Integrating Forecasting for Weather-Optimized Crop Cutting for Zalliant

Skye Crocker, Colin Hascup, Ahmed Malik, Lam Nguyen, Ruilin Zhang
Team 2, DSC 383 / 483
Goergen Institute for Data Science and Artificial Intelligence
University of Rochester

1. Introduction

Efficient crop harvesting and drying are critical for preserving forage quality and minimizing post-harvest losses. When stored as hay or silage, forage crops such as alfalfa must be adequately dried to prevent spoilage and ensure optimal nutritional value. While mechanical drying is available, environmental or field drying is generally preferred due to its sustainability and cost-effectiveness. However, natural drying is highly dependent on favorable weather conditions. Therefore, the ability to accurately predict optimal harvest windows and drying timelines is essential for farmers seeking to maintain forage quality while minimizing operational costs.

Environmental or field drying is the most cost-effective and sustainable approach to forage preservation, leveraging solar radiation, ambient temperature, and wind speed to reduce crop moisture to safe storage levels. Mechanical drying, while effective, is energy-intensive and generally used only when weather conditions are unfavorable. To maximize the efficiency of field drying, it is crucial to align harvest timing with upcoming periods of favorable weather. Studies have demonstrated that strategic harvesting based on weather forecasting can significantly reduce the need for mechanical drying and lower the incidence of spoilage due to high moisture content (Craig et al., 2009).

This project was conducted in collaboration with Zalliant, an agricultural technology company that develops wireless sensor systems for monitoring and managing agricultural processes. Zalliant's core technologies focus on time series monitoring of animal environments, feed processing, and grain storage conditions, including temperature tracking and moisture control, to prevent post-harvest damage and improve crop storage outcomes. Their systems are particularly relevant for crops such as canola, rice, wheat, barley, and alfalfa, where improper drying or storage can result in significant losses. By leveraging sensor data and forecasting tools, Zalliant seeks to help farmers optimize dry matter management and avoid heat-related damage in storage facilities.

To support this mission, we built two complementary models to help Zalliant determine when to harvest crops and drying decisions based on weather conditions: a Growing Degree Day (GDD) calculator to estimate crop maturity and guide harvest timing, and a drying calculation algorithm to forecast post-harvest moisture reduction. Together, these models provide Zalliant with a more integrated, data-driven approach to harvesting decisions. Although further validation and live testing are planned, the project lays a solid foundation for expanded crop support, improved prediction granularity, and long-term cost savings through optimized field operations.

2. Dataset

2.1 Weather Data

Accurate and timely weather data is essential for modeling crop growth, predicting drying conditions, and informing harvest decisions. For this project, historical and forecasted weather data were acquired using the Visual Crossing Weather API, a robust platform that provides granular, location-specific meteorological data. This data served as the foundation for both the Growing Degree Day (GDD) calculator and the drying calculation algorithm.

Weather data was collected for multiple locations, such as Ithaca of New York, and Alder Springs of California, chosen for their relevance to crop harvesting locations. For each location, data were queried for specific date ranges corresponding to typical planting, growing, and harvesting periods. Although these datasets were relatively small in size, for example, one dataset contained only 93 rows and 33 columns, they were well-structured and included various weather variables, making them sufficient for model development.

The API offers access to a wide range of weather variables, including temperature, humidity, dew point, wind speed, precipitation, solar radiation, and vapor pressure, on daily and hourly time scales (see table 1). While many data features were accessible through the API's free tier, a premium subscription sponsored by Zalliant was used to access advanced data fields, such as real-time soil moisture data. This granularity allowed us to tailor data inputs to the needs of each model. For example, the GDD calculator utilized daily minimum and maximum temperatures to estimate heat accumulation, while the drying model relied on harvest date and soil moisture data to predict the crop's moisture content over time.

datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	humidity	precip	precipprob	precipcover	preciptype	snow	snowdepth	windgust	windspeed
2025-01-01	42.1	36.4	39	40	26.9	34.3	34.1	82.8	0.467	100	62.5	rain	0.7	0	36.4	19.8
2025-01-02	36.6	29.9	34.1	27.4	20.4	24.1	22.4	62.1	0.025	100	16.67	rain,snow	0	0.6	42.7	23.3
2025-01-03	33	26.8	30.1	24.3	17.5	21.1	17.6	59.2	0	0	0	rain,snow	0.1	2.4	32.2	16
2025-01-04	27.1	20.1	24	25.6	6.3	13.4	12.9	62.2	0.002	100	4.17	snow	0	2.4	38.3	21.2
2025-01-05	27.3	15.1	24.6	18.6	4.8	14.8	15.2	67	0	0	0	snow	0.7	0.2	30	15.6

Table 1 Example of the Dataset Structure (First 5 Rows and 17 Variables)

2.2 USDA California Direct Hay Report

The United States Department of Agriculture (USDA) releases bi-weekly direct hay reports for the California region [3]. Each report is typically 2-3 pages in PDF format and presents text summaries alongside structured trade data. It contains information such as hay type, grade (e.g., premium or good), bale size, quantity, pricing per bale or ton, and regional freight/use classifications. Here is a snapshot of part of the report on Dec 20th, 2024 (see figure 1). Of all the features that were provided in the report, the primary ones that this project concentrates on are Location (combined with weather data obtained and presented in section 2.1), Crop type (only alfalfa for this project), Crop Age, and Date of the Report (time). Although most variables themselves are very straightforward, we would like to clarify more about Crop Age. The set of values for this variable in the report includes: {New Crop, Old Crop, First, Second, etc.}, referring to newly planted crops, old crops (not harvested yet), first time cutting, second time cutting, etc., of the crops. This feature is critically important for the team to determine the timeline of harvest status of the crops.

For this project, we extracted data (from the USDA website) between 2021 and 2025, with full-year data for 2023, 2024 until April 2025, and for the years from 2021-2022, we extracted mainly summer months (March-September) data. In total, we gathered 102 PDF files for 4 years of the California Direct Hay Report from the USDA website [3].

North San Joaquin Valley

Hay (Conventional)

Alfalfa - Premium (Trade/Per Bale)

	Qty	Price Range	Wtd Avg	Freight/Use	Description	Crop Age
Small Square 3 Tie	512	14.00	14.00	F.O.B.-Retail	Covered	New Crop

Sacramento Valley

Hay (Conventional)

Alfalfa - Premium (Trade/Per Bale)

	Qty	Price Range	Wtd Avg	Freight/Use	Description	Crop Age
Small Square 3 Tie	512	12.00	12.00	F.O.B.-Retail		New Crop

Fig. 1 Snapshot of USDA California Direct Hay Report (Alfalfa entry, Dec 20, 2024)

2.3 Exploratory Data Analysis (EDA)

2.3.1 Weather Data Patterns:

Across locations that we sampled later in this project (Ithaca, NY and multiple locations in California), cumulative GDD over time showed a steady increase from early spring to midsummer, as expected. We also noted seasonal trends in temperature, solar radiation, and precipitation that informed harvest timing and drying conditions. Generally, weather data seems consistent with the climate of Northern America (U.S.) in each season (e.g: little rain during winter months, higher temperatures and more rain during summer months between June-August)

2.3.2 Harvest Timing Trends

By analyzing 102 USDA Direct Hay Reports between 2021–2025, we identified consistent alfalfa harvest cycles. Reports categorized as "First", "Second", or "Third", etc. We recognized that the pattern of cutting seems consistent over the years (similar harvest status for similar months of the years), suggesting that harvesting might be consistent with the climates of the regions, and is a common practice among farmers.

2.3.3 Missing Data and Data Quality

Weather data had minimal missing values, and USDA report data was clean due to its structured publication format. We performed preprocessing steps including unit conversions, derived variables (e.g., VPD), and ensured matching between crop and weather data by region and time.

3. Data Engineering

3.1 Weather Data Data Engineering

The API's flexibility and scalability enabled seamless integration with Python-based data processing tools, returning clean and structured pandas DataFrames suitable for modeling and visualization. To ensure data quality and consistency, several preprocessing steps were applied. These included:

- Checking missing values
- Standardizing categorical fields (e.g., weather descriptions and condition labels)
- Converting units (e.g., Fahrenheit to Celsius)
- Deriving additional features, such as vapor pressure deficit and solar insolation, which were essential for calculating drying rates

3.2 Direct Hay Report Data Transformation and Cleaning

For the direct hay report, the team performed the process of data extraction of structured data from 102 PDF files (as mentioned in the Section 2.2 of this report) by selecting these following variables to include in a new CSV file (which reports crop harvest status): Date (of the report), Location, Crop Age, Crop quality (Premium, Standard, etc.). For the sake of simplicity for the project, we chose data from the alfalfa crop only. The CSV file contains 105 rows of data.

Since data was extracted directly from USDA official reports, there was no missing data. After extraction, we added weather variables from corresponding California locations to the dataset for further analysis.

4. Model Development

Having established our data foundation through comprehensive research, acquisition, and preprocessing, we now present the two models developed for this project. The Growing Degree Day (GDD) calculator and drying algorithm leverage these processed datasets to deliver actionable insights for harvest optimization.

4.1 Growing Degree Day (GDD) Calculator

One of the most critical milestones of this project was the development of a user-friendly, convenient, and robust GDD calculator implemented in Python. This tool processes weather data from the Visual Crossing API to predict crop maturity and optimize harvest timing. At its core, the GDD metric quantifies a crop's cumulative heat values, which directly correlate with its developmental stages. The fundamental daily GDD calculation is expressed as:

$$GDD = \frac{T_{max} + T_{min}}{2} - T_{base}$$

Where Tmax and Tmin represent the daily maximum and minimum temperatures, respectively, and Tbase denotes the crop-specific base temperature (e.g., 41°F for alfalfa). This daily value is accumulated over time to assess maturity, with alfalfa typically reaching harvest readiness at 750-755 GDD.

4.1.1 Establishing the GDD Calculator Tool

To enhance accuracy across diverse climates and crops, the calculator supports multiple GDD computation methods: average, modified average, single sine, symmetrical triangle, and double sine/triangle method. To ensure wide usability, the GDD calculator was implemented as a Python-based tool with a clean and intuitive interface (see figure 2(a)). See figure 2(b) for the cumulative GDD plot for Ithaca, NY, generated by the existing calculation tool. Farmers can input any U.S. location, select from multiple GDD calculation methods, and define the crop-specific base temperature (T_base), which proves the tool's flexibility for any crop beyond just alfalfa. Additionally, the calculator offers configurable upper and lower thresholds for modified GDD calculations, making it adaptable to future crop types with different growth requirements.

The tool fetches weather data automatically from the Visual Crossing API and plots cumulative GDD over time. During the progress of this project and presented later in this report (sections 4.1.2 and 5.1), we verified the applicability and accuracy of this GDD tool on real harvest data to determine if this tool is reliable for farmers in predicting optimal harvest time.

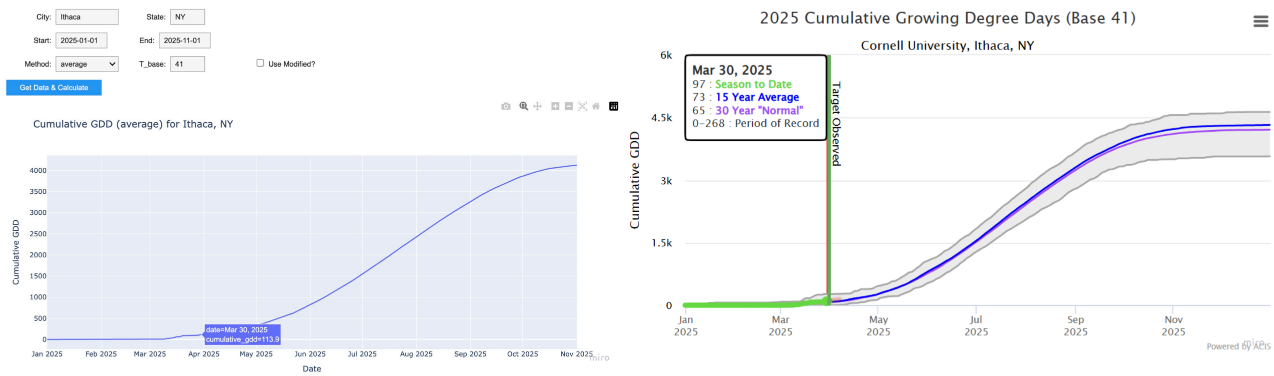

Fig. 2(a) Interface of Tool

Fig. 2(b) Tool-built Cumulative GDD

4.1.2 Analysis and verification of Cumulative GDD data provided by GDD calculator

We used the mentioned-above GDD calculator tool and verified with real data gathered USDA hay report (as mentioned in Sections 2.2 and 3.2). Literature research revealed that farms are encouraged to harvest (cut) their crop of alfalfa once cumulative GDD reaches between 700-750. [4] Our team tested on various locations and during multiple time ranges in California to see if the GDD calculator that we built could take in weather data (from API) and plot appropriate cumulative GDD plots for these regions. Then, using GDD cumulative plots, we compared the moments that California farms harvested (as shown in the CSV file created in section 3.2) with the moments that cumulative GDD reached about 750 in the corresponding location. If these times match, it indicates that farms follow the recommended practice of harvesting at cumulative GDD of 750 (instead of harvesting at a random time), and more importantly, the GDD calculator tool that we built in Section 4.1.1 can now serve as an accurate, easy and convenient tool for any farmer to predict their optimal harvesting time based on an open access weather data API (they may just need to simply input their specific location and croptype's T-base temperature).

4.2 Drying Calculation Algorithm

Once a harvest date is established, either post-harvest or forecasted using the GDD Calculator, it becomes critically valuable to know the estimated moisture content of the crop as it dries over the next couple of days. The drying model developed by the team allows users to enter a harvest date and location as input and provides forecasted moisture content for Alfalfa for each day following the harvest until it reaches a threshold value or reaches the end of the API's weather forecasts. Due to lack of training data containing the target variable—moisture content (%)—the model relies on previously validated work from Rotz & Chen [2] in the form of an equation which has been implemented as a modern algorithm. The general time-series formula for the moisture content of Alfalfa is expressed as:

$$M_{t+1} = M_t \cdot e^{-k_t}$$

Where M_t is the moisture content (%) of the crop on day t, and k_t is the drying rate constant for day t. The drying rate constant k (or DR) is influenced by various variables related to the environment which is calculated as follows:

$$DR = \frac{SI\,(1. + 9.03\,(AR)) + 43.8\,(VPD)}{61.4\,(SM) + SD\,(1.82 - 0.83\,(DAY))\,(1.68 + 24.8\,(AR)) + 2767}$$

where: SI = solar insolation, W/m^2
VPD = vapor pressure deficit, kPa
DAY = 1 for first day, 0 otherwise
SM = soil moisture content, % dry basis
SD = swath density, g/m^2
AR = application rate of chemical solution, g of solution/g of dry-matter

Typical values for each of these variables, as well as their effect on the resulting value of k were provided in the original paper and used as a sanity check for validating our implementation.

Model variable	Units	Variable Values Min.	Variable Values Typical	Variable Values Max.	Predicted k Min.	Predicted k Max.	Predicted k Diff., %
Model with vapor pressure deficit (equation [4]):							
Application rate	g liquid/g DM	0	0.075	0.25	–	—	–
Vapor pressure deficit	kPa	0	2.5	4.5	0.154	0.197	28
Swath density	g/m^2	150	450.0	1500.0	0.199	0.128	-36
Day		0	1.0	1.0	0.156	0.178	14
Solar insolation	W/m^2	0	700.0	950.0	0.024	0.232	867
Soil moisture content	%db	10	17.0	25.0	0.196	0.160	-18

Table 2 Typical Variable Ranges in Rotz & Chen (1985) Model Along with Corresponding k Value Range

4.2.1 Drying Algorithm Variables

Various challenges presented themselves in acquiring each variable and ensuring consistency with the recorded literature values. Here is a complete breakdown of each variable.

- **Solar insolation W/m^2** - This variable is directly provided by the Visual Crossing weather API. However, values provided by the API were consistently much lower than the typical value mentioned in the paper. After careful scrutiny, it was found that the value used in the paper is in fact peak solar insolation while the weather api provided average solar radiation. The solution to this was to extract hourly data from the api and find the max hourly solar insolation for each day and append this to our daily dataframe.

- **Vapor Pressure Deficit kPa -** This variable is not directly provided by the Visual Crossing weather API, but can be calculated as a derivative of temperature and dew point which are provided. The equation for this [6] is:

$$VPD = 0.6108 * (\exp((17.27 * T) / (T + 237.3)) - \exp((17.27 * Td) / (Td + 237.3)))$$

Where *T* is the daily temperature and *Td* is the daily dew point temperature both in Celcius.

- **Day -** Simple binary mechanism incorporated into algorithm. Helps account for the fact that the most drying occurs the first day right after the crops are cut.

- **Soil Moisture Content -** This variable is provided in the weather API; however, it requires a corporate/premium subscription.

- **Swath Density g/m^2 -** This variable is not provided in the weather dataset and would ideally be input by the user as it relates to the procedure of how the crops were harvested. Advice from domain experts suggest that usually the standard for modern values is 25-30 or 70 plants/ft^2. Attempts to convert these into units of g/m^2 resulted in values very different from the range provided in the chart, so until further clarification or a more reliable conversion can be provided, we will use the typical value provided in the chart above and use a constant 450 g/m^2.

- **Application Rate of Chemical Solution -** This is another variable that would need to be user defined as input and not part of the weather API. According to the paper, the standard value for this is 0 (i.e. no chemical solution applied) so this is what we stuck with for now.

5. Performance and Results

5.1 Verification of Harvest Timing Predictions Using the Cumulative GDD Calculations

Below is an example of a verification we performed to compare between cumulative GDD calculator plot and real harvest data (the task presented in Section 4.1.2). Figure 4 shows cumulative GDD for alfalfa in Alder Springs, California, from early January till late April, 2025. As shown from the plot, cumulative GDD reached 750 around the date of March 29th, 2025. Upon comparing with the harvest data (the CSV file), shown in Table 3, we found that this crop of alfalfa was harvested somewhere between March 28th and April 11th, showing harvesting at the literature-recommended cumulative GDD of 750 [4]. In Table 3, till March 28th, the crop was still categorized as “Old Crop”, meaning having been planted for some time but not yet harvested; however, on April 4th, it was already harvested (First-cut).

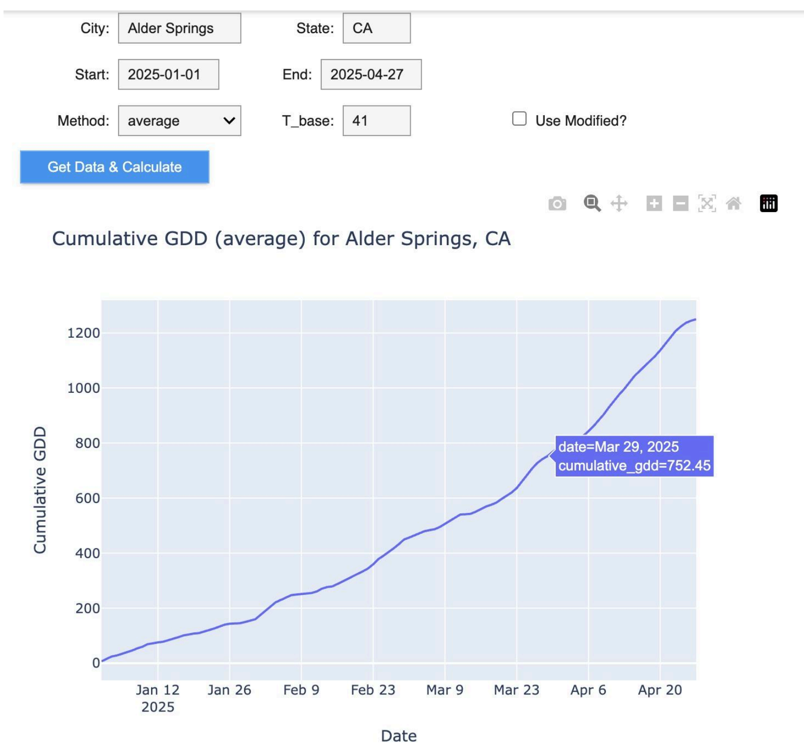

Fig. 4 Cumulative GDD accumulation in Alder Springs, CA from Jan to Apr 2025, generated by GDD calculator

2025-03-28	92443	Alfalfa	Premium	Old	0
2025-04-11	92443	Alfalfa	Premium	1	1

Table 3 Alfalfa Harvest Status (Mar 28-Apr 11, 2025) - Alder Springs, CA

We went further with data from the second cut and onwards (3rd cut, 4th cut, etc.) of the same specific alfalfa crops (same seeding time and location) in two locations in California: Alder Springs and San Joaquin Valleys for a longer period (from 9 months to 1 year, as shown in the Fig. 5(a) and 5(b)). We found out that the time intervals between any increment of 750 units in cumulative GDD generally coincided with the time intervals (from the CSV file data) between the previous and the next cut (the ith cut and the i+1th cut). In figures 5(a) and 5(b), the red lines indicate the dates that cuts (harvest) were performed, with the left-most lines representing the dates of the earlier cuts (1st, 2nd, 3rd, etc.). We recognized the pattern that during the intervals when the slope is steeper (quick increase of GDD), the time interval between two consecutive harvests is shorter or much shorter. (i.e, Farmers generally harvested every 750 unit increase in cumulative GDD, so if it increased faster, they harvested "sooner").

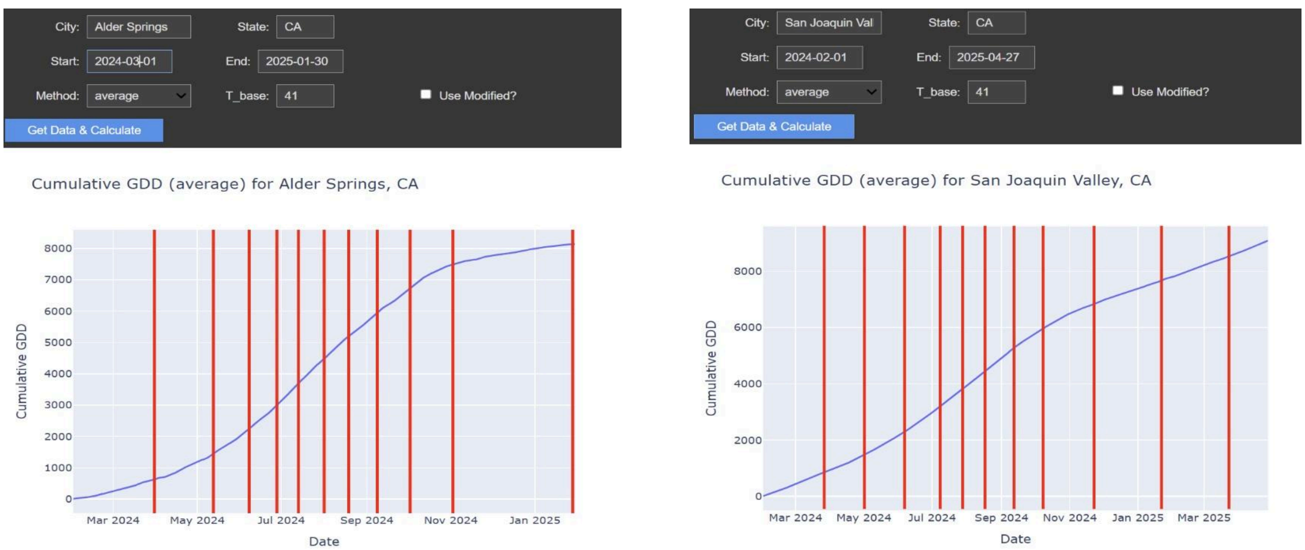

Fig. 5 (a) and 5(b) Cumulative GDD vs. Actual Harvest Dates at Two California Locations

(Blue line indicates cumulative GDD plot, Red lines indicate dates of real harvest)

We determined that this result is expected and shows a consistency between our GDD calculator model with real harvest data, proving a consistency between theoretical knowledge (harvest recommended every 750 GDD unit increase) and real data (Farmers do generally harvest accordingly). Since our calculator tool plots Cumulative GDD based on open-source, convenient API weather data, farmers can now use this tool to easily predict an optimal harvest time based on weather forecast (which is highly accurate nowadays) [5].

5.2 Drying Algorithm Model Results

As noted before, no proper validation data was available to the team for checking the moisture content with a ground truth. However, after inputting our variables, we do see values of k that are consistently aligned with the ranges provided in table 2. Given that this is a well-established paper and model, we are optimistic that our implementation is reliable. Zalliant will also validate the model over the summer as crops are harvested and live moisture content data becomes available. Figure 3 shows an example of output provided by the model.

Fig.7 Forecasted moisture content of alfalfa harvested on April 28, 2025, in Alder Springs, CA

Figure 7 shows the forecasted moisture content for Alfalfa scheduled to be harvested on April 28, 2025. It shows that the crop will reach moisture content levels safe for storage (12-18%) on May 8. One can also examine a dramatic decrease in the rate of drying after May 3, potentially indicating a good day to switch from environmental drying to mechanical drying.

6. Conclusion and Next Steps

This project successfully developed two models to optimize crop harvesting and drying decisions for Zalliant. The Growing Degree Day (GDD) calculator, validated against USDA harvest reports, demonstrated robust accuracy in predicting optimal harvest windows for alfalfa at the established threshold of 750-755 GDD. Its flexibility, supporting multiple calculation methods (average, sine, triangle) and customizable temperature thresholds, ensures adaptability across

U.S. regions. Complementing this, the drying algorithm, grounded in the work of Rotz and Chen (1985), provided reliable moisture decay forecasts to guide mechanical drying transitions. Together, these tools address a critical gap in agricultural decision-making by integrating weather data with crop physiology.

The project's validation phase revealed key insights: harvest intervals closely aligned with cumulative GDD increments, and steeper GDD slopes correlated with shorter harvest cycles. These findings not only confirm the models' efficacy but also highlight the potential for data-driven farming practices. However, challenges such as limited validation data and the need for domain-specific research (e.g., vapor pressure deficit, solar radiation) underscored the complexity of translating theoretical models into practical tools. Notably, although we successfully USDA Direct Hay Report to verify dates of harvest, this data source was not the optimal source for data validation. Since the report only came out every two weeks, the harvest date was only estimated with limited accuracy, with an absolute error of up to 14 days.

To build on this foundation, immediate priorities include live validation of the drying model during the summer of 2025. Mid-term goals involve expanding model applicability to other crops (e.g., wheat, barley) by incorporating crop-specific parameters and refining algorithms with hourly data granularity. Long-term objectives focus on scalability and accessibility. Additionally, comprehensive documentation of code and methodologies will ensure reproducibility and facilitate future iterations.

References

[1] Pitt, R. E. *Silage and Hay Preservation*. NRAES-5, Cooperative Extension, Natural Resource, Agriculture, and Engineering Service (NRAES), Cornell University, August 1990.

[2] Rotz, C.A. & Chen, Yi. (1985). Alfalfa Drying Model for the Field Environment. Transactions of the ASAE. American Society of Agricultural Engineers. 28. 1687-1672. 10.13031/2013.32500.

[3] U.S. Department of Agriculture. (2025). *California Direct Hay Report*. My Market News. https://mymarketnews.ams.usda.gov/viewReport/2904

[4] Fick, W.H., & Kilgore, A. (2012). *Estimating first cutting alfalfa quality using growing degree days*. Forage and Grazinglands, 10(1), 1–6. https://doi.org/10.1094/FG-2012-0162-RS

[5] Mass, C.F. (2021). *The Increasing Skill of Numerical Weather Prediction and the Growing Impact of Weather Forecasts*. Bulletin of the American Meteorological Society, 102(3), E549–E565. https://doi.org/10.1175/BAMS-D-20-0200.1

[6] Omni Calculator. (n.d.). *Vapor Pressure Deficit Calculator*. Retrieved April 26, 2025, from https://www.omnicalculator.com/biology/vapor-pressure-deficit