MDPI

Article

Characterizing Prediction Uncertainty in Agricultural Modeling via a Coupled Statistical–Physical Framework

John C. Chrispell [1,*], Eleanor W. Jenkins [2], Kathleen R. Kavanagh [3] and Matthew D. Parno [4,*]

1 Department of Mathematical and Computer Sciences, Indiana University of Pennsylvania, Indiana, PA 15705, USA
2 School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC 29634, USA; lea@clemson.edu
3 Department of Mathematics, Clarkson University, Potsdam, NY 13699, USA; kkavanag@clarkson.edu
4 Department of Mathematics, Dartmouth College, Hanover, NH 05755, USA
* Correspondence: john.chrispell@iup.edu (J.C.C.); Matthew.D.Parno@dartmouth.edu (M.D.P.)

Abstract: Multiple factors, many of them environmental, coalesce to inform agricultural decisions. Farm planning is often done months in advance. These decisions have to be made with the information available at the time, including current trends, historical data, or predictions of what future weather patterns may be. The effort described in this work is geared towards a flexible mathematical and software framework for simulating the impact of meteorological variability on future crop yield. Our framework is data driven and can easily be applied to any location with suitable historical observations. This will enable site-specific studies that are needed for rigorous risk assessments and climate adaptation planning. The framework combines a physics-based model of crop yield with stochastic process models for meteorological inputs. Combined with techniques from uncertainty quantification, global sensitivity analysis, and machine learning, this hybrid statistical–physical framework allows studying the potential impacts of meteorological uncertainty on future agricultural yields and identify the environmental variables that contribute the most to prediction uncertainty. To highlight the utility of our general approach, we studied the predicted yields of multiple crops in multiple scenarios constructed from historical data. Using global sensitivity analysis, we then identified the key environmental factors contributing to uncertainty in these scenarios' predictions.

Keywords: stochastic modeling; Gaussian processes; agricultural modeling; crop yield; water management; global sensitivity analysis

1. Introduction

Characterizing uncertainty in model predictions is important for understanding risk and constructing robust models which aid in the decision making process. The risks of agricultural business decisions, which can include planting and harvesting schedules, irrigation needs, and market prices, depend heavily on environmental factors, which are inherently uncertain. In this work, we describe the development of a flexible framework for modeling aspects of the agricultural system and understanding the impact of meteorological variability on crop yields. Our approach relies on coupling stochastic process models of important meteorological factors (e.g., temperature, wind speed, and solar radiation) with a physical model of crop yield based on the well-known Penman–Monteith equation. We demonstrate the utility of our combined statistical–physical modeling framework through a probabilistic characterization of the impact of irrigation limits on yield and by leveraging global sensitivity analysis to identify the meteorological factors that contribute most to uncertainty in yield predictions.

The importance of making appropriate agricultural decisions cannot be overstated. Worldwide increases in population have led forecasters to predict a 70% increase over 2018 levels of food production will be needed to meet demand by 2050 [1]. As farmers

work towards meeting this goal, they are faced with difficulties imposed by climate change and increasingly limited resources. Mean values for environmental parameters are being shifted (e.g., average temperature is increasing), and the variability in these parameters is increasing (e.g., the variance of daily temperature is growing) [2,3]. In addition, farmers have a vested interest in adopting economically and environmentally sustainable agricultural practices. Consumer interest in local agriculture is also continuing to grow [4,5] and there are many environmental and socioeconomic advantages to smaller scale local production [6–8]. The COVID-19 pandemic in particular has highlighted the resilience of local food supply networks [9].

Significant improvements in data collection and reporting help farmers adjust their practices and incorporate strategies to maintain their livelihoods [1]. Measurements for a wide variety of environmental metrics, taken in an almost continuous time frame, are reported from stations located throughout the country. This data inundation can lead to information overload, where there is no clear understanding of how these metrics and their associated uncertainties couple to affect dynamic processes. Decision support systems help extract useful information from big data and can thus guide decision making and increase efficiency (see, e.g., [10,11]). However, there are indications that the majority of advanced digital agriculture products benefit large producers and agribusinesses more than small producers and individual farmers [12]. This benefit limitation highlights the need for easy-to-use automated decision support tools that can enable all farmers to benefit from advanced data analysis, modeling, and machine learning.

Our focus is on the development of support mechanisms for small, local farming operations, which tend to be resource limited but who are playing increasingly important roles in community food supplies [4]. Our proposed framework could help bring the digital agriculture revolution to these important smaller scale farmers, and the framework used in this work is easily scaled to an agricultural plot of any size. With climate change adaptation in mind, we focus on characterizing uncertainty in water usage and the impacts of different crop choices and watering strategies on yield. While climate change will likely have impacts on the economic aspect of cropping decisions, there is a more direct connection with environmental processes. This work therefore focuses on characterizing uncertainties in environmental processes and propagating those into uncertainties into predicted yield. The techniques developed here could also be easily extended to problems with other decision inputs.

This work was motivated by our earlier efforts to help farming communities simultaneously manage crop and water portfolios [13–16]. Unlike those previous efforts, the focus in this work was on the development of a data-driven framework that is computationally efficient and amenable to uncertainty quantification. We leveraged publicly available historical data, advanced statistical modeling of temporally varying environmental variables, and process-based physical models of soil hydrology to provide probabilistic predictions of crop yield under varying climate conditions and farmer decisions (e.g., crop choice and irrigation strategies). In particular, we focused on uncertainties stemming from variable meteorological conditions and their impacts on available soil moisture and crop yield. In addition to quantifying prediction uncertainty, we used global sensitivity analysis to identify the most impactful environmental variables under different climate conditions (wet or dry). This can be seen as a first step towards a more all-encompassing framework to analyze farming choices under uncertainty. Techniques similar to those presented here could also be used for other decision inputs.

In the following sections, we describe the models used to compute the metrics needed to evaluate farming decisions and the strategies used to provide a statistical characterization of parameters which feed into the associated metrics. We demonstrate the robustness of our framework by evaluating yields under different irrigation strategies using modeled environmental forecasts. The framework predicts crop responses under stressed conditions, which can show farmers the tipping points for the crops they have planted.

2. Physical Model

This section provides an overview of the analytical physics-based models and the environmental parameters used to predict yield. Our overall approach is to provide forecast information for these parameters and evaluate the impacts on yield and the associated water use. In the subsequent section, we describe the strategy for statistical incorporation of environmental time series and the use of observations to calibrate the statistical representation to a given location.

The FAO guidelines model yield as a function of evapotranspiration [17], which is the transfer of water from the land to the atmosphere through evaporation from the soil and transpiration from plants. The FAO model equation for yield is given as [18]

$$\left(1 - \frac{Y_a}{Y_m}\right) = K_y\left(1 - \frac{ET_a}{ET_m}\right) \tag{1}$$

where Y_a is the actual yield, Y_m is the maximum yield, K_y is the yield response coefficient, ET_a is the actual crop evapotranspiration, and ET_m is the maximum crop evapotranspiration in unstressed conditions. The coefficient K_y measures the sensitivity of a plant to a water deficit and changes over a growing cycle. For instance, values for K_y are higher for plants during the flowering period and smaller during the ripening stage [18].

The maximum crop evapotranspiration is defined as [17]

$$ET_m = ET_0 K_c \tag{2}$$

where K_c is the crop coefficient and ET_0 is the baseline reference evapotranspiration. The crop coefficient also depends on the growth stage of the plant; it adjusts the baseline evapotranspiration to account for the presence of evapotranspiration due to crops.

Combining Equations (1) and (2) and solving for actual yield Y_a gives

$$Y_a = Y_m\left(1 - K_y\left(1 - \frac{ET_a}{ET_0 K_c}\right)\right). \tag{3}$$

2.1. Modeling Evapotranspiration

Evapotranspiration appears in Equation (3) as a main predictor of yield; it serves as a proxy for the amount of water consumed by the plant, which is an indicator of its health. The plant water consumption depends on the amount of water available in the soil, which in turn depends on the soil type, soil conditions, and several environmental parameters.

The association between reference evapotranspiration and environmental parameters is defined by the Penman–Monteith equation [18], which models ET_0 as

$$ET_0 = \frac{0.408\delta(R_n - G) + \gamma\frac{900}{T+273}u_2(e_s - e_a)}{\delta + \gamma(1 + 0.34u_2)}. \tag{4}$$

The parameters on the right are described in Table 1. Weather stations collect observations of the meteorological parameters temperature (T), wind speed (u_2), atmospheric pressure (P_a), and actual vapor pressure (e_a). The remaining parameters are derived from these observable variables using the relationships described below. We obtained predictions for ET_0 by training on historical records of these variables and using statistical modeling to build forecasts.

Table 1. Definitions of parameters in the Penman–Monteith equation. We used historical data to develop stochastic models for radiation, temperature, wind speed, and vapor pressure in the framework.

Variable	Definition
ET_0	Reference evapotranspiration [mm day^{-1}]
R_n	Net radiation at the crop surface [MJ m^{-2} d^{-1}]
P_a	atmospheric pressure [kPa]
G	Soil heat flux density [MJ m^{-2} d^{-1}]
T	Mean daily air temperature at 2 m height [°C]
u_2	Wind speed at 2 m height [m s^{-1}]
e_s	Saturation vapor pressure [kPa]
e_a	Actual vapor pressure [kPa]
δ	Slope vapor pressure curve [kPa °C^{-1}]
γ	Psychrometric constant [kPa °C^{-1}]

Following [18], the saturation vapor pressure e_s is given by

$$e_s = 0.6108 \exp\left[\frac{17.27T}{T + 237.3}\right].$$

The slope of this curve, $\delta = \frac{\partial e_s}{\partial T}$ is given by

$$\delta = \frac{\partial e_s}{\partial T} = \frac{4098 e_s}{(T + 237.3)^2}.$$

According to [18] the psychrometric constant γ is proportional to the atmospheric pressure

$$\gamma = 0.665 \times 10^{-3} P_a.$$

The net radiation R_n is estimated using the Stefan–Boltzman law with the corrections for humidity described further in [18]. The expression is

$$R_n \approx \sigma T^4 (0.34 - 0.14\sqrt{e_a}),$$

where $\sigma = 4.903 \times 10^{-9}$ is the Stefan–Boltzman constant.

The value of the soil heat flux density G is positive when the soil is warming and negative when cooling. It is small relative to the net radiation at the crop surface and is often ignored ($G = 0$). Given the stochastic nature of net radiation, we adopt the common assumption of $G = 0$ in our model [18].

2.2. Modeling Soil Water Availability

A plant's access to water is determined by soil water availability. After a precipitation event, an amount of water is held in the soil until it is removed through plant uptake or evaporation. The field capacity of water is the amount of water left in the soil once drainage decreases. As water is removed from the soil, the soil forms a stronger bond with the remaining water, impeding the plant's ability to use it. This water quantity, known as the wilting point, is the water content in the soil matrix where plants will continuously wilt. Thus, the total water available, TAW, depends on the difference between the water content at field capacity, θ_{FC} and the water content for the wilting point, θ_{WP}, multiplied by the root depth Z_r [18]:

$$TAW = 1000(\theta_{FC} - \theta_{WP})Z_r$$

Water content is a dimensionless variable, and Z_r has units [m], meaning *TAW* has units [mm]. This relationship signifies *TAW* depends on both the rooting depth and the soil type. Common values of θ_{FC} and θ_{WP} for different soil types are listed in [18,19].

In addition to having access to water, a plant must be able to pull the water quickly enough from the soil to satisfy the evapotranspiration requirements of its growth stage. The stronger bonds developed between the soil matrix and the water as the water content is reduced deter this transfer and cause the plant to experience water stress before the wilting point is reached. Thus, in reality a plant can access only a fraction *p* of *TAW* before suffering water stress. This is known as readily available water *RAW*; that is,

$$RAW = pTAW.$$

The root zone depletion rate D_r is the water shortage relative to field capacity [18]; at field capacity, $D_r = 0$. Once evapotranspiration begins (i.e., a plant uses available water), field capacity is reduced—or depletion increases. This continues until D_r reaches *RAW*, or when plants undergo water stress. After this point, evapotranspiration is reduced as plants can no longer access needed water supplies. Thus, the actual transpiration ET_a differs from the reference transpiration ET_0. Assuming evapotranspiration from the soil is negligible, ET_a can be determined using [18]

$$ET_a = K_s K_c ET_0 \tag{5}$$

where K_s is the (dimensionless) water stress coefficient which depends on the soil type and available water. When the root zone depletion rate D_r is above the readily available water *RAW*, K_s is determined using

$$K_s = \frac{TAW - D_r}{(1 - p)TAW}, \tag{6}$$

Otherwise, $K_s = 1$. The seasonally changing K_c values lead to each crop having four distinct growing phases over a given growing season. We use piece-wise linear interpolation to obtain intermediate K_c values when needed [18].

The root zone depletion rate D_r is modeled by the differential equation

$$\frac{\partial D_r}{\partial t} = -(P - RO) - I - CR + ET_a(D_r) + DP. \tag{7}$$

which defines the depletion rate as the difference between water entering the system and water exiting the system. Water sources include precipitation *P*, runoff from the soil surface *RO*, the net irrigation depth *I* that infiltrates the soil, and capillary rise from the groundwater table *CR*. Water leaves the system through evapotranspiration $ET_a(D_r)$, and water loss out of the root zone through deep percolation, *DP*. We emphasize the explicit dependence of evapotranspiration on D_r, which occurs through the water stress coefficient as seen in Equations (5) and (6).

We simplify the model by assuming both *DP* and *RO* are zero, and by assuming the water table is deep enough for *CR* to also be zero [18]. In our framework, we represent precipitation stochastically and note that the impact of these assumptions could potentially be accounted for by modifying the statistical description of the precipitation. We update the root zone depletion using the explicit temporal discretization of the simplified version of Equation (7). Let the subscript *i* denote the day of the simulation. The update of D_r from day $i - 1$ to day *i* then takes the form

$$D_{ri} = \max\{0, D_{ri-1} - P_{i-1} - I_{i-1} + ET_a(D_{ri-1})\}, \tag{8}$$

where the max is included to ensure that even after discretization, root zone depletion is also negative, which it should be by definition.

2.3. Strategies for Irrigation

We examine the impacts of three distinct irrigation procedures to generate the value of the irrigation, I, and measure the total amount of water used annually. Specifically, we consider no irrigation, limited irrigation, and full irrigation. In the no irrigation case, $I = 0$. In the full irrigation case the daily value of I is set for each crop using the following process:

Step 1: The K_s value is updated based on the previous day's zone depletion, D_r:

$$K_s = \min\left\{1.0, \frac{TAW - D_r}{(1.0 - p)TAW}\right\}$$

Step 2: The value of ET_a is assigned using K_s as defined in Equation (5).

Step 3: The necessary irrigation is computed as the difference between the root zone depletion (as computed in Equation (8)) and readily available water:

$$I = \max\{D_r - RAW, 0\}.$$

In the limited irrigation case, a cap is specified on the amount of water that can be used per acre of planting per day, I_{max}, and the daily irrigation value satisfies:

$$I = \min\{I_{max}, \max\{D_r - RAW, 0\}\}.$$

3. Statistical Characterization of the Environment

The reference evapotranspiration ET_0 defined in (4), and thus the yield, depends on the environmental parameters listed in Table 1. Historical meteorological observations from systems such as the California Irrigation Management Information System (CIMIS) [20] can be used to drive the model described in Section 2, but these deterministic simulations are limited in their utility. They cannot predict into the future without accurate forecasts of the weather, and being deterministic, cannot be used to characterize uncertainty in future predictions. The following sections describe our approach for overcoming these challenges with additional statistical models of the environmental parameters. These models were trained on historical CIMIS data and allowed us to construct a coupled physical-statistical model that is amenable to uncertainty quantification and global sensitivity analysis, enabling us to assess which meteorological components contribute the most to prediction uncertainties.

3.1. Gaussian Process Background

Gaussian processes (GPs) are a key building block in our statistical modeling approach. A GP is a continuous analog of a vector-valued Gaussian distribution. Whereas a Gaussian distribution is completely characterized by a mean vector μ and a covariance matrix Σ, in the GP setting, this becomes a mean function $m(t)$ and a covariance function $k(t, t')$, both of which depend on an independent variable $t \in \mathbb{R}^D$. For a process $u(t) \sim GP(m(t), k(t, t'))$ and any finite collection of points $t_1, t_2, \ldots, t_N$, the vector $[u(t_1), u(t_2), \ldots, u(t_N)]$ will have a multivariate Gaussian distribution. In our applications, t represents time and $D = 1$. GPs are a flexible way of describing random fields. The form of the covariance kernel, in particular, can be chosen to reflect the underlying structure of the field (e.g., periodicity and multiscale behavior). The sections below demonstrate how we leverage this to model the environmental fields needed to compute the reference evapotranspiration ET_0. For variables such as precipitation, which are not naturally represented as Gaussian random variables, we use Gaussian processes to capture temporal correlations in parameters of more complicated statistical models. Note that while we model each environmental variable independently, the models are trained on the same data, and the predictions of different parameters are only conditionally independent; the predictions will be correlated. The historical observations and periodic structure tie future values of the different environmental parameters together.

We designed covariance kernels to represent our prior knowledge about the environmental parameters. To do this, we leverage the fact that sums and products of valid covariance kernels are also valid covariance kernels. This allows us to capture relatively complex structure by combining canonical kernels with closed forms. In particular, we consider combinations of four different parameterized families of covariance kernels: radial basis kernels, Matern kernels, periodic Matern kernels, and constant kernels. All of these kernels are available in the GPy python package [21]. Parameters in the kernels are optimized by GPy to maximize the likelihood of historical CIMIS training data. When optimizing the GP hyperparameters, it is possible for length scales in the covariance kernels to become small and therefore cause the GP to behave like white noise as it tries to capture structure in noisy data. To prevent this from occurring, we place Gamma hyperprior distributions on the length scales that encourage the optimized length scales to capture the annual correlations observed in the data.

3.1.1. RBF Kernel

The radial basis function (RBF) kernel is given by

$$k_{\text{RBF}}(t, t'; \sigma, L) = \sigma^2 \exp\left(-\frac{1}{2}\frac{\|t - t'\|^2}{L^2}\right), \tag{9}$$

where σ^2 is the marginal variance of the process and L is a length scale that controls how quickly samples of the GP vary over time. The form of the RBF kernel guarantees all samples of a GP with kernel $k_{\text{RBF}}(t, t'; \sigma, L)$ are infinitely differentiable.

3.1.2. Matern Kernel

The Matern family of covariance kernels are a generalization of the RBF kernel where the differentiability can be varied by a parameter ν. In this work we consider Matern kernels with $\nu = 3/2$, which results in continuously differentiable functions. The form of the kernel is given by

$$k_{32}(t, t'; \sigma, L) = \sigma^2 \left(1 + \frac{\sqrt{3}\|t - t'\|}{L}\right) \exp\left(-\frac{\sqrt{3}\|t - t'\|}{L}\right), \tag{10}$$

where again σ^2 is the marginal variance and L is a length scale parameter.

3.1.3. Periodic Matern Kernel

In the GP setting, a covariance kernel defines a reproducing kernel Hilbert space (RKHS) of functions. The periodic Matern class of covariance kernels comes from the considering a subspace of the Matern RKHS that contains periodic functions. Following [22], the family of periodic Matern kernels takes the form

$$k_p(t, t'; \sigma, L, \tau) = F(t)G^{-1}F(t'), \tag{11}$$

where

$$F(t) = \left[\sin\left(\frac{\pi x}{\tau}\right), \cos\left(\frac{\pi x}{\tau}\right), \sin\left(\frac{2\pi x}{\tau}\right), \cos\left(\frac{2\pi x}{\tau}\right), \ldots, \cos\left(\frac{N_f \pi x}{\tau}\right), \sin\left(\frac{N_f \pi x}{\tau}\right)\right]$$

is a vector containing evaluations of a $2N_f$ trigonometric basis functions with the same period τ and defined on some interval $[a, b]$. The matrix G is given by

$$G_{ij} = \frac{L^3}{12\sqrt{3}\sigma^2} \int_a^b \left(\frac{3}{L^2}F_i(t) + 2\frac{\sqrt{3}}{L}\frac{dF_i}{dt} + \frac{d^2F_i}{dt^2}\right)\left(\frac{3}{L^2}F_j(t) + 2\frac{\sqrt{3}}{L}\frac{dF_j}{dt} + \frac{d^2F_j}{dt^2}\right)dt$$
$$+ \frac{1}{\sigma^2}F_i(a)F_j(a) + \frac{L^2}{3\sigma^2}\frac{dF_i(a)}{dt}\frac{dF_j(a)}{dt},$$

which is the Gram matrix of the trigonometric basis computed in the Matern kernel's RKHS. Note that this periodic kernel is different than the simpler periodic kernel $\sigma^2 \exp[-2\sin^2(\pi\|t-t'\|/P)/L^2]$ described in [23]. A discussion of the differences can be found in [22].

3.1.4. Constant Kernel

We also employ a constant kernel, which takes the form

$$k_c(t, t'; \sigma) = \sigma^2. \tag{12}$$

The term "constant" is used to describe this kernel because all realizations of a GP defined by k_c will be constant functions. This kernel is useful when combined with other kernels because it allows modeling constant, but unknown, offsets.

3.1.5. White Noise Kernel

A GP with no temporal correlation (i.e., white noise) has a covariance kernel of the form

$$k_w(t, t'; \sigma) = \sigma^2 \delta(t - t') \tag{13}$$

This kernel is used to capture observation noise and small scale variations with length scales shorter than our daily observations can capture.

3.2. Stochastic Process Models for Environmental Parameters
3.2.1. Temperature

The air temperature $T(t)$ has both quasi-periodic annual cycles (higher in summer, lower in winter) and long term trends (e.g., a week-long heat wave). To capture the quasi-periodic structure, we use the product of a periodic kernel k_p and RBF kernel k_{RBF}. We then add an RBF kernel to capture the long term trends. White noise and constant kernels are used for observation noise and mean temperatures, respectively. The resulting kernel takes the form

$$\begin{aligned} k_T(t, t'; \theta_T) = k_p(t, t'; \sigma_1, L_1, \tau = 1)\, k_{\mathrm{RBF}}(t, t'; \sigma_1, L_2) \\ + k_{\mathrm{RBF}}(t, t'; \sigma_3, L_3) + k_c(t, t'; \sigma_4) + k_w(t, t'; \sigma_5), \end{aligned} \tag{14}$$

which has seven free parameters $\theta_T = [\sigma_1, L_1, L_2, \sigma_3, L_3, \sigma_4, \sigma_5]$ that can be tuned to the historical data. The period of the periodic kernel is fixed to 1 year to capture the known annual cycles of the temperature.

3.2.2. Net Radiation

The net radiation $R_n(t)$ has a similar structure to temperature; it has annual cycles and long term trends. We therefore employ a kernel with the same form as k_T, but with different parameters:

$$\begin{aligned} k_R(t, t'; \theta_R) = k_p(t, t'; \sigma_1, L_1, \tau = 1)\, k_{\mathrm{RBF}}(t, t'; \sigma_1, L_2) \\ + k_{\mathrm{RBF}}(t, t'; \sigma_3, L_3) + k_c(t, t'; \sigma_4) + k_w(t, t'; \sigma_5), \end{aligned} \tag{15}$$

where again $\theta_R = [\sigma_1, L_1, L_2, \sigma_3, L_3, \sigma_4, \sigma_5]$.

3.2.3. Atmospheric Pressure

Unlike the temperature and net radiation, the atmospheric pressure must, by definition, be positive. Instead of using a Gaussian process directly on the pressure, we treat

pressure as a lognormal process. In particular, $\log(P_a) \sim GP(0, k_{P_a}(t, t'))$, where k_{P_a} again takes the form

$$k_{P_a}(t, t'; \theta_{P_a}) = k_p(t, t'; \sigma_1, L_1, \tau = 1) \, k_{\text{RBF}}(t, t'; \sigma_1, L_2)$$
$$+ k_{\text{RBF}}(t, t'; \sigma_3, L_3) + k_c(t, t'; \sigma_4) + k_w(t, t'; \sigma_5). \quad (16)$$

Figure 1 shows that the atmospheric pressure observations from CIMIS have a quasi-periodic structure to the temperature and net radiation.
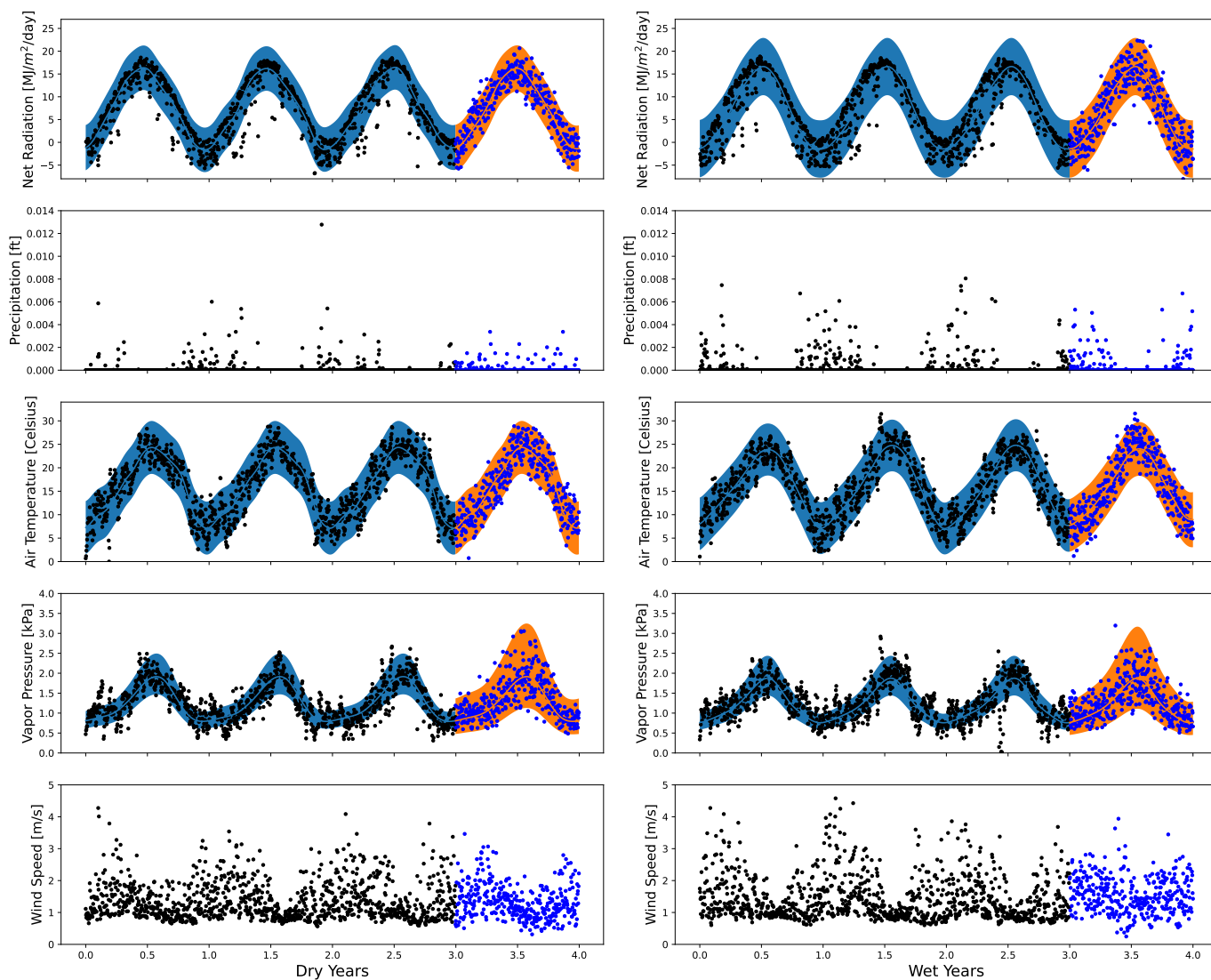

Figure 1. Training and forecast data for environmental parameters. The training data (in black) were chosen from a sequence of dry or wet years according to the SPI in Figure 2. The forecast is shown in blue dots, with the envelopes around the data representing 2 standard deviations of the mean. The key distinction between these model training scenarios is the seen in the wet years having noticeably more precipitation events.

3.2.4. Modeling Precipitation

The daily precipitation $P(t)$ is more complicated to model than the previous environmental variables; there are many days where no precipitation occurs and $P(t)$ cannot be well-represented by a GP directly. Instead, we employ a similar approach to [24] and use a GP to describe the probability of a rain event occurring. Let b_t be a Bernoulli random variable that is 1 when nonzero precipitation occurs on day t and 0 otherwise. The probability

of a rain event on day t is then $\mathbb{P}[b_t = 1]$. Now, let $z(t) \sim GP(0, k_z(t, t'))$ be a latent GP and assume that

$$\mathbb{P}[b_t = 1] = \Phi(z(t)), \tag{17}$$

where $\Phi(\cdot)$ is the CDF of a standard normal distribution. This is commonly called a probit model and is used extensively in Gaussian process classification problems.

The latent GP $z(t)$ and Bernoulli random variable b_t provide a statistical model of when rain events will occur, but do not describe the amount of precipitation. If a rain event occurs, we randomly sample a precipitation amount using the empirical cumulative distribution function of nonzero precipitation amounts $\hat{\Phi}_P : \mathbb{R} \to [0, 1]$, which can be estimated from historical CIMIS data. The precipitation $P(t)$ is given by

$$P(t) = b_t \Phi_u^{-1}(u_t), \tag{18}$$

where $u_t \sim U[0, 1]$ is an independent uniform random variable. Notice that (18) is a mixture model for the precipitation where the probabilities in the mixture are time-dependent and use the latent GP $z(t)$. The kernel used to construct $z(t)$ is given by

$$k_P(t, t'; \theta_{pre}) = k_p(t, t'; \sigma_1, L_1, \tau = 1) \, k_{\text{RBF}}(t, t'; \sigma_1, L_2)$$
$$+ k_{\text{RBF}}(t, t'; \sigma_3, L_3) + k_c(t, t'; \sigma_4). \tag{19}$$

Using GPy [21], the parameters are estimated from historical observations of precipitation events by maximizing the Bernoulli likelihood.

3.2.5. Modeling Wind Speed

Like precipitation, adequately capturing the structure in wind speed observations requires a more sophisticated statistical model. Here we employ a hierarchical model combining a Gamma distribution for marginal wind speeds with temporal GP processes for the parameters in the Gamma distribution.

The average daily wind speed is modeled as a Gamma random variable with time-varying parameters $\alpha(t)$ and $\beta(t)$. These parameters are related to the time-varying mean and variance of the Gamma distribution via

$$\alpha(t) = \frac{m^2(t)}{v(t)},$$

and

$$\beta(t) = \frac{m(t)}{v(t)}.$$

where $m(t)$ is the mean of the Gamma distribution and $v(t)$ is its variance. Modeling the wind speed is now a problem of modeling the mean $m(t)$ and variance $v(t)$ of the Gamma distribution. Given the ability of Gaussian processes to capture both annual cycles and medium term trends, we model both $m(t)$ and $v(t)$ with conditionally independent Gaussian processes (given the observations). In particular, we assume priors $m(t) \sim GP(0, k_m(t, t'))$ and $v(t) \sim GP(0, k_v(t, t'))$, where the kernels are constructed as before:

$$k_m(t, t'; \theta_m)$$
$$= k_p(t, t'; \sigma_1, L_1, \tau = 1) \, k_{\text{RBF}}(t, t'; \sigma_1, L_2) + k_{\text{RBF}}(t, t'; \sigma_3, L_3) + k_c(t, t'; \sigma_4) \tag{20}$$

$$k_v(t, t'; \theta_v)$$
$$= k_p(t, t'; \sigma_1, L_1, \tau = 1) \, k_{\text{RBF}}(t, t'; \sigma_1, L_2) + k_{\text{RBF}}(t, t'; \sigma_3, L_3) + k_c(t, t'; \sigma_4). \tag{21}$$

As training data for $m(t)$ and $v(t)$, we took the 50 day running mean and running variance of the CIMIS wind speed data. Random realizations of the wind speed, such as those in Figure 1, can be obtained by sampling $m(t)$ and $v(t)$, then computing $\alpha(t)$ and $\beta(t)$, and finally drawing (conditionally) independent realizations of the Gamma random variable for each time.

4. Numerical Results

In this section we demonstrate the capability of the framework by comparing irrigation strategies for the yields of several crops using weather information from the California Irrigation Management Information System in Butte County, CA [20]. Forecasts were developed by training on historical data and projecting for a specified set of years into the future. The uncertainty in the training set was projected into the forecast information.

An overview of the modeling framework is given in Figure 3. The user has the ability to designate parameters that effect the modeling on a regional, farm-specific, or municipal level. At the regional level, the training data can be designated to reflect any environmental trends desired. We tested these capabilities by training data on years designated as wet or dry based on historical precipitation events; see Figure 2.

For the farm-specific component of the model, the soil and the crop portfolio parameters are required to reflect what would typically be grown over the the desired forecast period. Municipally or self-enforced limits on the amount of water an individual farm is allotted per acre may be adjusted to predict the effects these limitation have on annual yield and water usage.

The framework was written in python and run using an interactive python notebook. In addition to the use of the standard numerical python and pandas libraries, extensive use was made of the Gaussian processes framework in python, Gpy, maintained by the Sheffield machine learning group [21].
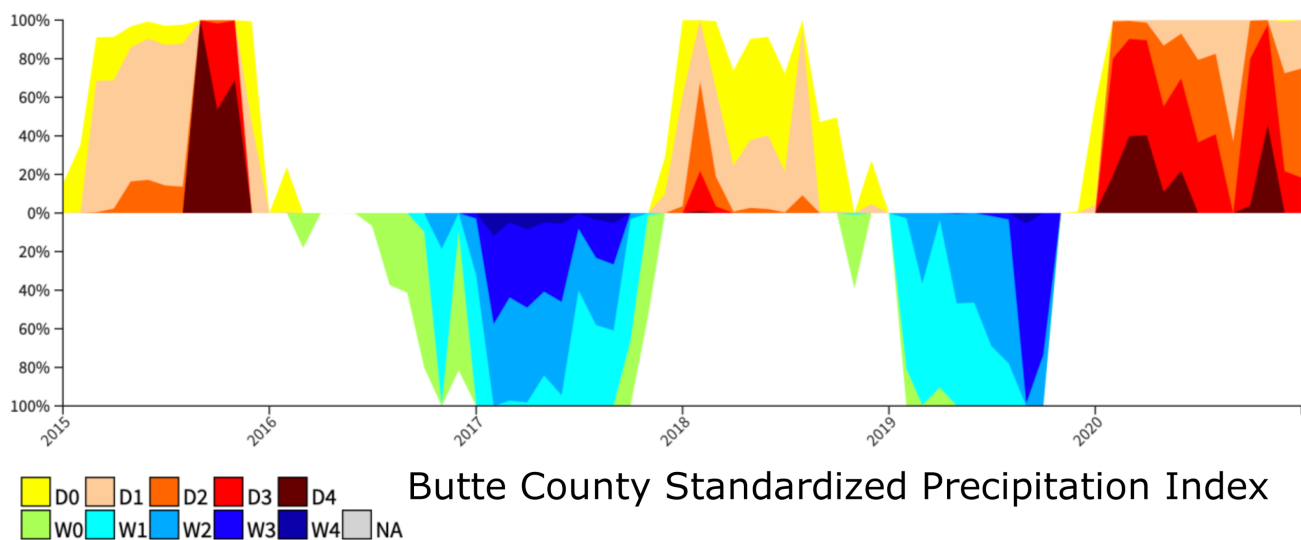

Figure 2. Monthly standardized precipitation index (SPI) from the National Integrated Drought Information System for Butte County, CA. Prefixes D and W denote dry and wet; indices 0–4 correspond to abnormal, moderate, severe, extreme, and exceptional wet and dry periods.
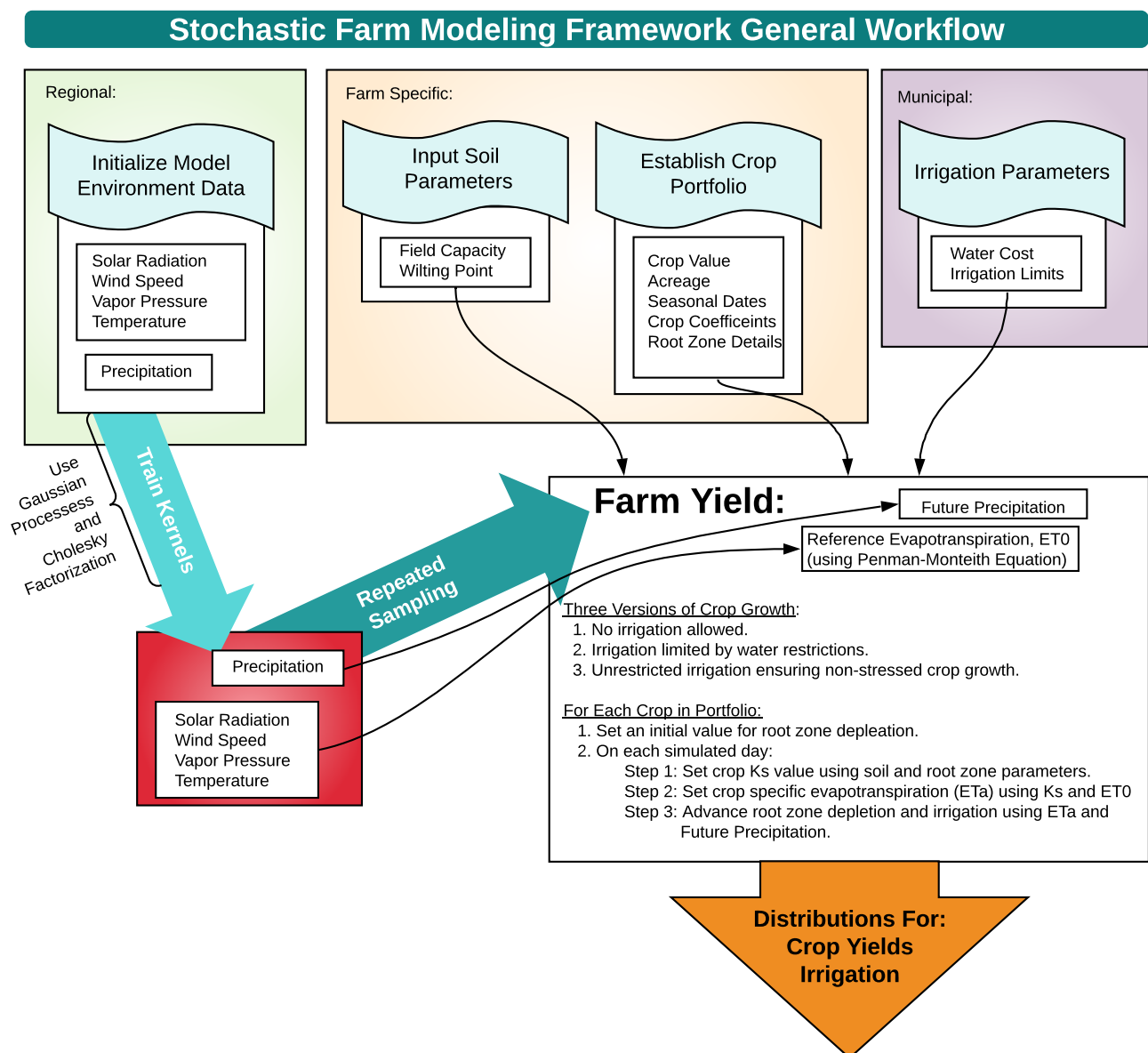
Figure 3. Overview of the stochastic modeling workflow. Note that repeated sampling of trained environment-specific parameters leads to distributions for a modeled crops yield and paired irrigation use in three distinct situations: no irrigation, restricted irrigation, and unrestricted irrigation.

4.1. Classification of Forecasting Data

We culled model training data for both wet and dry periods from the annual CIMIS data recorded from 2015–2020 at Station 12 in Butte County, CA. Wet and dry periods are categorized using the monthly standardized precipitation index (SPI) from the National Integrated Drought Information System [25]. The SPI index characterizes meteorological drought on a timescale ranging from 1 to 72 months, with the SPI depicting the standard deviations that observed cumulative precipitation in a region deviates from the given climatological average. The resulting classification for Butte County is depicted in Figure 2. The modeling framework was then trained using data from 2015, 2018, and 2020 designated as dry years, and 2016, 2017, and 2019 designated as wet training years.

Once the model was trained using environmental data, a collection of forecasts (or samples) were used to predict the yield and corresponding water use for the different irrigation strategies. The projected environment contains uncertainty which mimics the uncertainty seen in its training information.

Examples of the training and forecast environmental data are shown in Figure 1. Amounts of daily precipitation seem to be roughly the same for wet and dry years, but the number of precipitation events is much larger in the wet years (on the right). The forecast data are in blue, and the training data are in black. The forecasts mimic the variability in the training data but are slightly higher.

Figure 4 shows the subtle differences observed in the predictive kernels for the model for radiation, temperature, and air pressure. The expected values for the maximum net radiation and air temperature kernels occurred slightly earlier in the year when trained with the dry year data. With the designated training datasets established, simulated scenarios could be used to study how these slightly different environmental events impact yield and water used under distinct irrigation strategies.
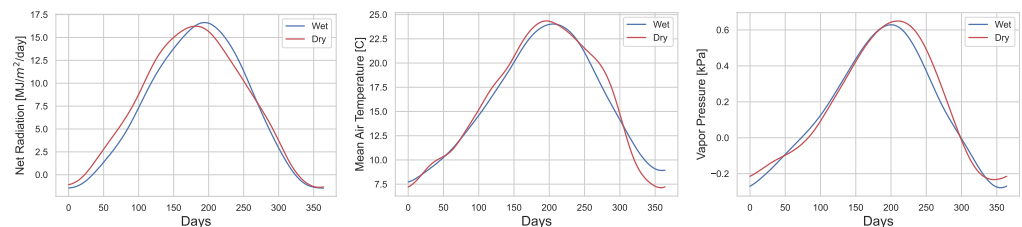

Figure 4. Plots of daily averages for environmental parameters for wet and dry seasons. Net radiation is on the (**left**), temperature is in the (**middle**), and vapor pressure is on the (**right**).

4.2. Example Farming Environment

Once the environmental forecasting has been trained, the framework integrates specific farm properties via the types of soil and crop portfolio. For the purposes of this work, the farm had three representative crops grown, using the irrigation scenarios discussed in Section 2.3. Alfalfa, lettuce, and strawberries were chosen, as they have distinct water needs and growing seasons.

Recall that three scenarios for irrigation were considered: full irrigation ensures each crop gets the full amount of water it needs; no irrigation has only precipitation as the available water source; and limited irrigation has a cap value on the maximum allowed daily irrigation. Unless specified, the default maximum daily irrigation was set to 0.0075 acre-feet per day. The crop coefficients for alfalfa, lettuce, and strawberries used in the model are given in Table 2. The planting dates given in the table were established in pairs, denoting when a crop is put into the ground and subsequently harvested. Lettuce was modeled using biannual growing seasons.

Table 2. Crop parameters used in this study [17,18].

Crop	Ky	Y_m [metric tons/acre]	$K_{c_{beg}}$	$K_{c_{mid}}$	$K_{c_{end}}$	Z_r [m]	p	Planting Dates (start, end)
Alfalfa	1.1	9.072	0.4	0.95	0.9	1.0	0.55	(0, 365)
Lettuce	1.0	2.722	0.0	1.0	0.95	0.4	0.3	(15, 120), (258, 349)
Strawberry	0.85	4.536	0.4	0.85	0.75	0.25	0.2	(60, 273)

Soil characteristics on individual farms can be defined to reflect effects of available groundwater on crop growth. In all modeling presented here, the values of field capacity and wilting point, θ_{FC}, and θ_{WP}, were set to 0.12 and 0.045 respectively, reflecting a sandy

soil structure; however, any soil structure may be specified. The rooting depth, Z_r, for each crop was fixed, but could be adjusted over a growing season.

Predicted growth for an acre of strawberries during a dry year forecast is shown in Figure 5. The projected annual precipitation is shown in the bottom graph in order to establish a connection between available water and strawberry growth. When root zone depletion (D_r), shown just above precipitation, is above the threshold value of readily available water (RAW), crops do not have enough water for full growth. This is seen in the corresponding reduction in ET_a (graphed in the top panel) during this same time period. In our simulations, the initial value of D_r was randomly initialized to a value between 0 and the maximum root zone depletion determined by the rooting depth of the crop. The initial increase in ET_a, which occurs a fifth of the way into the year, corresponds to strawberry planting. Values of ET_a increase according to plant growth, and the leveling of ET_a shown in the dotted black line indicates the lack of available water to ensure continued growth. Irrigation levels for the planted strawberries can be seen in the second panel from the top. When irrigation is unlimited, strawberries are allowed to take as much water as needed to ensure maximum growth. Dips in maximum irrigation requirements correspond to spikes in precipitation events. Under the limited irrigation scenario, the maximum daily threshold was met during a sequence of drier days, when crops need water for growth and precipitation events fall short of supplying this requirement.

The aggregated yield for strawberries over the same forecast event is shown in Figure 6. Restrictions on irrigation have a substantial impact on annual yield, with a significantly lower harvest under the no irrigation strategy. Limited irrigation resulted in an overall difference of around 1 metric ton, whereas no irrigation reduced yield by three times as much.

To further understand the response of the yield model to the irrigation strategies and forecast data, we generated 1000 realizations of the environmental parameters and computed water use and yield for three different crops. The annual water use and yields for alfalfa, lettuce, and strawberries are shown in Figures 7 and 8, for both wet and dry training data.

As seen in Figure 7, on average, more water is required during a dry forecast (the grey distributions), regardless of the crop or the irrigation strategy. For a crop with a low water requirement, such as lettuce, limiting the amount of irrigation has little effect on the annual water consumption. The amount of water needed by lettuce fell below the allotted irrigation for much of the growing season.

Yield distributions for the crops under each irrigation strategy are shown in Figure 8. Results for each crop occupy a row, going from full to no irrigation left to right. For each crop, the variation in yield was smallest using a full irrigation strategy, regardless of whether a wet or dry forecast was used. Variability increased as irrigation amounts were reduced; greater variability can be seen in crops requiring more water (i.e., strawberries and alfalfa), especially with a dry forecast. As expected, average yields for each crop were larger in the no irrigation scenario with a wet forecast. The average yields for each crop were also relatively close as long as some irrigation strategy was used; significant reductions occurred when no supplemental water was provided, regardless of the forecast environment. The average predicted yield for strawberries with limited irrigation was larger for dry years than wet years, despite the extra stress put on the crops in dry years. As discussed later in Section 5.2, this was likely caused by the greater sensitivity of strawberry yield to net radiation; dry years tend to see more solar radiation and thus larger yields.
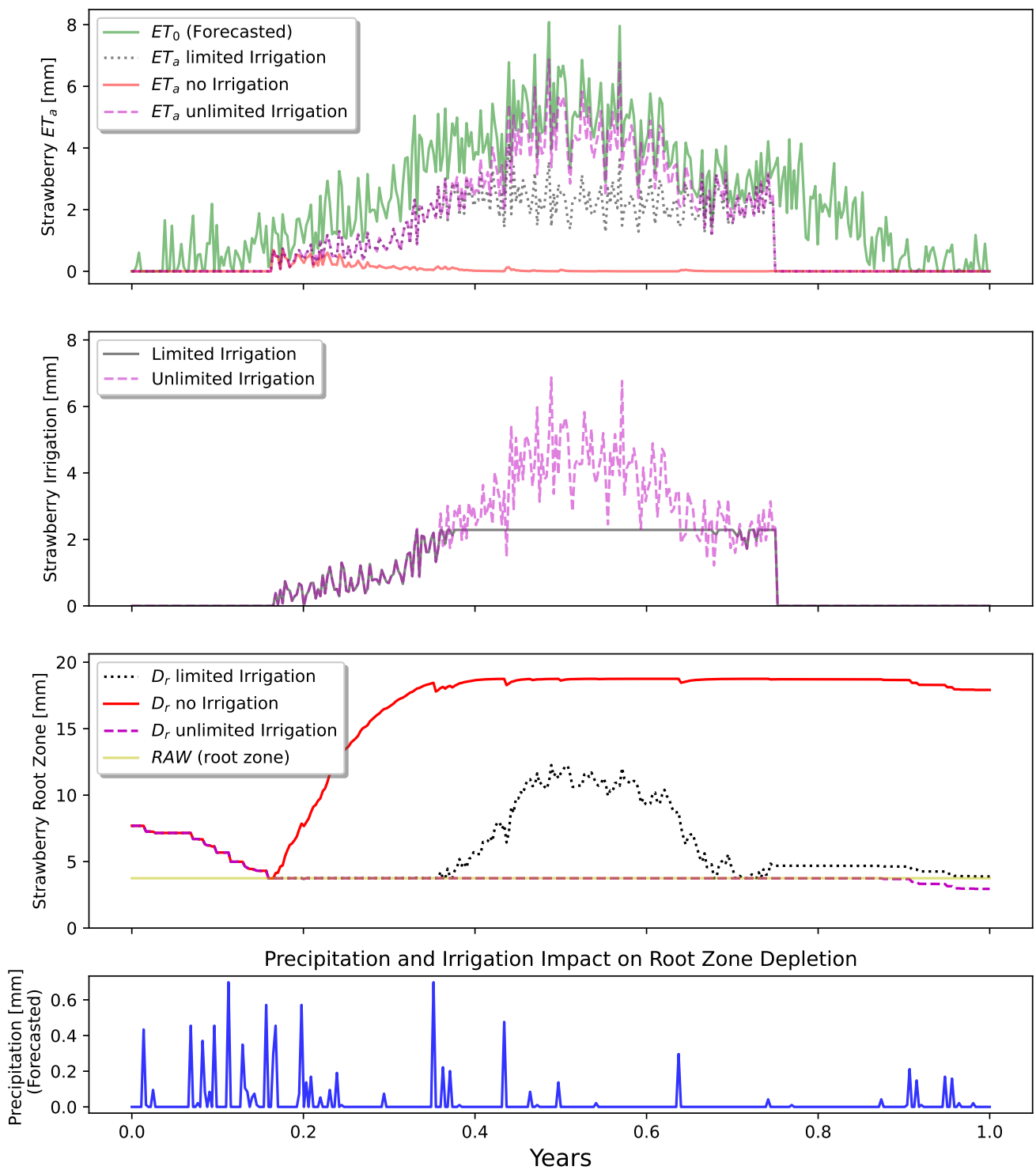
Figure 5. Depiction of root zone water use, irrigation, and *ET* for strawberries over a single growing season forecast. Root zone depletion is visible for dry periods when irrigation was limited, resulting in reduced crop growth.
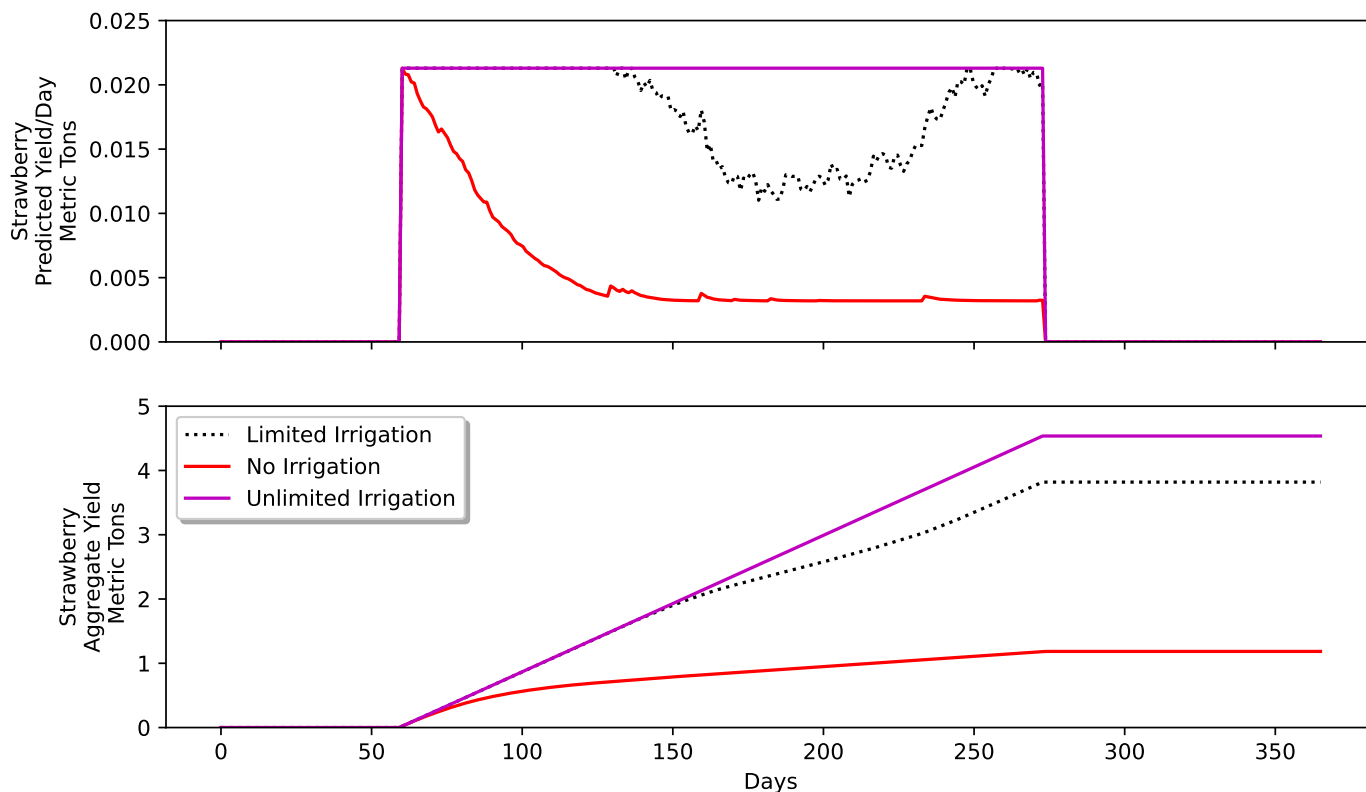
Figure 6. Strawberry yield rates for a single growing season with varying irrigation settings. As expected, crop yield diminishes when water needs are not met. Despite variability in environmental parameters, the constant yield rate in the unlimited irrigation scenario indicates that in our model, supplemental water is sufficient to ensure maximal yields. In our model, the yield rate (**top** plot) varies continuously in time and represents the amount of of biomass added to the crop that could be harvested on a a future date. The aggregate yield (**bottom** plot) is simply the cumulative sum of the yield rate and represents the potential size of a future harvest.

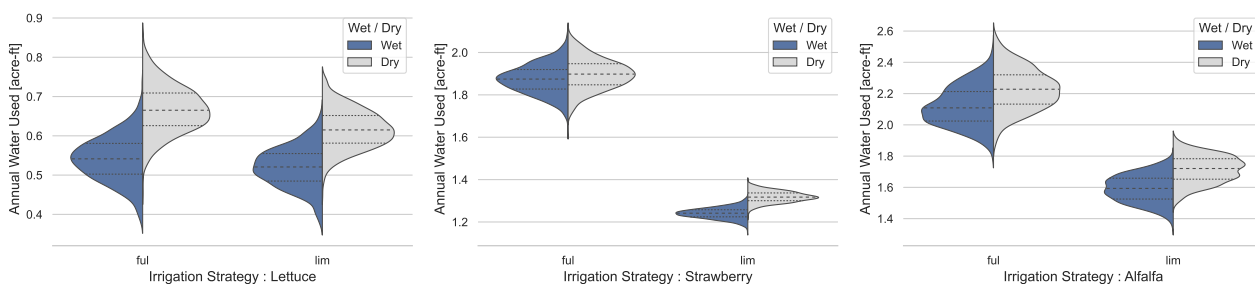**Figure 7.** Distributions of annual water use under full (distributions on the **left**) and limited irrigation (distributions on the **right**) strategies for lettuce, strawberries, and alfalfa. The annual water use for dry forecasts was larger, regardless of the crop or irrigation strategy. Water use by lettuce is the least affected by the irrigation strategy; strawberry yield is sensitive to limited irrigation.
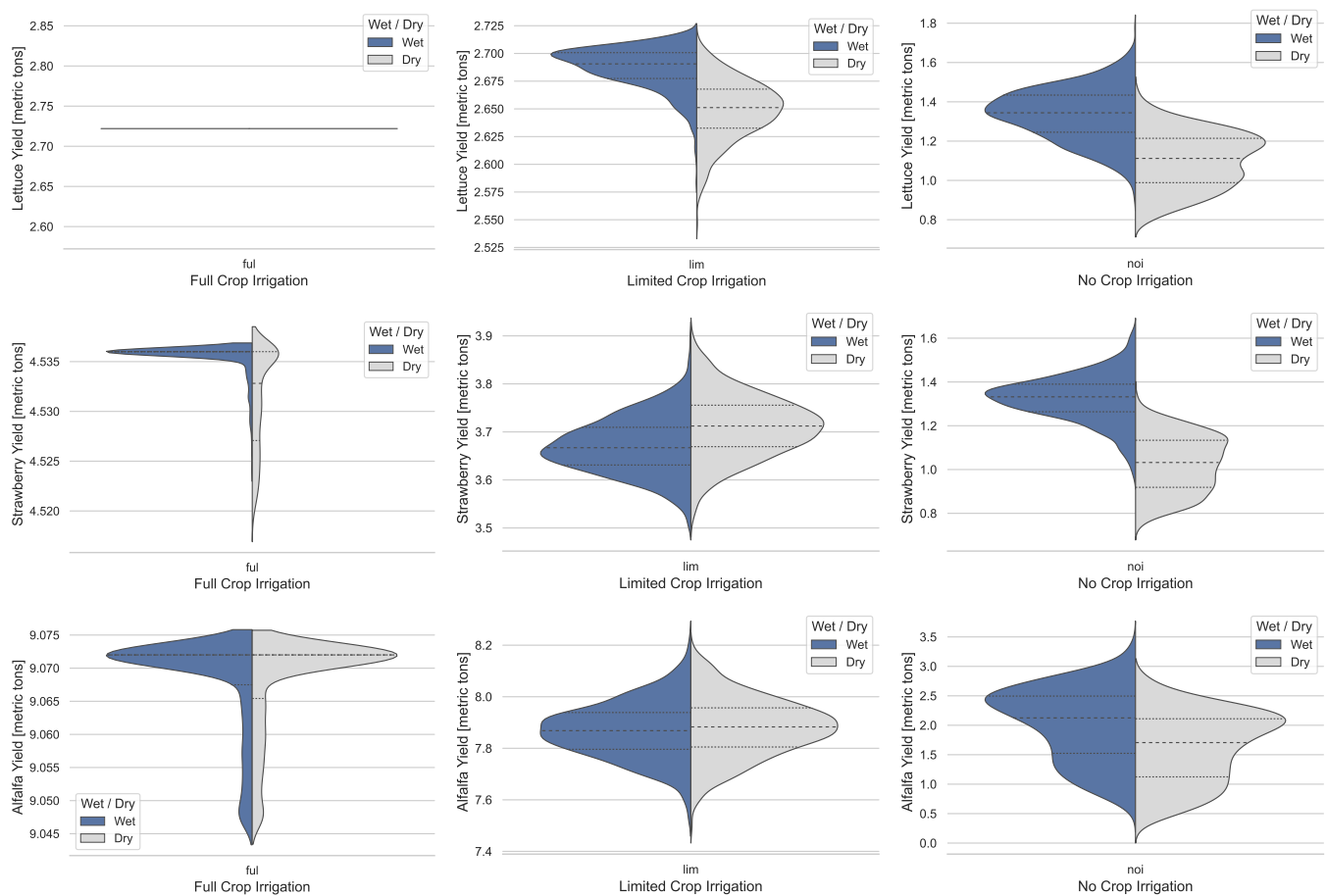
Figure 8. Distributions of annual yield (in metric tons) for alfalfa, strawberries, and lettuce under all three irrigation strategies: full (ful), limited (lim), and no irrigation (noi). Both wet and dry forecasts were used; yields were maximized under full irrigation. Average yields were cut in half when no irrigation was used.

4.3. Varying Limits on Water Use

We also considered the impact of the irrigation cap on the water usage and yield. These caps were modeled with a uniform distribution; water use samples were combined with wet and dry forecasts to generate information on water use and yield as the cap changed. Water caps ranged from 0 acre-feet per day to 0.025 acre-feet per day. For comparison, the results in the previous section used an irrigation limit of 0.0075 acre-feet per day.

The increase in Figure 9 shows the change in water use as the maximum additional water increases. As expected, with full irrigation, regardless of forecast environment, water use was maximized. The graph of annual water use for a limited irrigation strategy shows an increase until the irrigation cap is around 0.015 acre-feet per day. Once above this value, annual water use never exceeds more than 2 acre-feet. Water use in the dry environment exceeds that of the wet environment in both full and limited irrigation options.
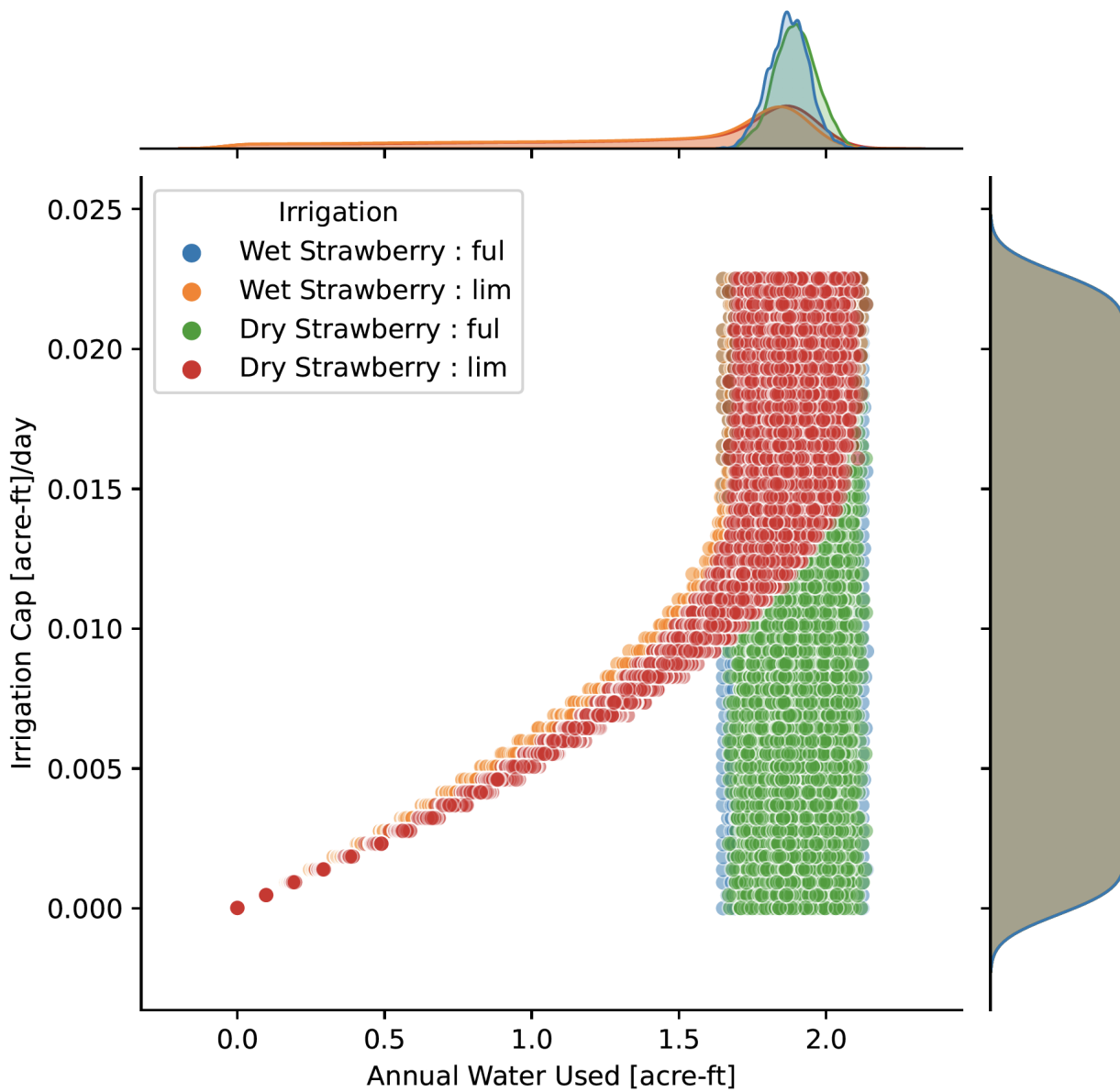
Figure 9. Annual irrigation used for strawberries with water limits drawn from a uniform distribution (on the **right**) along with sampled wet and dry forecast environments (along the **top**). Water use was maximized after a threshold water limit was reached.

The plot for sampled yield in Figure 10 tells a similar story. The caps on water use were again modeled with a uniform distribution, and strawberry yields were computed for samples of the water cap and wet and dry forecasts.

The water use cap affected strawberry yield when irrigation was limited, with yields increasing (nonlinearly) with the cap. However, once the cap exceeded 0.015 acre-feet per day, the yield became constant, indicating this is a threshold value for water use. In addition, when no irrigation was used, there were slightly higher annual yields in the wet environment as compared to the dry.
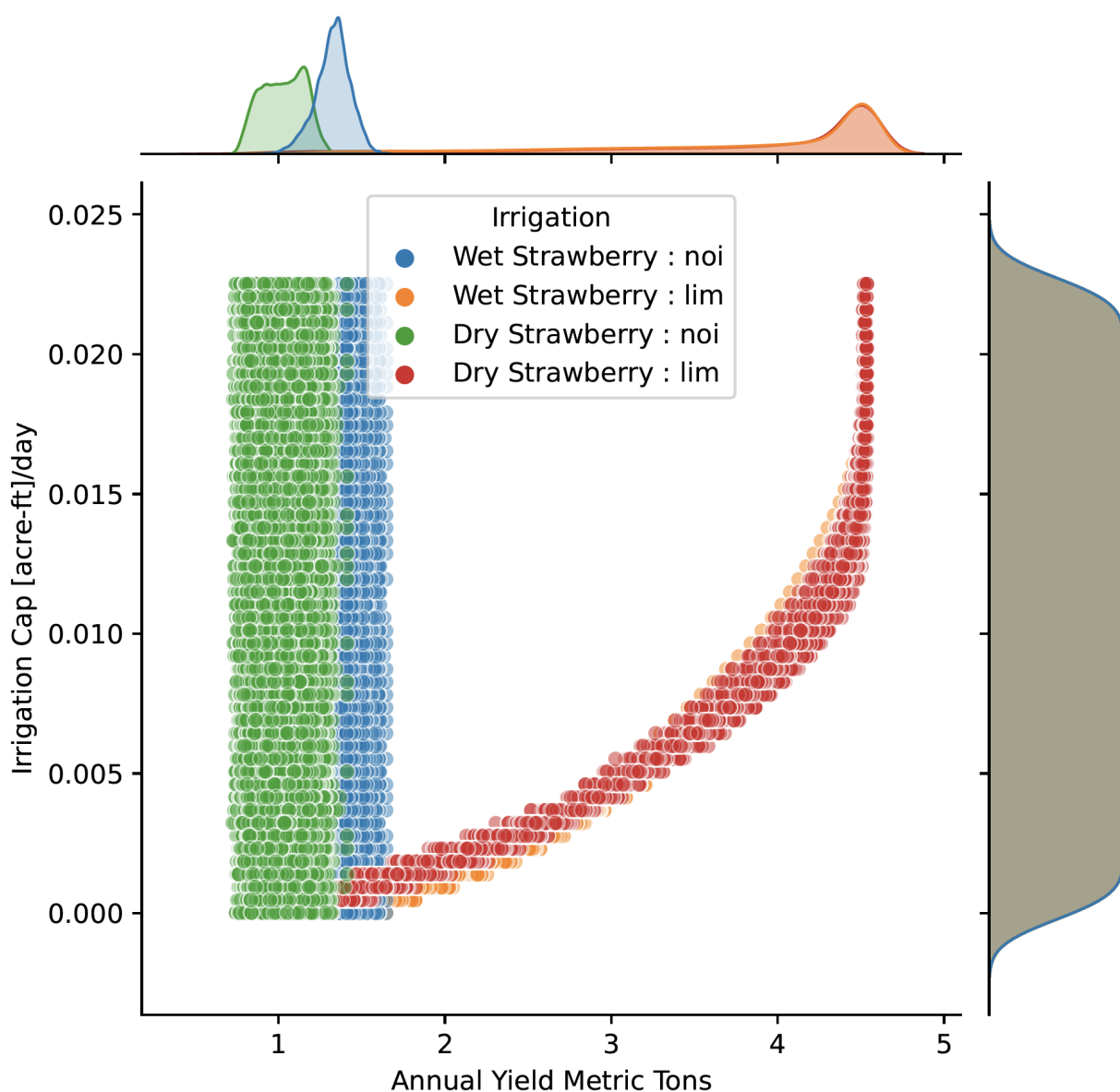
Figure 10. Sampled values of strawberry yield with water caps drawn from uniform distribution (on the **right**) and wet and dry forecasts (on the **top**). An irrigation cap 0f 0.015 acre-feet per day seems to be a threshold value.

5. Global Sensitivity Analysis

5.1. Background

The statistical models of the environmental parameters not only allow us to characterize the uncertainty in predictions of yield; they also enable us to identify the environmental processes that contribute most to the prediction uncertainty. In particular, we are able to compute two types of global sensitivity information: the main effects and the total effects. Main effects measure the fraction of predicted yield coming solely from a single parameter and do not capture higher order interactions between the parameters.

Let θ_i denote the ith vector-valued model parameter (e.g., daily precipitation) and let Y_j denote the predicted yield of crop j. The main effect of parameter θ_i on Y_j is given by

$$M_{ji} = \frac{\mathrm{Var}_{\theta_i}\left[\,\mathbb{E}_{\theta_{\sim i}}\left[Y_j|\theta_i\right]\,\right]}{\mathrm{Var}[Y_j]} \tag{22}$$

These expectations do not have closed-form expressions because Y_j depends on the nonlinear physics-based yield and hydrologic models described in Section 2. The expectations are also high dimensional because each parameter θ_i corresponds to a daily time series of one environmental parameter. To overcome these challenges, we employed one of the Monte Carlo approximations described in [26]. Let $A = \{\theta^{(a,1)}, \ldots, \theta^{(a,N)}\}$ and $B = \{\theta^{(b,1)}, \ldots, \theta^{(b,N)}\}$ be two sets of parameter samples, each with N independent samples. Let $\theta^{(ab_i,k)}$ denote the vector $[\theta_1^{(a,k)}, \ldots, \theta_{i-1}^{(a,k)}, \theta_i^{(b,k)}, \theta_{i+1}^{(a,k)}, \ldots, \theta_1^{(a,k)}]$, which is the same as the sample $\theta^{(a,k)}$ except that the ith component $\theta_i^{(a,k)}$ has been swapped with the corresponding $\theta_j^{(b,k)}$ from B.

Following [26], we approximate the main effect M_{ji} using the estimator

$$\widehat{M}_{ji} = \frac{1}{N\hat{\sigma}_j^2} \sum_{k=1}^{N} Y_j(\theta^{(b,k)}) \left(Y_j(\theta^{(ab_i,k)}) - Y_j(\theta^{(a,k)}) \right), \tag{23}$$

where $\hat{\sigma}_j \approx \mathrm{Var}[Y_j]$ is a Monte Carlo estimate of the variance of the yield for crop j using all $2N$ samples from both A and B. Mathematically, $\hat{\sigma}_j^2$ is given by

$$\frac{1}{2N-1} \sum_{k=1}^{N} \left(Y_j(\theta^{(a,k)}) - \hat{\mu}_j \right)^2 + \left(Y_j(\theta^{(b,k)}) - \hat{\mu}_j \right)^2, \tag{24}$$

where $\hat{\mu}_j$ is an estimate of the mean yield

$$\hat{\mu}_j = \frac{1}{2N} \sum_{k=1}^{N} Y_j(\theta^{(a,k)}) + Y_j(\theta^{(b,k)}). \tag{25}$$

While main effects only measure the first order impact of a parameter on the output variance, total effects include interactions to measure the total impact of each parameter on the prediction variance. Mathematically, the total effect of parameter θ_i is given by

$$T_{ji} = \frac{\mathbb{E}_{\theta_{\sim i}}\left[\mathrm{Var}_{\theta_i}\left[Y_j | \theta_{\sim i} \right] \right]}{\mathrm{Var}[Y_j]}. \tag{26}$$

We again employed a Monte Carlo estimate of T_{ji} from [26]; the total effects estimator is given by

$$\widehat{T}_{ji} = \frac{1}{2N\hat{\sigma}_j^2} \sum_{k=1}^{N} \left(Y_j(\theta^{(a,k)}) - Y_j(\theta^{(ab_i,k)}) \right)^2. \tag{27}$$

In this work we used independent random sampling in the estimators. More efficient quasi Monte Carlo (QMC) estimators are commonly employed in (23) and (27) (e.g., [26]), but these methods are not suitable for our high dimensional setting where θ has 1826 components.

Both the main effects and total effects can be interpreted as a fraction of the prediction variance $\mathrm{Var}[Y_j]$ stemming from one of the environmental variables θ_i, either without considering interactions or with interactions. They therefore quantify the relative impact that uncertainty in each of the environmental parameters has on uncertainty in the yield.

5.2. Sensitivity Results

To compute the global sensitivity indices M_{ji} and T_{ji} for each crop and parameter, we used $N = 10^4$ samples to construct the sets A and B. The process was repeated for both wet and dry scenarios to understand how the sensitivities change with environmental conditions. Our test used limited irrigation with an irrigation cap of 0.0075, which tends to slightly stress crops (see Figures 9 and 10).

For the six vector-valued parameters in our model, computing the sensitivity indices required a total of $2(2N + 6N) = 16 \times 10^4$ model evaluations. The main effects and total effects for alfafa, lettuce, and strawberries are shown in Tables 3 and 4. The strawberry sensitivities are also visualized in Figure 11.

Table 3. Main effects for each crop in both wet years and dry years. MaxIPerDay = 0.0075.

	Crop	Net Radiation	Temperature	Wind Speed	Pressure	Precipitation	Depletion
Wet	Alfalfa	5.32×10^{-01}	4.73×10^{-02}	0.00×10^{-00}	3.44×10^{-01}	4.81×10^{-02}	3.44×10^{-02}
	Lettuce	0.00×10^{-00}	4.08×10^{-02}	2.94×10^{-01}	0.00×10^{-00}	4.14×10^{-01}	6.62×10^{-03}
	Strawberries	5.29×10^{-01}	0.00×10^{-00}	0.00×10^{-00}	4.37×10^{-01}	0.00×10^{-00}	6.45×10^{-02}
Dry	Alfalfa	0.00×10^{-00}	7.24×10^{-02}	0.00×10^{-00}	9.17×10^{-01}	0.00×10^{-00}	1.36×10^{-01}
	Lettuce	0.00×10^{-00}	0.00×10^{-00}	0.00×10^{-00}	1.47×10^{-00}	0.00×10^{-00}	0.00×10^{-00}
	Strawberries	4.22×10^{-01}	2.02×10^{-01}	0.00×10^{-00}	3.86×10^{-01}	0.00×10^{-00}	6.42×10^{-02}

Table 4. Total effects for each crop in both wet and dry years. MaxIPerDay = 0.0075.

|	Crop	Net Radiation	Temperature	Wind Speed	Pressure	Precipitation	Depletion
Wet	Alfalfa	2.81×10^{-01}	3.86×10^{-01}	2.31×10^{-02}	1.90×10^{-01}	2.63×10^{-02}	1.18×10^{-01}
	Lettuce	3.60×10^{-01}	4.30×10^{-01}	3.19×10^{-02}	2.60×10^{-01}	6.14×10^{-02}	1.94×10^{-05}
	Strawberries	3.09×10^{-01}	4.56×10^{-01}	2.53×10^{-02}	2.00×10^{-01}	2.63×10^{-02}	2.79×10^{-03}
Dry	Alfalfa	6.10×10^{-01}	1.27×10^{-01}	1.37×10^{-02}	1.62×10^{-01}	1.10×10^{-02}	1.16×10^{-01}
	Lettuce	5.55×10^{-01}	2.28×10^{-01}	3.16×10^{-02}	2.38×10^{-01}	1.40×10^{-02}	7.26×10^{-07}
	Strawberries	6.71×10^{-01}	1.20×10^{-01}	1.16×10^{-02}	1.48×10^{-01}	1.35×10^{-02}	5.43×10^{-02}
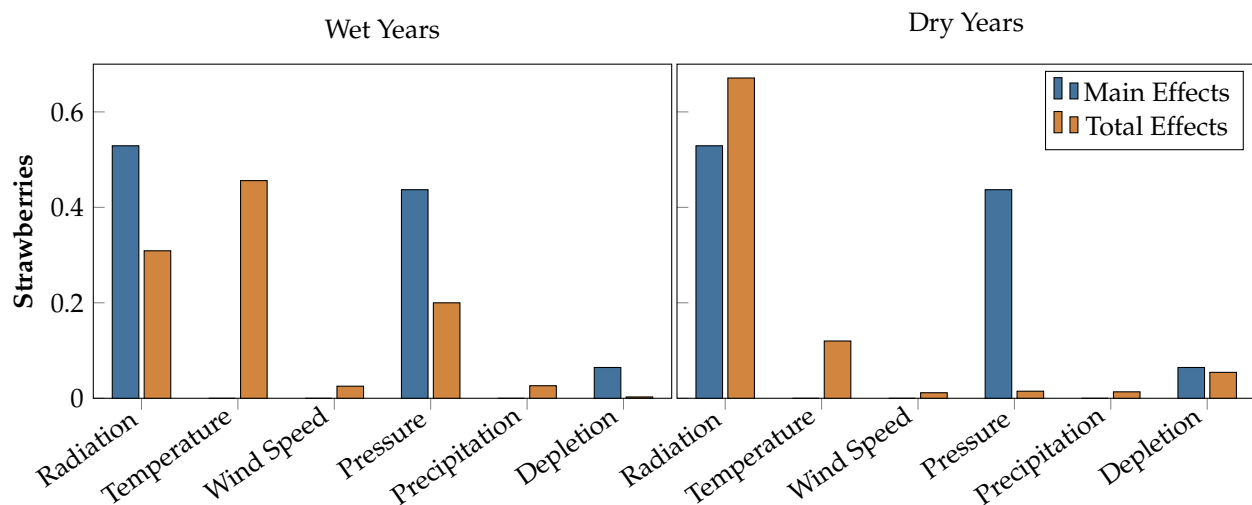
Figure 11. Comparison of global sensitivities for strawberries under wet and dry conditions. In dry years, the net radiation contributes more to uncertainty in yield predictions than it does in wet years. As shown in the total effects, the temperature has a smaller relative impact on the yield in dry years. Sensitivities for other crops can be found in Tables 3 and 4.

In both wet and dry conditions, uncertainty in net radiation caused a large fraction of the variance in yield: from Table 4, we see that the total effects in wet years were about 0.3 for all crops and about 0.6 for all crops in dry years. The total effects estimator also shows that uncertainty in air temperature had a large impact on yields during wet years ($\approx$0.4); the impact of temperature was less dramatic in dry years (0.12–0.23). It is also informative that in wet years the main effect of temperature was small ($\approx$0.04), though the total effect was relatively large ($\approx$0.4). This is an indication that the impact of temperature variability is predominantly through interactions with other environmental conditions.

From Tables 3 and 4, it is also interesting to note that the initial root zone depletion had a large impact on predictions of yield for Alfafa (total effect of ≈ 0.1), but less of an impact on the yields of lettuce and strawberries (total effect of < 0.06). This is an indication that the yield of Alfafa is more sensitive to initial root zone conditions than the other crops.

6. Conclusions and Future Directions

Our coupled statistical–physical framework outlines a strategy for incorporating meteorological uncertainty into physics-based agricultural yield models. Our approach avoids computationally expensive, high-fidelity simulations that require high levels of user expertise or purely data-driven approaches that do not incorporate known physical relationships. We combined historical data with well-established physical models of soil hydrology and crop growth to create an inexpensive computational engine for predicting yields which requires limited user input and expertise. While our physics-based model is less complex than state-of-the-art simulators such as MODFLOW-OWHM [27], the model is efficient and requires minimal preprocessing.

This approach also allows for uncertainty quantification and quantitative risk assessment. Numerous samples can be computed quickly, making it feasible to characterize predictive distributions and evaluation metrics for multiple parameters of interest. The computational efficiency is also critical for enabling global sensitivity analysis, which requires tens-of-thousands of model evaluations to obtain estimates of sensitivity indices. The global sensitivities reported in Section 5.2 provide context to the predictive distributions shown in Section 4.2; the sensitivity of strawberry yield to solar radiation indicates that, despite being more stressed, strawberries will tend to have larger yields in dry conditions because wet conditions are accompanied by less solar radiation.

The data needed for our framework can be obtained from a wide variety of weather stations throughout the country. Additional crop-specific parameters could also be obtained with local expertise available through many agricultural extensions and county farm agents. It could therefore be possible for our framework to be developed into a forecasting system for small farmers that, unlike large corporations, do not have the resources for extensive in-house data collection and analysis. Rigorously quantifying predictive uncertainties can also facilitate more formal risk analyses in the future. Various risk metrics, e.g., value at risk or conditional value at risk, require probabilistic predictions similar to those provided by our framework. Moreover, the flexibility to choose historical data representing different climatic scenarios (such as the wet and dry conditions discussed above), enables individualized exploration of worst- and best-case scenarios.

These ideas are not limited to the parameters studied here and can be extended to considering other distributions and other variables affecting farm management and harvests. Our future plans are to incorporate decision tools to help farmers use these analytics to estimate risk and manage other aspects of farming while weather patterns due to climate change continue to evolve.

Author Contributions: Conceptualization, J.C.C., E.W.J., K.R.K. and M.D.P.; methodology, J.C.C., E.W.J., K.R.K. and M.D.P.; software, J.C.C. and M.D.P.; validation, J.C.C. and M.D.P.; formal analysis, J.C.C., E.W.J., K.R.K. and M.D.P.; writing—original draft preparation, J.C.C., E.W.J., K.R.K. and M.D.P.; writing—review and editing, J.C.C., E.W.J., K.K., and M.D.P.; visualization, J.C, and M.D.P. All authors have read and agreed to the published version of the manuscript.

References

1. De Clercq, M.; Vats, A.; Biel, A. Agriculture 4.0: The future of farming technology. In Proceedings of the World Government Summit, Dubai, United Arab Emirates, 11–13 February 2018.
2. NOAA. Global Climate Indicators. Available online: https://www.climate.gov/news-features/understanding-climate/global-climate-indicators (accessed on 9 December 2021).

3. Gray, E. NASA at Your Table: Climate Change and Its Environmental Impacts on Crop Growth. *NASA Earth Science News*, 1 September 2021. Available online: https://www.nasa.gov/feature/goddard/esnt/2021/nasa-at-your-table-climate-change-and-its-environmental-impacts-on-crop-growth (accessed on 8 December 2021).

4. Tropp, D.; Moraghan, M. Local food demand in the U.S.: Evolution of the marketplace and future potential. In *Harvesting Opportunity: The Power of Regional Food System Investments to Transform Communities*; Dumont, A., Davis, D., Wascalus, J., Wilson, T., Barham, J., Tropp, D., Eds.; Federal Reserve Bank of St. Louis and the Board of Governors of the Federal Reserve System: St. Louis, MO, USA, 2017; pp. 15–42.

5. Tropp, D. From anecdote to formal evaluation: Reflections from more than two decades on the local food research trail at USDA. *J. Agric. Food Syst. Community Dev.* **2019**, *9*, 13–30. [CrossRef]

6. Robinson, J.; Farmer, J. *Selling Local: Why Local Food Movements Matter*; Indiana University Press: Bloomington, IN, USA, 2017.

7. Shelef, O.; Fernández-Bayoand, J.; Sher, Y.; Ancona, V.; Slinn, H.; Achmon, Y. Elucidating local food production to identify the principles and challenges of sustainable agriculture. In *Sustainable Food Systems from Agriculture to Industry*; Elsevier: London, UK, 2018; pp. 47–81.

8. Kriewald, S.; Pradhan, P.; Costa, L.; Ros, A.G.C.; Kropp, J.P. Hungry cities: How local food self-sufficiency relates to climate change, diets, and urbanisation. *Environ. Res. Lett.* **2019**, *14*, 094007. [CrossRef]

9. Thilmany, D.; Canales, E.; Low, S.; Boys, K. Local Food Supply Chain Dynamics and Resilience during COVID-19. *Appl. Econ. Perspect. Policy* **2021**, *43*, 86–104. [CrossRef]

10. Udias, A.; Pastori, M.; Dondeynaz, C.; Carmona Moreno, C.; Abdou, A.; Cattaneo, L.; Cano, J. A decision support tool to enhance agricultural growth in the Mékrou river basin (West Africa). *Comput. Electron. Agric.* **2018**, *154*, 467–481. [CrossRef]

11. Zhai, Z.; Martínez, J.F.; Beltran, V.; Martínez, N. Decision support systems for agriculture 4.0: Survey and challenges. *Comput. Electron. Agric.* **2020**, *170*, 105256. [CrossRef]

12. Bronson, K. The Digital Divide and How it Matters for Canadian Food System Equity. *Can. J. Commun.* **2019**, *44*, 63–68. [CrossRef]

13. Chrispell, J.; Fowler, K.; Howington, S.; Jenkins, E.; Minik, M.; Sendova, T. Mathematical modeling, simulation, and optimal design for agricultural water management. In Proceedings of the 2012 SC Water Resources Conference, Columbia, SC, USA, 10–11 October 2012.

14. Bokhiria, J.; Fowler, K.; Jenkins, E. Modelling and optimization for crop portfolio management under limited irrigation strategies. *J. Agric. Environ. Sci.* **2014**, *3*, 209–237.

15. Fowler, K.; Jenkins, E.; Ostrove, C.; Chrispell, J.; Farthing, M.; Parno, M. A decision making framework with MODFLOW-FMP2 via optimization: Determining trade-offs in crop selection. *Environ. Model. Softw.* **2015**, *69*, 280–291. [CrossRef]

16. Fowler, K.; Jenkins, E.; Parno, M.; Chrispell, J.; Colón, A.; Hanson, R. Development and use of mathematical models and software frameworks for integrated analysis of agricultural systems and associated water use impact. *AIMS Agric. Food* **2016**, *1*, 208–226. [CrossRef]

17. Steduto, P.; Hsiao, T.; Fereres, E.; Raes, D. *Crop Yield Response to Water. FAO Irrigation and Drainage*; Technical Report 66; Food and Agriculture Organization of the United Nations: Rome, Italy, 2012.

18. Allen, R.; Pereira, L.; Raes, D.; Smith, M. *Crop Evapotranspiration Guidelines for Computing Crop Water Requirements FAO 56*; Technical Report 56; Food and Agriculture Organization of the United Nations: Rome, Italy, 1998.

19. Cornell University Cooperative Extension. (Fact Sheet 29) Agronomy Fact Sheet Series: Soil Texture. Available online: http://nmsp.cals.cornell.edu/publications/factsheets/factsheet29.pdf (accessed on 9 December 2021).

20. California Department of Water Resources. California Irrigation Management Information Systems. Available online: https://cimis.water.ca.gov/ (accessed on 9 December 2021).

21. GPy. GPy: A Gaussian Process Framework in Python. 2012. Available online: http://github.com/SheffieldML/GPy (accessed on 9 December 2021).

22. Durrande, N.; Hensman, J.; Rattray, M.; Lawrence, N.D. Gaussian process models for periodicity detection. *PeerJ Comput. Sci.* **2016**, *2*, 1–18. [CrossRef]

23. Williams, C.K.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006.

24. Kleiber, W.; Katz, R.W.; Rajagopalan, B. Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. *Water Resour. Res.* **2012**, *48*, W01523. [CrossRef]

25. U.S. Government. National Integrated Drought Information System. Available online: https://www.drought.gov/states/california/county/butte (accessed on 9 December 2021).

26. Saltelli, A.; Annoni, P.; Azzini, I.; Campolongo, F.; Ratto, M.; Tarantola, S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput. Phys. Commun.* **2010**, *181*, 259–270. [CrossRef]

27. Hanson, R.T.; Boyce, S.E.; Schmid, W.; Hughes, J.D.; Mehl, S.W.; Leake, S.A.; Maddock, T., III; Niswonger, R.G. *One-Water Hydrologic Flow Model (MODFLOW-OWHM)*; Technical Report; US Geological Survey: San Diego, CA, USA, 2014.