Nannochloropsis sp. algae for use as biofuel: Analyzing a translog production function using data from multiple sites in the southwestern United States.

Archambault, Steven[A]*
Email: sarchamb@nmsu.edu

Starbuck Downes, Cara Meghan[B,C]
Email: cdownes@nmsu.edu

Van Voorhies, Wayne[C]
Email: wavanvoo@nmsu.edu

Erickson, Christopher A.[B]
Email: chrerick@nmsu.edu

Lammers, Peter[C]
Email: plammers@nmsu.edu

[A]Agricultural Economics and Agricultural Business
New Mexico State University
MSC 3169 NMSU, P.O. Box 30003
Las Cruces, NM 88003-8003
*Primary Contact

[B]Economics, Applied Statistics, and International Business
New Mexico State University
MSC 3CQ
PO Box 30001
Las Cruces, NM 88003-8001

[C]Algal Bioenergy Program
New Mexico State University
Laboratory, Box 30001, MSC 3RES
Las Cruces, NM 88003-8001

Abstract

This paper investigates the production of *Nannochloropsis sp.* algae at five different sites located in the southwestern region of the United States. Studies of the economic viability of algae production typically calculate the Capital and Operating Expenses of stylized algal production firms with minimal understanding of the linkages between production and input variables that drive the costs being estimated. These results work towards filling this gap by estimating several production functions using real world data. Our dataset includes 10,316 days of algae growth, from which we generate 495 growth period observations. Particularly, the study analyzes the relationship between variation in input factors over a growth period and the resulting alage production measured by ash free dry weight. We carry out several multivariate econometric regression analyses. The variables photosynthethcially active radiation (PAR), length of growth periods, and the growth of *Nannochloropsis salina* result in increased algae production. Algae production at the Texas AgriLife at Texas A&M University in Pecos, Texas, and Flour Bluff, Texas, resulted in higher algae production than the three sites in New Mexico. Increases in the initial algae inoculation levels and average precipitation consistently indicated a negative relationship with algae production in our model. These results should be useful for further studies aiming to connect real world algae production decisions with measures of costs and profitability.

Keywords: algae, biofuel, ***Nannochloropsis*** *sp.*, econometrics, production function, translog

1. Introduction

1.1 Microalgae suitability for bioenergy

Considerable interest has been expressed in policy circles regarding the potential of microalgae biofuels as an alternative source of clean energy [1]. Microalgae are diverse unicellular microorganisms that can convert sunlight and CO_2 into carbohydrates, protein, and natural oils, using photosynthesis [2]. As much as 75% of body weight in some species is made up of natural oils [1, 3, 4]. These oils can be processed into numerous products through transesterification [5], hydrothermal liquefaction [6, 7], or gasification [8]. Microalgae lipids have been upgraded to jet fuel, diesel fuel, gasoline, green diesel, or biodiesel through many of the same processes used to convert petroleum crude into finished fuel products [9] [10]. These products have the advantage, in contrast to ethanol, of being energy dense fuels that are compatible with existing energy infrastructure [11]. Algal based biofuels have the potential to be produced with a smaller carbon footprint than traditional fuels and can be produced with water, land, and nutrient inputs that do not compete with food production, unlike other feedstocks, such as corn, sorghum, and sugarcane [12]. Algae also have a much faster rate of growth and smaller land footprint due to the increased photosynthetic efficiency relative to land crops [13].

The first generation of biofuel production focused on *Nannochloropsis salina*, which are a coldwater marine species [14, 15] shown to be tolerant of brackish water [16] and suitable for CO_2 fixation [16]. *Nannochloropsis* are also high in triglycerides and have a relatively high growth rate. Thus, this species was thought to be a good candidate for use as a biofuel species. While continued research has found additional species that are more viable for production scale, much has been learned from the initial cultivation experience with *Nannochloropsis* [11]. It has been used as the base organism in many of the Life Cycle Assessments and first generation techno-economic models, and many of the growth and nutrient predictions for greenhouse gas and land use change calculations have been done using *Nannochloropsis*. [2] [13] [17] [18] [19] [20] [21]. Many algae cultivation studies have used techno-economic assessment (TEA) to analyze the potential economic viability of algae production and to calculate the Capital and Operating Expenses (CAPEX and OPEX) of stylized algal production firms [11, 22, 23, 24, 25, 26, 27, 28], with minimal understanding of the linkages between production and input variables that drive the costs being estimated. This research works towards bridging this gap with an applied algae production analysis that estimates the relationships between a selection of critical environmental and control variables and the impact on biomass production using 10,316 days of outdoor *Nannochloropsis* production data from five sites in the southwestern United States. Using econometric analysis, production functions are estimated, allowing for the examination of the role of various environmental and control inputs in the production of algae. Both Cobb-Douglas and translog functional forms of production are estimated. The research provides a systematic analysis of the relationship between biomass productivity and the explanatory variables of temperature, PAR, production cycle length, and initial inoculation, using real world data. The methodology can identify inputs that are over- and under-utilized. The results allow simulation of the impact from changes to the quantity of algae production input variables, and provide a comprehensive analysis of microalgae production data. The results should be useful for the development of additional models concerned with financial and environmental viability of algal fuel production.

1.2 Production and economic efficiency

Understanding the relationship between inputs and outputs is a critical step in accurately determining economic feasibility, and more importantly, can be used to direct research and development toward reducing costs and increasing output in order to increase economic viability of the use of algae as a biofuel [29]. Any given production process can be represented by a production function:

$$Y = f(X) \tag{1}$$

Equation (1) gives the combination of inputs (*X)* and outputs (*Y*) that are technologically feasible at a specified point in time, and allows the flow of inputs and outputs for a given time period to be tracked through a production system or process (see, e.g., [30, 31, 32]). An applied production analysis focuses on defining the elements and relationships in $Y = f(X)$ such that profit can be estimated and sensitivity analyses for the various production inputs can be investigated [33, pp. 54-75].

To further understand $Y = f(X)$, it is useful to divide this input vector into three categories. First are elements of X that are under the operational control of management and can be varied in the short-run. The second category includes capital inputs that are under the control of management, but can only be varied in the long run, between growing cycles or when longer-term management strategies are being considered. Third are environmental factors that are important for the production process but are not under the direct control of management. These environmental variables are stochastic in nature. While management does not directly control these environmental variables, many of the capital and operating expenses incurred will be related to mitigating the adverse impact of these environmental stochastic variables on production. Thus, stochastic non-control variables enter into the choice set of the firm through decisions regarding the use of capital and operating systems and processes. Thus, the production function can be represented as follows:

$$Y = f(o, \kappa, \varepsilon) \tag{2}$$

where $\mathbf{o}$ is a vector of inputs under operational control that can be varied in the short run, κ is a vector of capital inputs that are fixed in the short run, and ε contains stochastic environmental variables not under the direct control of management. Equation (2) captures the basic elements of algae lipid production, which can be used to derive the revenues, costs, and profit or loss of the firm. More directly, the stylized production function captures the production based variables and their interdependencies.

The conceptual framework defined by Equation (2) needs to be translated into a functional analysis. Typically TEAs do this by using mathematical equations to populate a spreadsheet with the economic and financial metrics of interest. Parameters for these equations are typically derived using lab bench experiments or other prototypes. Often, idealized operation is assumed. An alternative procedure, which is pursued in this paper, is to estimate a production function from actual data generated from experiments. In particular, a production function for *Nannochloropsis sp.* is estimated using a panel data set created by pooling data from five experimental production facilities [34].

2. Material and Methods

2.1 Description of Data

The authors use 10,316 days of algae growth from five sites located in the southwestern United States collected from 2009-2012. From this sample, 495 growth period observations were generated. Data was collected from the following sites and partners: (1) Sapphire Energy in Las Cruces, NM (SAP); (2) New Mexico State University Energy Research Laboratory, in Las Cruces, NM (NMS); (3) Center for Excellence in Hazardous Materials Management in Atoka, NM (CHM); (4) Texas A&M AgriLife Extension in Pecos, Texas (PEC); and (5) Texas A&M AgriLife Extension in Flour Bluff, Texas, near Corpus Christi, Texas (COR). The cultivation data was collected over a four year period in outdoor reactors similar to traditional Oswald raceways. Cultivation volume was from 1,000 liters to 100,000 liters and more than 50% of the observations are drawn from cultivation volumes in excess of 25,000 liters.

Table 1 provides descriptive statistics for the variables included in our study. AFDW is a uniform measure of organic content that eliminates the variability that may arise from samples with differing water content or ash content [35]. In many instances, including the measuring of initial values that were non-zero, AFDW was extrapolated from a recorded value of AFDW density (g/l). For other cases, optical density at 750 nm (OD750) was used to determine AFDW [35]. For the latter case, an observed relationship between OD750 and AFDW was determined via an ordinary least squares regression analysis for each site. From this analysis, the AFDW values are determined.

The growth periods were a number of days of growth, which began with an initial measurement of AFDW, and ended with a final measurement of AFDW. The final measurement of AFDW was recorded from a measurement of harvested biomass, a final reading of AFDW density in the pond, or from a combination of the two. In some growth periods, for example with the PEC site, biomass was not harvested, yet the batch was moved to a different pond, diluted, and a new growth period began. In the case of CHM, and in some of the SAP growth periods, biomass was partially harvested, then growth was allowed to continue. The day of harvesting, or the last day of consecutive days of harvesting if harvest occurs over multiple days, is considered the final day of a growth period. For each growth period in which biomass was harvested throughout the growth period, the harvested quantity was added to the final growth quantity. The following equation summarizes the AFDW calculation:

$$\textbf{AFDW = Ending biomass - Initial biomass + Harvested biomass} \tag{3}$$

The average daily-integrated photosynthethcially active radiation (PAR) over the growth period is taken from data collected in three-minute intervals by Colorado State University (CSU) [36]. Several sites did collect PAR onsite, but the CSU data set provides a uniform methodology to collect PAR. The CSU PAR sensors closest to the production site were used [27, 37, 38].[1] The use of CSU PAR sites introduces measurement error, but researchers felt that PAR is a critical variable and that this proxy measure was preferable to excluding PAR as a production variable. At the beginning of each growth period, the initial density of algae (INI) is measured as AFDW (g/l). A nonlinear relationship between INI and AFDW was hypothesized. A zero value of INI would result in no growth, as there would be no parent algae. On the other hand, a high value of INI would result in excessive competition for nutrients as well as self-shading. Growth periods varied in length over time at individual sites, and also across different sites. The number of days in each growth period (DAY) was included to control for growth period variation. It was expected that very short growth periods, and very long growth periods, would result in lower overall per day productivity, providing a non-linear relationship between productivity and DAY.[2] The average range in daily ambient air temperature over the growth period by site (TEM) is a proxy for water temperature fluctuation. Ideally, direct measures of water temperature would be used [38], but this data was not measured consistently at each of the sites. Air temperature is an acceptable proxy, as no site in the study mechanically controlled water temperature. Average participation per day during the growth period (PRE) is included to account for storm events, which are associated with the invasive species events.

A number of dummy variables are included in the analysis. First among these is NAN, which is a dummy variable indicating that the species is *Nannochloropsis salina*. All of the observations that were not *Nannochloropsis salina* were from the genus *Nannochloropsis*, but included various strains other than *N. salina* such as *Nannochloropsis occulata*. In some instances, the strain was not identified.

[1] The NMS site was 38 km from the PAR sensor, located at the Jornada long-term agricultural research site near Las Cruces, New Mexico. This sensor also provided data for SAP (43 km distance) and CHM (221 km distance). The PAR sensor in Seguin, Texas, provided the COR PAR data (227 km distance). The PEC PAR observations were taken from the PAR sensor in Big Bend, Texas (253 km distance).
[2] Seven observations with fewer than two days in the growth period were eliminated as being two short a time period to be considered full growth cycles. Two additional observations of 595 and 600 days were eliminated because they were considered unrealistic growth scenarios.

Dummy variables for location were also included in the analysis.[3] Location dummies are expected to have a significant effect on production stemming from geographical location, from physical design of ponds and raceways [39], and from systematic differences in production techniques across sites.

Table 1. Descriptive statistics.

Variable	Units	Description	Obs	Mean	SD	Min	Max	CV
AFDW	g/m^2	Ash free dry weight generated over growth period per area	495	77.6	67.2	-61.0[a]	353.6	0.866
PAR	$\mu mol/(m^2\ sec)$	Average daily integrated PAR over the growth period (in thousands)	495	36277.1	12033.1	14919.9	60129.6	0.332
INI	g/l	Initial ash free dry weight density for growth period	495	0.31	0.24	0.02	1.00	0.778
DAY	#	Number of days in the growth period	495	20.8	20.4	3.0	146.0	0.980
TEM	F	Average range of daily ambient air temperature fluctuation over the growth period	495	21.7	8.7	7.3	41.0	0.400
PRE	in/d	Average precipitation per day over the growth period	495	0.02	0.04	0.00	0.56	2.618
NAN	dummy	Dummy variable indicating algae species as *Nannochloropsis salina*	495	0.72	0.45	0	1	0.622
SAP	dummy	Dummy variable indicating growth at Sapphire Energy in Las Cruces, New Mexico	495	0.09	0.28	0	1	3.245
PEC	dummy	Dummy variable indicating growth at Texas AgriLife at Texas A&M University in Pecos, Texas.	495	0.17	0.37	0	1	2.230
COR	dummy	Dummy variable indicating growth at Texas AgriLife at Texas A&M University in Flour Bluff, Texas, near Corpus Christi, Texas.	495	0.48	0.50	0	1	1.040
CHM	dummy	Dummy variable indicating growth at the Center for Excellence in Hazardous	495	0.12	0.33	0	1	2.670

[3]The dummy variable takes on the value 1 when the data is from the indicated location, and is zero otherwise.

		Materials Management in Atoka, NM.						
NMS	dummy	Dummy variable indicating growth at New Mexico State University Energy Research Laboratory, in Las Cruces, NM.	495	0.14	0.35	0	1	2.467

[a]Growth was negative for some observations, arising from pond crashes in which a significant portion of the algae died prior to harvest.

Daily productivity at each site is provided in Figure 1, measured as ash free dry weight (AFDW) per day (g/m^2/d), by site and overall. The PEC site had the highest average productivity, but also the most variation. CHM was least productive while NMS had the least variation in output. Daily AFDW varies from an average of 0.803 g/m^2/d in CHM to an average of 8.513 g/m^2/d in relatively nearby PEC.[4]

Figure 1 Box plot of daily algae production by site, and overall production for all sites.

2.2 Data relationships

Figure 2 displays scatter diagrams plotting the natural log of algae production as measured by average ash free dry weight generated over the growth period (ln AFDW) to the natural log of the various potential determinates, with different determinants displayed in each of the panels. Also included in each panel is a fitted value determined using ordinary least squares. Logarithms were used to account for potential nonlinearity in the data. One difficulty with this approach is that some observations for growth were negative, arising from pond crashes in which a significant portion of the algae died prior to harvest. Values less than or equal to zero cannot be transformed into natural log form. A common solution is to add a factor to all observations of a variable that sufficiently brings all values above zero. Doing so does not change the relationship between the dependent and independent variables. [40]. Following this approach, 61 was added to each AFDW observation.

[4] The growth period data at the CHM site was not clearly delineated, as the growth was carried out in ongoing pond growth periods spanning multiple years. See discussion below.

Similarly a one was added to the independent variables INI and PRE, to eliminate values less than zero, and negative log values. Panel A in Figure 2 relates ln AFDW to the natural log of average PAR over the growth period (ln PAR). A positive relationship is expected [41]. In fact, a weak negative relationship is observed. Panel B shows algae production increases with days over which growth occurs (ln DAY). It is expected that over longer grow periods, production will remain positive, but the growth rate will begin to decline due to self-shading [42]. Panel C shows the relationship between ln AFDW and the natural log of initial density (ln INI). A negative relationship is observed indicating over inoculation may be occurring [42]. Panel D shows the relationship of the natural log of the mean daily range in ambient air temperature (ln TEM) to be negatively related to algae production [41, 42]. A constant, controlled temperature appears to promote growth. In Panel E, it is apparent that the natural log rainfall during the growth period (ln PRE) is associated with declining algae production. This is likely due to storms causing pond crashes as wind and rain can contaminate open ponds.

Figure 2 Log-log relationship between algae production and the determinants of algae production. Panel A illustrates a positive relationship between ln AFDW and ln PAR. Panel B illustrates a positive relationship between ln AFDW and ln DAY. Panel C illustrates a negative relationship between ln AFDW and ln INI. Panel D illustrates a negative relationship between ln TEM and ln AFDW. Panel E illustrates a negative relationship between ln PRE and ln AFDW.

3. Econometric Modeling

The two-way correlation in Figure 2 provides an indication of the relationship between algae growth and production factors. However, multivariable regression analysis permits examining the role of the various factors simultaneously in influencing production. In this section, econometric methodology is laid out in full.

The production function for *Nannochloropsis sp.* can be represented by $Y_{it} = f(X_{it1}, X_{it2}, \ldots X_{itM}; \eta_i, \upsilon_{it})$, where i = 1,2,..,5 is an index of locations, t is a time index, Y_{it} is output at time t for location i, X_{itm} are factors that affect the algae production also indexed for time and location, η_i is an unobservable site-specific effect, and v_{it} is a random component. In what follows, F(·) is assumed to be approximated as log-linear. The natural logarithm of $Y_{it,}$ and X_{itm} are denoted by q_{it} and x_{itm}, respectively. The specific form of the production equation can be approximated as a log-linear function defined as follows.

$$y_{it} = \alpha_0 + \sum_{m=1}^{M} \alpha_m x_{itm} + \eta_i + \upsilon_{it} \qquad (4)$$

This is the Cobb-Douglas production function, which is frequently used in economics, as it illustrates with ease the trade-off between input variables in order to achieve production output. It has been shown to appropriately estimate a wide variety of production relationships [30] [33] [34]. The term α_m is the production elasticity for the input x_{itm} and M is the number of inputs. Thus, given our specification, a 1 percent increase in x_{itm}causes an $\alpha_m$percent increase in y_{it}. Equation (4) is estimated using an unbalanced pooled data[5] with three different techniques—ordinary least squares (OLS), ordinary least squares with fixed effects (OLS-FE), and instrumental variables (IV) [34].

[5] The data is pooled in the sense that data from all five sites are used to estimate the regressions. The data is unbalanced in the sense that there are a different number of observations for different sites and the observations may not correspond to each other in time.

Table 2. Cobb-Douglas Production Function.

Dep. Variable: ln AFDW	Model 1 OLS		Model 2 FE-OLS		Model 3 IV	
Dependent Variables[a]	Coefficient	Robust S.E.	Coefficient	Robust S.E.	Coefficient	Robust S.E.
CON	3.841***	(0.927)	1.964	(1.458)	1.078	(1.440)
ln PAR	0.157**	(0.074)	0.277***	(0.092)	0.343***	(0.094)
ln INI	-1.057***	(0.331)	-0.163	(0.231)	0.090	(0.306)
ln DAY	0.056	(0.069)	0.159***	(0.053)	0.361**	(0.144)
ln TEM	-0.242***	(0.062)	0.090	(0.099)	0.051	(0.100)
ln PRE	-0.660*	(0.370)	-1.004***	(0.380)	-1.362**	(0.555)
NANO	0.029	(0.041)	0.105***	(0.037)	0.106***	(0.040)
SAP			-0.293***	(0.094)	-0.502*	(0.217)
COR			0.326**	(0.137)	0.154	(0.191)
CHM			-1.295***	(0.261)	-1.521***	(0.409)
NMS			-0.248***	(0.064)	-0.441***	(0.152)
N	495		495		495	
Std. Dev. of the Residuals	0.41		0.36		0.38	
R^2	0.35		0.51		0.47	
Adj R^2	0.34		0.50		0.45	
AIC[b]	538.9		406.2		449.3	
F[c]	54.3***		60.3***		45.6***	
Kleib-Paap LM[d]					19.86***	
Kleib-Paap F[be]					16.14[i]	
Hansen J (X^2)[f]					3.41	
Endog (X^2)[g]					0.400	

[a] PAR is daily-integrated photosynthethcially active radiation, INI is the initial concentration of algae at the time production is commenced, DAY is the number of days over which production occurred, TEM is the average daily variation in temperature, PRE is average daily precipitation, and NANO indicates that the species cultivated is *Nannochloropsis salina.* and zero otherwise.

[b]AIC: Goodness-of-fit measure considering the trade-offs between accuracy and complexity. A lower value indicates a preferred model.

[c]F-test: Statistic examining the significance of the explanatory variables, as a group, in the model. The null hypothesis is that the variable groups are not significant. The results reject the null at the 1% level in each model.

[d]Kleib-Paap LM test: Under identification (test t, with the null hypothesis that instruments are not independent, therefore, invalid. This indicated that the instruments used are appropriate.

[e]Kleib-Paap F: Weak identification test of instruments. [i] indicates test stat exceeds the critical value of 5% relative bias and 15% maximal IV size distortion [43].

[f]Hansen J: Over identification test, with the null that instruments are over identified and valid.

[g]Endog (chi-sq): Tests exogeneity of the questioned explanatory variable, with the null hypothesis that

the variable is exogenous. The null is not rejected.

Table 2 presents results using the Cobb-Douglas specification given in equation (4). For each model, the natural log of AFDW is the dependent variable and included are six explanatory variables—the natural log of PAR, INI, DAY, TEM, and PRE, and the dummy variable NANO. Time effects are controlled for using dummy variables for each year. The Cobb-Douglas model relates the inputs to the output in such a way that the coefficients can interpreted as elasticities. For example, a one-percent increase in TEM will cause a -0.242% change in production. Model 2 differs from Model 1 by adding location dummies. Comparing the two models, the inclusion of location dummy variables improves measures of goodness of fit, indicating that Model 2 is preferred. The significance of ln INI and ln TEM drops out in the FE model, but NANO gains significance. The coefficient of ln TEM, a measure of temperature flux may be anticipated to have a negative sign, as it does in Model 1, but is not significant in Model 2. The adjusted R^2 indicates that Model 2(OLS-FE), which includes location fixed effects, performs better than Model 1. The OLS-FE model captures the systematic differences between sites including weather, managerial skill, and physical facilities.

Model 3 is the same as model Model 2 accept in using the estimation technique of instrumental variables to account for potential endogeneity of DAY. In particular, managers may change inputs under their control so as to mitigate random fluctuations in production, thus, potentially creating a feedback loop between the regressors and the error term. In the context of the current setting, ln DAY, which is under the control of management, could be endogeneous as managers could vary the length of the production cycle to offset other factors. To test for endogeneity, Model 3 is estimated using instrumental variable (IV) for ln DAY. This requires choosing instrument variables that are correlated with the potential endogenous variable, ln DAY, but not correlated with the error term of the model [44]. The dataset contained additional variables that were able to be used for the IV model test. The natural log of the number of days taken to harvest (ln HARV), the natural log of the surface area of the tanks used in production (ln ARE), and a dummy indicating a winter month (WIN), were selected as instruments. It is expected that the values of these variables may influence the number of days of a growth period. The instruments were checked for appropriateness using the Hansen J over identification test, Kleibergen-Paap under identification test, and the Kleibergen-Paap weak identification tests (which are reported in Table 2) [38]. All three of these instrument tests indicate the chosen instruments are appropriate. The key test statistic for the appropriateness of IV, Endog, does not reject OLS, indicating that IV is not necessary. The IV model (Model 3) is not necessary, as the test statistic (Endog) listed in Table 2, fails to reject OLS. This indicates that an instrumental variables technique is not necessary. Thus, for the Cobb-Douglas specification, the OLS model with Fixed Effects is the preferred estimator.

Table 3 reports estimations of Equation (4) using a translog specification. The translog is of the form:

$$q_{it} = \alpha_0 + \sum_{m=1}^{M} \alpha_m x_{itm} + \sum_{l=1}^{M} \sum_{m=1}^{M} \alpha_{ml} x_{itm} x_{itl} + \eta_i + \upsilon_{it} \qquad (5)$$

The translog is a more flexible form than the Cobb-Douglas, and allows flexibility in the relationships between the variables. Indeed, the Cobb-Douglas is a special case of the translog, where the coefficients of the double summation in Equation (5) are zero. More generally, the translog can be considered to be a second order approximation of an arbitrary production function [45]. Again, models are analyzed with and without the location dummy variables. Table 3 gives F-tests for the joint significance of the coefficients on the PAR, INI, DAY, TEM, and PRE, and associated interactive terms. All variable groups were found to be jointly significant. The NANO term was insignificant in the translog specifications. The goodness-of-fit measures suggest the model including location dummies (i.e., Model 5 (OLS-FE)) is the preferred model.

Table 3. Translog Production Function.

Dep. Variable: ln AFDW	Model 4: Coefficient	Robust S.E.	Model 5: Coefficient	Robust S.E.
CON	62.32**	(24.119)	63.83***	(21.685)
ln PAR	-11.56**	(5.078)	-11.95***	(4.540)
ln INI	11.49**	(5.035)	21.24***	(4.730)
ln DAY	2.798**	(1.221)	1.863	(1.131)
ln TEM	-2.539	(2.871)	-2.547	(2.546)
ln PRE	65.85*	(36.603)	-1.401	(34.161)
½ $(\ln PAR)^2$	1.175**	(0.550)	1.261**	(0.491)
½ $(\ln INI)^2$	-12.91***	(2.333)	-8.863***	(2.150)
½ $(\ln DAY)^2$	-0.143**	(0.056)	-0.150***	(0.053)
½ $(\ln TEM)^2$	-0.679	(0.418)	-0.0268	(0.377)
½ $(\ln PRE)^2$	8.255	(11.141)	-8.697	(10.143)
ln PAR x ln INI	-1.048**	(0.512)	-1.692***	(0.461)
ln PAR x ln DAY	-0.305**	(0.123)	-0.247**	(0.112)
ln PAR x ln TEM	0.233	(0.279)	0.0951	(0.248)
ln PAR x ln PRE	-2.439	(3.125)	1.627	(2.939)
ln INI x ln DAY	-1.882***	(0.237)	-1.474***	(0.217)
ln INI x ln TEM	2.361***	(0.764)	0.998	(0.718)
ln INI x ln PRE	-6.224	(7.550)	-20.30***	(6.889)
ln DAY x ln TEMP	0.462***	(0.124)	0.563***	(0.115)
ln DAY x ln PRE	-5.821***	(1.716)	-4.187***	(1.542)
ln TEM x ln PRE	-8.981***	(3.151)	-0.257	(2.892)
NANO	0.0576	(0.051)	0.0532	(0.046)
SAP			-0.463***	(0.107)
COR			0.468***	(0.149)
CHM			-1.190***	(0.122)
NMS			-0.343***	(0.080)
N	495		495	
Std. Dev. of the Residuals	0.368		0.320	
R^2	0.49		0.61	
Adj R^2	0.47		0.58	
AIC[b]	448.5		330.8	
F-Test[c]				
F-Joint	18.9***		25.5***	
F-PAR	4.4***		5.8***	
F-INI	24.1***		15.1***	
F-DAY	16.4***		16.9***	
F-TEM	10.5***		4.3**	
F-PRE	3.8***		3.2***	

[a] PAR is daily-integrated photosynthethcially active radiation, INI is the initial concentration of algae at the time production is commenced, DAY is the number of days over which production occurred, TEM is the average daily variation in temperature, PRE is average daily precipitation, and NANO indicates that the species cultivated is *Nannochloropsis salana.* and zero otherwise.

[b] AIC: Goodness-of-fit measure considering the trade-offs between accuracy and complexity. A lower value indicates a preferred model.

[c] F-Test: Statistic examining the significance of the explanatory variables, as a group, in the model. The null hypothesis is that the variable groups are not significant. The results strongly reject the null in each model.

4. Discussion

4.1 Estimation of Elasticities

As previously stated elasticities measure the percentage change in one variable that is attributable to a 1% in another variable. Elasticities are useful measures of how a variable of interest, in this case biomass productivity, is related to input variables such as sunlight and temperature or initial concentration. Input elasticities measure the sensitivity of output to an increase in inputs. Table 4 shows input elasticities of production (calculated using the Cobb-Douglas and translog specifications) reported in Table 3 and Table 4. The elasticities are evaluated at the mean value of the inputs and are reported with 95% confidence intervals calculated using bootstrapping techniques.[6] For the Cobb-Douglas equation, the coefficient of the input is the elasticity, which can be taken directly from Table 2. Calculating the elasticity for the translog specification is more complicated as it requires giving values to the other inputs as these terms influence the value of the

[6] The bootstrapping of Regression Coefficient method was used [46]. The residuals from the original regression are randomly added back to the estimated values of the dependent variable, thereby, creating a pseudo dependent variable. The pseudo dependent variable is then used to estimate the regression. This was repeated 1,000 times. The results of the regression were then used to calculate 1,000 elasticity measures, which were then used to calculate the upper and lower limits of the 95% confidence interval.

translog elasticity via the interaction terms. It was decided to use the mean values in doing calculations of the elasticities.

Table 4 tells a fairly consistent story with the exception of INI. INI has a negative and significant elasticity both overall and individually for four out of five sites. The exception is COR, which had a positive elasticity. This indicates that INI is systematically too high for optimal production. One particular explanation for high overall INI is likely an incentive to avoid pests that may compromise algae growth.

Table 4. Input elasticities of production for the Cobb-Douglas and Translog fixed effects model with confidence intervals calculated using bootstrapping.[a]

Variable	Measure	Cobb-Douglas[b] (Model 2 OLS-FE)	Translog[c,d] (Model 5 OLS-FE)					
		All Data	All Data	CHM	COR	SAP	NMS	PEC
ln PAR	Elasticity	0.404	0.404	0.187	0.228	0.476	0.669	0.815
	95% L.L.[e]	0.402	0.402	0.178	0.230	0.468	0.661	0.799
	95% U.L.[f]	0.409	0.409	0.187	0.230	0.481	0.680	0.845
	S.E. of CI[g]	0.096	0.096	0.134	0.108	0.135	0.143	0.180
ln INI	Elasticity	-0.629	-0.629	-2.319	0.234	-0.892	-0.790	-0.157
	95% L.L.	-0.654	-0.654	-2.911	0.160	-0.923	-0.874	-0.449
	95% U.L.	-0.594	-0.594	-1.675	0.320	-0.870	-0.710	0.145
	S.E. of C.I .	0.208	0.208	0.433	0.408	0.303	0.225	0.262
ln DAY	Elasticity	0.090	0.090	-0.065	0.038	0.226	0.186	0.194
	95% L.L.	0.089	0.089	-0.064	0.038	0.225	0.184	0.193
	95% U.L.	0.090	0.090	-0.064	0.040	0.226	0.185	0.193
	S.E. of C.I.	0.032	0.032	0.050	0.051	0.050	0.038	0.046
ln TEM	Elasticity	0.340	0.340	0.508	0.261	0.646	0.365	-0.177
	95% L.L.	0.324	0.324	0.429	0.261	0.623	0.337	-0.200
	95% U.L.	0.351	0.351	0.572	0.263	0.667	0.388	-0.165
	S.E. of C.I.	0.129	0.129	0.210	0.141	0.209	0.188	0.196

ln PRE	Elasticity	-3.399	-3.399	-7.725	-1.488	-4.994	-3.497	-0.859
	95% L.L.	-4.339	-4.339	-7.790	-2.810	-7.157	-5.321	-2.659
	95% U.L.	-2.545	-2.545	-7.783	-0.216	-2.960	-1.769	0.889
	S.E. of C.I.	1.444	1.444	2.307	0.710	2.677	2.268	1.746

[a]The Bootstrapping on Regression Coefficient method was used [46, p. 17], p. 17.

[b]The formula for the input elasticity of production for the Cobb-Douglas is given by $E_i = \frac{d \ln Y}{d \ln X_m} = \frac{d\,y}{d\,x_i} = \alpha_m$ where α_m is the coefficient on input *m* from the Cobb-Douglas specification in Equation (4) [44].

[c]The formula for the input elasticity of production for the translog is given by $E_i = \frac{d \ln Y}{d \ln X_m} = \frac{d\,y}{d\,x_i} = \alpha_m + \sum_{l=1}^{M} \alpha_{ml} x_l$ where α_m is the coefficient on input *m* and α_{ml} is the coefficient on the cross interactive terms of input *m* form the translog specification in Equation (5) [44].

[d]Elasticities are calculated at the mean value of the regressors.

[e]L.L.: Lower limit of the confidence interval.

[f]U.L.: Upper limit of the confidence interval.

[g]S.E. of C.I.: Standard error of the confidence interval.

4.3 Simulation

The estimated elasticities presented in Table 4 are used to simulate the effect of a one standard deviation change from the mean values for PAR, INI, DAY, and TEM input variables on the output variable (AFDW). These are presented in Table 5. The simulation occurs while holding all other inputs constant. The standard deviation changes were given a positive or negative sign depending on the sign of the elasticity measure. The aim was to demonstrate the impact of input changes that would lead to positive output changes. One of the strongest simulated changes comes from the adjustment of starting density levels—a decrease in output of approximately 49% for the overall sample for a one standard deviation increase in INI. Higher net output with lower initial stocking densities could be an important economic result. It should be noted that the COR simulation suggests starting density should go up to improve production—an increase in output of approximately 17% for each standard deviation increase in INI. All sites, except CHM, simulate higher production levels with longer growing periods. The PAR simulation provides the expected result that increases in PAR will produce more algae. This is likely indicating that times of year with longer days are more conducive to production. The simulation regarding TEM indicates increases in TEM will lead to increased production. While these are important findings from a technological perspective, without reliable cost data it is unclear if such changes would be economically reasonable.

Table 5. Simulating input adjustments: Percent increases in production given changes of one standard deviation in the value of explanatory variables from mean values.[a]

Variable	Change (+ or -)	$\%\Delta$ AFDW	Change (+ or -)	$\%\Delta$ AFDW	Change (+ or -)	$\%\Delta AFDW$
	Total Data		CHM		COR	
PAR	+	13.4%	+	4.7%	+	6.6%
INI	-	48.9%	-	35.9%	+	16.9%
DAY	+	8.8%	-	9.6%	+	2.9%
TEM	+	13.6%	+	8.8%	+	5.0%
	SAP		NMS		PEC	
PAR	+	11.3%	+	14.2%	+	6.4%
INI	-	36.4%	-	44.9%	-	4.2%
DAY	+	21.3%	+	7.5%	+	7.9%
TEM	+	8.9%	+	5.3%	-	1.6%

[a]The formula for the simulation is given by $\%\Delta \text{AFDW} = E_i * (\frac{\Delta X}{X})$.

5. Conclusions

There is considerable interest in determining the feasibility of production of biofuels from microalgae, but such evaluations require assessment of productivity. To address this issue, a pooled time series, cross sectional data set is created using observations from five different production locations. This data set is believed by the authors to be the most extensive collected to date on algae production. The data are used to estimate a production function for outdoor cultivation of *Nannochloropsis sp*. Input elasticities of production are estimated that allow the evaluation of production efficiency. The results indicate that for the sample of production analyzed, the initial concentration of algae is too high and should be adjusted downward. This analysis, when combined with economic cost data, will provide more accurate insight into the economic feasibility of algae production.

The methodology used in this study is the first step towards developing more realistic economic models and assessments of the environmental impacts of algae production. It is clear from the work completed on this unique dataset that much remains to be done in terms of collecting reliable data on productivity and pond cultivation conditions. Differences in data collection on the key variables of biomass productivity and basic site conditions resulted in the use of proxy variables that introduce significant measurement error. As the elasticity measures show, it is possible to construct direct measures of the impact of changing input conditions on productivity. If improvements are made in the measurement of productivity and in the evaluation of which control parameters impact productivity, more accurate measures of profitability and environmental impact will be possible.

This study illustrates how applied production analysis techniques can provide vital information to those seeking to cultivate algae for commercial purposes. The production function approach allows for the measurement on productivity (and profitability) of changes in operating conditions. Better predictions of the impact of weather, water depth, temperature management strategies, predator and weed control strategies and the like can be rigorously analyzed using production analysis methods. The resulting elasticities provide specific control metrics for optimizing production and can provide a powerful toolset for reducing costs and environmental impact from large scale algae cultivation.

Acknowledgements

Funding was provided by the National Alliance for Advanced Biofuels and Bioproducts (NAABB), through the Department of Energy contract EE00030406, a public-private partnership of universities, companies, and U.S. national laboratories. The authors would like to thank Bryn Davis at Sapphire Energy for providing access to large scale cultivation data as well as numerous tours and discussions on the production of algae for fuel. Additionally, partners at CEHMM, NMSU, and Texas A&M were vital to the data collection efforts presented in this research. Thank you to the reviewers of this manuscript, who have provided many useful comments that have improved our work.

Bibliography

[1] Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Seibert and A. Darzins, “Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances,” *Plant,* vol. 54, no. 4, p. 621–639, 2008.

[2] M. Rickman, J. Pellegrino, J. Hock, S. Shaw and B. Freeman, "Life-cycle and techno-economic analysis of utility-connected algae systems," *Algal Research,* vol. 2, no. 1, pp. 59-65, 2013.

[3] J. Sheehan, T. Dunahay, J. Benemann and P. Roessler, "A look back at the us department of energy’s aquatic species program biodiesel from algae," NREL TP-580-24190, 1998.

[4] F. O. Holguin and T. M. Schaub, “Characterization of Microalgal Lipid Feedstock by Direct Infusion FT-ICR Mass Sjpectrometry,” *Algal Research,* p. in review, Submitted July 2, 2012.

[5] P. D. Patil, H. Reddy, T. Muppaneni, A. Mannarswamy, T. Schuab, P. Lammers, N. Nirmalakhandan, P. Cooke and S. Deng, “Power Dissipation in Microwave-Enhanced In-Situ Transesterification of Algal Biomass.,” *Green Chemistry,* vol. 14, pp. 809-817., 2012.

[6] T. Brown, P. Duan and P. Savage, “Hydrothermal Liquefaction and Gasification of Nannochloropsis sp.,” *Energy & Fuels,* vol. 24, no. 6, pp. 3639-3646, 2010.

[7] C. Miao, M. Chakraborty and S. Chen, “ Impact of reaction conditions on the simultaneous production of polysaccharides and bio-oil from heterotrophically grown Chlorella sorokiniana by a unique sequential hydrothermal liquefaction process.,” *Bioesour,* vol. 110, pp. 617-627, 2012.

[8] D. Elliott, “Catalytic hydrothermal gasification of biomass,” *Biofuels Bioproducts & Biorefining-Biofpr,* vol. 2, no. 3, pp. 254-265., 2008.

[9] P. Biller, R. Riley and A. Ross, “Catalytic hydrothermal processing of microalgae: Decomposition and upgrading of lipids,” *Bioresource Technology,* vol. 102, no. 7, pp. 4841-4848, 2011.

[10] P. Duan and P. Savage, “Upgrading of crude algal bio-oil in supercritical wate,” *Bioresource Technology,* vol. 102, no. 2, pp. 102(2), 1899-1906, 2011.

[11] A. Sun, R. Davis, M. Starbuck, A. Ben-Amotz, R. Pate and P. T. Pienkos, "Comparative cost analysis of algal oil production for biofuels," *Energy,* vol. 36, pp. 5169-5179, 2011.

[12] J. W. Richardson, J. L. Outlaw and M. Allison, "The Economics of Microalgae Oil," *AgBioForum,* vol. 13, no. 2, pp. 119-130, 2010.

[13] L. Batan, J. Quinn, B. Willson and T. Bradley, "Net Energy and Greenhouse Gas Emission Evaluation of Biodiesel Derived from Microalgae," *Environmental Science and Technology,* vol. 44, no. 20, pp. 7975-7980, 2010.

[14] M. J. Griffiths and S. T. L. Harrison, "Lipid productivity as a key characteristic for choosing algal species for biodiesel production," *Journal of Applied Phycol,* vol. 21, p. 493–507, 2009.

[15] L. Rodolfi, G. C. Zittelli, N. Bassi, G. Padovani, N. Biondi, G. Bonini and M. R. Tredici, "Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor," *Biotechnology and Bioengineering,* vol. 102, no. 1, pp. 100-112, 2008.

[16] E. Suali and R. Sarbatly, "Conversion of microalgae to biofuel," *Renewable and Sustainable Energy Reviews,* vol. 16, p. 4316– 4342, 2012.

[17] L. Lardon, A. Helias, B. B. Sialve, J. P. Stayer and O. Bernard, "Life-cycle assessment of biodiesel production from microalgae," *Environmental Science and Technology,* vol. 43, p. 6475–6481, 2009.

[18] P. Campbell, BeerT and B. D, "Life cycle assessment of biodiesel production from microalgae in ponds," *Bioresource Technology,* vol. 102, pp. 50-56, 2011.

[19] A. Clarens, R. E, M. White and C. L, "Environmental life cycle comparison of algae to other bioenergy feedstocks," *Environmental Science and Technology,* vol. 44, no. 5, p. 1813–1819, 2010.

[20] R. Hoefnagels, E. Smeets and F. A, "Greenhouse gas footprints of different biofuel production systems," *Renewable and Sustainable Energy Reviews,* vol. 14, no. 7, p. 1661–1694, 2010.

[21] F. Møller, E. Slentø and P. Frederiksen, "Integrated well-to-wheel assessment of biofuels combining energy and emission LCA and welfare economic Cost Benefit Analysis," *Biomass and Bioenergy,* vol. 60, pp. 41-49, 2014.

[22] S. Zou, Y. Wu, M. Yang, I. Kaleem, C. Li and J. Tong , "Production and characterizationof bio-oil from hydrothermal liquefaction," *Energy,* vol. 35, pp. 5406-11, 2010.

[23] Y. Haik, M. Selim and T. Abdulrehman, "Combustion of algae oil methyl ester in an indirect injection diesel engine," *Energy,* vol. 36, pp. 1827-35, 2010.

[24] J. Benneman and W. Oswald, "Systems and economic analysis of microalgae ponds for conversion of CO2 to biomass," 1996.

[25] T. Lundquist, I. C. Woertz, N. W. T. Quinn and J. R. Benemann, "A realistic technology and engineering assessment of algae biofue.," UC Berkley, Berkley, 2010.

[26] P. T. Pienkos, "Historical overview of algal biofuel technoeconomic analysis," in *EERE algal biofuels technology roadmap workshop proceedings*, University of Maryland, 2008.

[27] E. Sforza, A. Bertucco, T. Morosinotto and G. Giacometti, "Photobioreactors for microalgal growth and oil production with Nannochloropsis salina: From lab-scale experiments to large-scale design," *Chemical Engineering Research and Design,* vol. 90, no. 9, p. 1151–1158, 2012.

[28] R. Slade and A. Bauen, "Micro-algae cultivation for biofuels: Cost, energy balance, environmental impacts and future prospects," *Biomass and Bioenergy,* vol. 53, pp. 29-38, 2013.

[29] C. M. Downes and Q. Hu, "First Principles of Techno-Economic Analysis of Algal Mass Culture," in *Handbook of Microalgal Culture*, 2nd ed., A. Richmond and H. Qiang, Eds., John Wiley & Sons, Ltd. Published by Blackwell Publishing Ltd., 2013, pp. 210-326.

[30] H. R. Varian, Microeconomic Analysis, 3rd Edition, New York, NY: W. W. Norton & Co., 1992.

[31] H. R. Varian, Intermediate Microeconomics: A Modern Approach, New York, NY: W. W. Norton & Co., 1996.

[32] B.-Y. Chen, "On some geometric properties of h-homogeneous production functions in microeconomics," *Kragujevac Journal of Mathematics,* vol. 35, no. 3, pp. 343-357, 2011.

[33] M. Morroni, Production Process and Technical Change, New York: Cambridge University Press, 1992.

[34] B. H. Baltagi, Econometrics of Analysis of Panel Data, 2nd Ed., New York: John Wiley and Sons, 2001.

[35] M. S. Chautona, Y. Olsen and O. Vadstein, "Biomass production from the microalga Phaeodactylum tricornutum: Nutrient stress and chemical composition in exponential fed-batch cultures," *Biomass and Bioenergy,* vol. 58, pp. 87-94, 2013.

[36] Natural Resource Ecology Laboratory, "UV-B Monitoring adn Research Program," Colorado State University, [Online]. Available: http://uvb.nrel.colostate.edu/UVB/index.jsf. [Accessed 10 June 2013].

[37] NASA, "Photosyntheetically Active Radiation," NASA, 2011 24 October. [Online]. Available: http://disc.sci.gsfc.nasa.gov/hydrology/data-holdings/nldas-parameters/photo-active-radiation. [Accessed 7 June 2013].

[38] J. Van Wagenenemail, T. W. Millermail, S. Hobbs, P. Hook, B. Crowe and M. Huesemann, "Effects of Light and Temperature on Fatty Acid Production in Nannochloropsis Salina," *Energies,* vol. 5, no. 3, pp. 731-740, 2012.

[39] B. Crowe, S. Attalah, S. Agrawal and e. al., "A Comparison of Nannochloropsis salina Growth Performance in Two Outdoor Pond Designs: Conventional Raceways versus the ARID Pond with Superior Temperature Management," *International Journal of Chemical Engineering,* vol. 2012, no. 2012, p. 9, 2012.

[40] R. Wicklin, "Log Transformations: How to handle negative data values," SAS, 27 April 2011. [Online]. Available: http://blogs.sas.com/content/iml/2011/04/27/log-transformations-how-to-handle-negative-data-values/ . [Accessed 31 December 2013].

[41] A. Sukenik, O. Zmora and Y. Carmeli, "Biochemical quality of marine unicellular algae with special emphasis on lipid composition. II. Nannochloropsis sp.," *Aquaculture,* vol. 117, no. 3-4, pp. 313-326,

1993.

[42] S. Boussiba, A. Vonshak, Z. Cohen, Y. Avissar and A. Richmond, "Lipid and biomass production by the halotolerant microalga Nannochloropsis salina," *Biomass,* vol. 12, no. 1, pp. 37-47, 1987.

[43] J. Hausman, J. H. Stock and Y. Motohiro, "Asymptotic properties of the Hahn–Hausman test for weak-instruments," *Economics Letters,* vol. 89, no. 3, pp. 333-342, 2005.

[44] W. H. Greene, Econometric Analysis, 5th ed., New York: Prentice Hall, 2003.

[45] H. G. Jacoby, "Shadow Wages and Peasant Family Labour Supply: An Econometric Application to the Peruvian Sierra," *The Review of Economic Studies,* vol. 60, no. 4, pp. 903-921, 1993.

[46] C. Z. Mooney and R. D. Duval, Bootstrapping: A Nonparametric Approach, Newbury Park, CA: Sage, 1995.

Figure 1

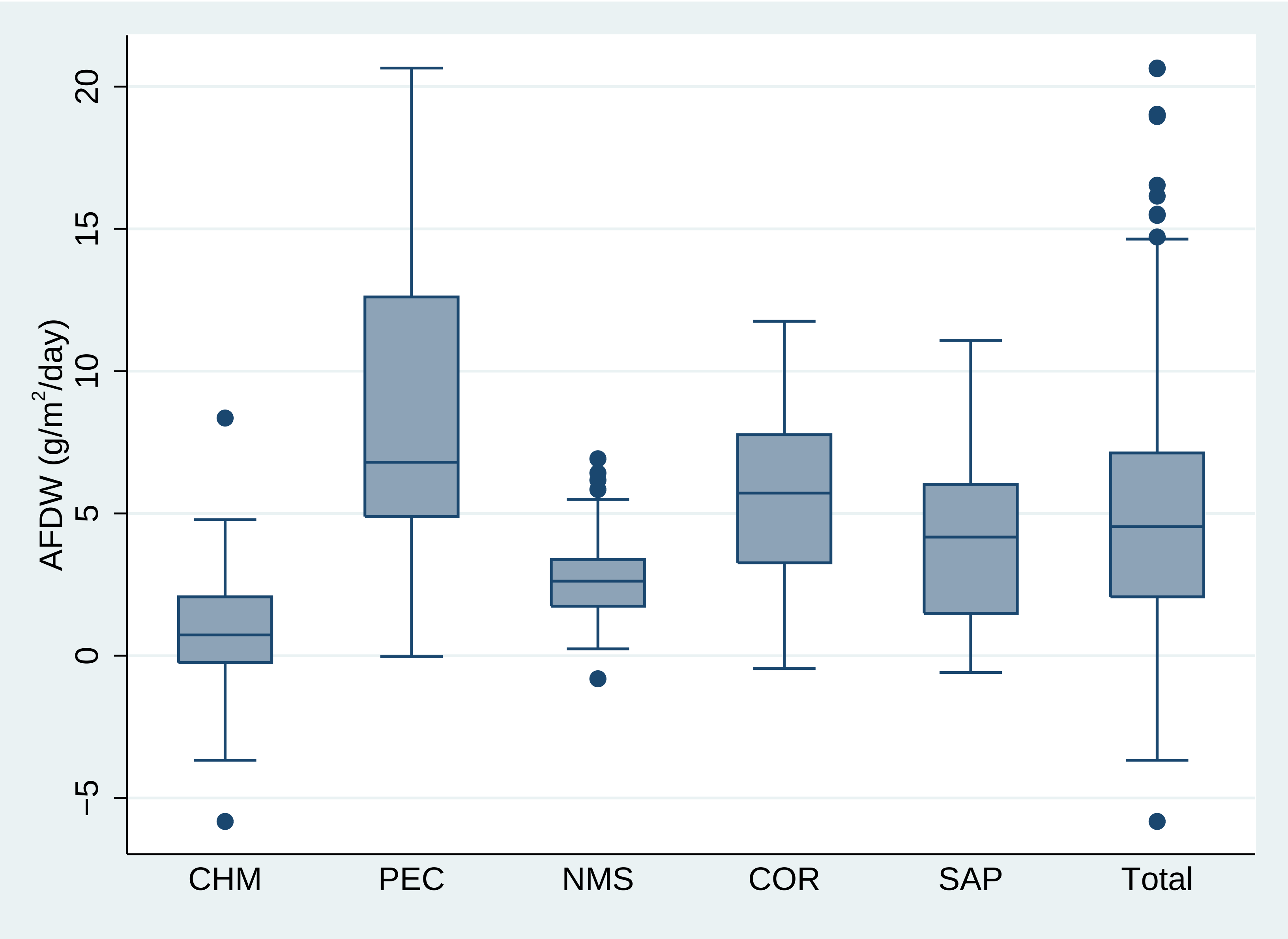
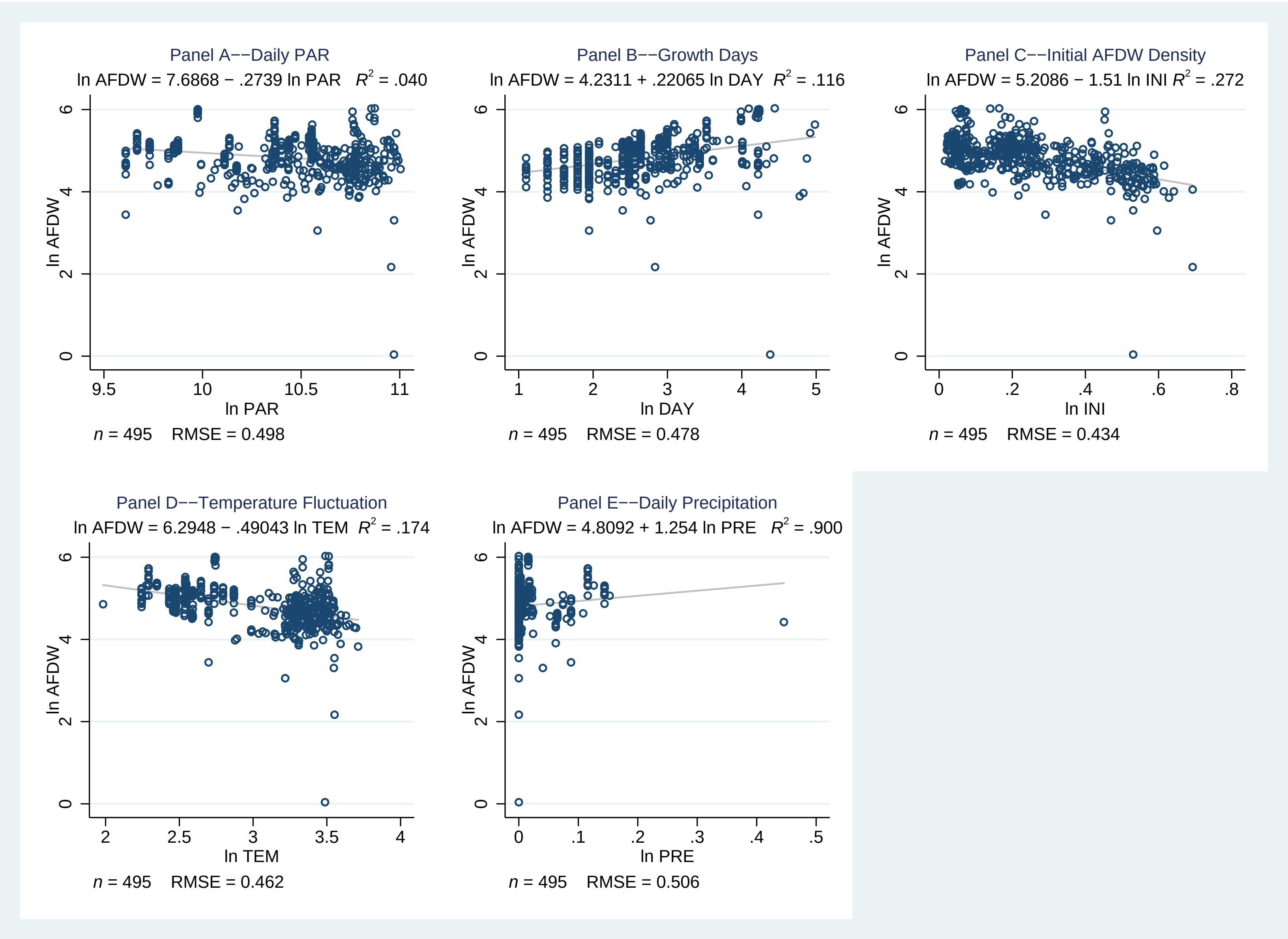


Figure 2



Panel A−−Daily PAR
ln AFDW = 7.6868 − .2739 ln PAR R^2 = .040
ln AFDW
ln PAR
n = 495 RMSE = 0.498
Panel B−−Growth Days
ln AFDW = 4.2311 + .22065 ln DAY R^2 = .116
ln AFDW
ln DAY
n = 495 RMSE = 0.478
Panel C−−Initial AFDW Density
ln AFDW = 5.2086 − 1.51 ln INI R^2 = .272
ln AFDW
ln INI
n = 495 RMSE = 0.434
Panel D−−Temperature Fluctuation
ln AFDW = 6.2948 − .49043 ln TEM R^2 = .174
ln AFDW
ln TEM
n = 495 RMSE = 0.462
Panel E−−Daily Precipitation
ln AFDW = 4.8092 + 1.254 ln PRE R^2 = .900
ln AFDW
ln PRE
n = 495 RMSE = 0.506