and then compare results to the baseline, where the same technology input time series remains in force. These steps can be repeated for the remaining technology variables and any associated with the remaining two non-climate factor categories (land use, management/regulation). Summary statistics can also be used to estimate the importance of each of the non-climate categories collectively. Results are repeated for other regions of interest.

A2.1: Compute all baseline assessment model output variables as a 40-year (for Approach A) or 10-year time series (Approaches B & C).

A2.2: Inactivate the technology-related input in question and re-run the assessment models.

A2.3: Compute a DELTA value for each output variable produced under scenario time series. This requires assembling the Baseline results recorded from A1.1 for each of the key output variables, using the DELTA metric shown below:

\[
\Delta = \sum_{i=T^1}^{T^2} \frac{(ncSFE_i - Baseline_i)}{\sum_{i=T^1}^{T^2} (ncSFE_i + Baseline_i)}
\] (A2.1)

To minimize the idiosyncratic nature of a particular year, T1 and T2 represent start and end years of any particular analysis. These could include full start and end decades, 1980-89 and 2010-2019, respectively, or individual year boundaries in the historical period, with the output variables computed as means over any of the chosen time periods. In some cases, we fixed the value of particular input variable for the entire historical time period to isolate its effect on assessment model outputs relative to Baseline (e.g., fixing engineered wastewater treatment at zero for N or reducing N fertilization to 0% of the contemporary value in corn cultivation). The \( \Delta \) calculation will show both negative and positive values. As necessary, we express absolute values for overall impact, but keeping the “+” and “-” sense of the metrics to convey the direction of impacts for each assessment model output of interest. (For use with Approach B & C the \( \Delta \) is computed over the 10-year period 2010-2019 only).
A2.4: **Compute the economic impacts.** Here we determine drought-related impacts in $S$ terms. For each element of the portfolio of services in the economic valuation model, we estimate the value of the change in conditions from the baseline run to the technology scenario. We then report the aggregate value of the changes.

A2.5: **Construct summary statistics and graphics depicting positive and negative ncSFE effects on TEI and NBI.** The statistics are drawn from the 10 or 40-year experiments and are evaluated using the $\Delta$ values (Eq. A2.1) from which means, aggregate values, ranges and standard deviations are computed. This enables statements to be made regarding the degree to which engineered or nature-based infrastructure variables were differentially sensitive to repeated droughts. From this analysis, statements can be made regarding the capacity of the non-climate factors to attenuate (or exacerbate) a multi-decadal history of drought extremes.
APPENDIX 3: MULTI-FACTOR EXPERIMENTS (MFEs)

These experiments are designed to explore major interactions across the climate events and non-climate input factors that together define aggregate FEWS behaviors. Constructing appropriate scenarios of these factors can help to better understand some of the key interactions across the different FEWS sectors. These experiments can be made in conjunction with Approach A, B, or C. Analysis can be made over past and future time domains but the explanations given here also apply specifically to initial analysis over the historical (1980-2019) period. Our methodology described below is also applied to future forecast periods and will be reported elsewhere. The general architecture of the experiments relies on the inactivation of chosen sets of variables. We use one specific example below to demonstrate how our methodology works. This example is drawn from a much larger matrix of interactions, is hypothetical, and used for demonstration purposes only.

EXAMPLE OF CLIMATE APPROACH B FOR DROUGHT PLUS A TECHNOLOGY-RELATED VARIABLE

The methodology computes long-term system response to different variables representing technology applied across the time domain of the historical past for a particular region, under the condition of accentuated drought. In this experiment, we apply the Approach B climate together with an inactivated individual non-climate variable (one of those that are technology-related), and then compare results to the baseline, where the recorded climate and the original historical technology input time series remain in force. Various combinations (including those with multiple non-climate variables) can be considered as the analysis unfolds. Summary statistics can also be used to estimate the importance of each of the non-climate categories collectively.

A3.1: Compute all baseline assessment model output variables as a 10-year time series (this is identical to Step 1, cSFE Approach A, but here for the shorter time series).

A3.2: Assemble climate Approach B time series and inactivate one of the technology-related input variables (Table 3 of main text).

A3.3: Compute all assessment model output variables as 10-year time series under this combined scenario.

A4.4: Compute an aggregate DELTA value for each output variable produced under the scenario time series. This requires assembling the Baseline results recorded from Step 1, Approach A for each of the key output variables, using the DELTA metric shown below:

$$\Delta = \frac{\sum_{t=T1}^{T2} (MFE_t - Baseline_t)}{\sum_{t=T1}^{T2} (MFE_t + Baseline_t)}$$  \hspace{1cm} A3.1$$

To minimize the idiosyncratic nature of a particular year, T1 and T2 represent start and end years of any particular analysis. These could include full start and end decades, 1980-89 and 2010-2019, respectively, or individual year boundaries in the historical period, with the output variables computed as means over any of the chosen time periods. As for some of the ncSFEs described in Appendix 2, we also can fix the values of particular combinations of input variables for the entire historical time period to isolate their effects on assessment model outputs relative to Baseline (e.g., fixing engineered wastewater treatment at zero for N while simultaneously reducing N fertilization to 50% of the contemporary value in corn cultivation). The Δ calculation will show both negative and positive values. As necessary,
we express absolute values for overall impact, but keeping the “+” and “−” sense of the metrics to convey the direction of impacts for each assessment model output of interest. (For Approach B & C the $\Delta$ is computed over the 10-year period 2010-2019 only).

A3.5: When a modified climate time series is part of the MFE scenario, compute a second category of metric (the Sensitivity Index) to assess and decouple of the role of climate from non-climate factors. That index combines two individual $\Delta$ measures and is computed as:

$$S_d = \Delta_{MFE} - \Delta_{NC},$$

where $MFE$ is aggregate system behavior under the combined scenario, here referring to Approach B drought extreme plus the single technology variable considered. $NC$ is the non-climate SFE analog operating under the Baseline climate (multiple non-climate factors can also be considered). The $S_d$ term gives a measure of the importance of the climate factor (drought in this case) in defining the aggregate system behavior, stripping away the impacts of the other non-climate factors that are in force.

A3.6: Compute the economic impacts. Here we determine aggregate drought-related impacts in $\$$ terms for the portfolio of services in the economic valuation model. We do this by using VM to estimate the values of each of the changes in the assessment model outputs for the baseline run and then for this combined multi-factor scenario. We then report the aggregate difference.

A3.7: Construct summary statistics and graphics depicting positive and negative effects on Green/Gray/Combined infrastructures. The statistics are drawn from the 10-year
experiments and are evaluated using the Δ values from which means, aggregate values, ranges and standard deviations are computed. This enables statements to be made regarding the degree to which green or engineered infrastructure variables were differentially sensitive to the absence of drought. From this analysis, statements can be made regarding the capacity of the non-climate factors to attenuate (or exacerbate) a multi-decadal history of drought extremes.
APPENDIX 4: C-FEWS PERFORMANCE MEASURICS AND MODEL CALIBRATION / VALIDATION

Each modeling sub-group within C-FEWS routinely executes extensive calibration and validation, and we detail the specific data resources below. We have built redundancy into our approach to enable cross-checks (e.g., hydrological flux estimates from ISAM, TEM, and WBM; ISAM [process-based] and TEM [reduced form] simulations intercompared to determine biofuel production uncertainties).

ISAM is validated for major food crops (e.g., corn, soybean, spring and winter wheat, rice) and 2nd-generation bioenergy crops (e.g., Miscanthus & two switchgrass cultivars) across an array of geographies and studies including those representing weather extremes (Song et al., 2013; Song et al., 2014; Song et al., 2016; Xu et al., 2021). In C-FEWS, we further evaluate ISAM results under extreme climate events against USDA county-level survey data on key phenological dates (planting, anthesis, maturity) as well as crop yields and production.

TEM simulations of C, N and water fluxes are validated at two spatial scales (Lu et al., 2013; Lu et al., 2015). At the finest, we compare modeled estimates to those from field studies on biomass, net primary production (NPP), and crop yield including those from long-term ecological research (LTER), long-term agro-ecosystem research (LTAR) and USDA ARS Conservation Effects Assessment Project (CEAP) sites. Similar to previous studies (Felzer et al., 2004; Tian et al., 2011; Lu et al., 2013), we also compare model estimates to net ecosystem exchange of C and latent heat flux from eddy covariance measurements (e.g. (Felzer et al., 2009; Luo et al., 2011; Melillo et al., 2011)) at several sites in a region of interest. At the state-level, model results for C sinks are compared to biomass increment estimates from the USFS Inventory and Analysis (FIA) (Lu et al., 2013; Lu et al., 2015).

The WBM/WTM, TP2M, and SPARROW models are tested against existing calibration and validation data, specifically: observed discharge, temperature and nitrogen flux at USGS gauges (Stewart et al., 2013), USGS-NAWQA (Moore et al., 2004), and multi-agency data syntheses (Saad et al., 2011). Three-tiered evaluation is made of land-to-water transfer predictions by ISAM and TEM using headwater sites (< 10 km2), and for SPARROW using larger stream and river sites validating river network loadings, point source inputs (i.e., wastewater treatment plants), and aquatic transformations regulating heat (Miara et al., 2013; Stewart et al., 2013), C, and nutrient fluxes. Model calibration and validation is executed using split time and location (one period/site for calibration,
another for validation), and by cross-validation using repeated sub-sampling of randomly selected sets of calibration/validation sites. We use rating-curve methods (USGS-LOADEST (Saad et al., 1991); recent refinements to WRTDS (Hirsch et al., 2010) to interpolate over gaps in water quality time series (e.g., (Raymond et al., 2008)).

VM cannot be conventionally validated, as there are no observable social values to which their economic estimates can be compared. However, the biogeophysical flows represented in the C-FEWS Services Portfolios can be validated as described above, and those uncertainties carried over into the economic valuation. In VM, confidence intervals are estimated for value estimates based on uncertainty regarding the parameter assumptions in the social surplus models and on confidence intervals associated with benefit transfer from previous non-market value estimates.