

# Fool-Proofing Recipe Measurements

By Samantha Maticka and Kurt Nelson

## Abstract

This text should be back in two columns

## 1 Introduction

Many of us casual cooks are not able to reproduce winning recipes when applying the Italian Grandma method (IGM), and instead we are left with one-hit-wonders followed by an onslaught of failed recreation attempts that pale in comparison and litter your fridge with dreadful leftovers. Our project seeks to end this dystopia.

The broader objective of our project is to create a smart recipe recorder and instructor. Essentially, you can either 1) make a recipe-free dish, in which you add ingredients at will and a smart device films and records the recipe, or 2) you can create a saved recipe, where the device tells you when to stop pouring a specified ingredients.

We therefore propose applying machine learning and computer vision to teach phones how to measure for us. As a first step to solving this burdensome problem, our project will focus on volume prediction of a poured liquid. The world will soon be a better place.

## 2 Goal

Given features about a flowing liquid, one can calculate the volume poured with basic principles. Namely

$$V = \int v(t, z) A(t, z) dt, \quad (1)$$

where  $V$  is the total poured volume,  $v(t, z)$  is the velocity of the pour stream, and  $A(t, z)$  is the stream area.  $v(t, z)$  and  $A(t, z)$  are both time dependent and require evaluation at a fixed reference height  $z$ . However the calculation can be simplified if the a representative time average area  $\bar{A}$  and velocity  $\bar{v}$  are

known over a pour duration  $\bar{T}$ , giving  $V = \bar{A}\bar{v}\bar{T}$ . Each of these features can be extracted from footage of liquid being poured using computer vision techniques. However, due to our crude estimate of a time-varying process this prediction method has error. The idea of using machine learning to estimate volume is to improve the accuracy of our volume prediction.

## 3 Experiments

To create our data, we filmed over 100 videos of liquid being poured into a bowl ( 10 video samples per volume, with volumes ranging from 1/4 cups (c) to 2 1/4 c, in 1/4 c increments). The experiments were kept highly controlled to minimize error resulting from the setup; we used blue water to enhance contrast, chose a setup that minimized glare and shadowing, and included a ruler to calibrate pixel lengths.

## 4 Features and labels

To process the videos, each film was broken into a series of still frames. Based on the contrast of consecutive image intensities, the frames were used to find the start and stop times of the poured liquid (Figure ??), which was then used to estimate the pour duration. The area was predicted from each video by first identifying the stream width at different heights based on red color intensity variations (Figure ??), which was then squared and averaged to compute the mean area. Velocity was estimated by tracking the front of the falling liquid using edge detection. After processing predictions of  $\bar{A}$ ,  $\bar{v}$ , and  $\bar{T}$  were extracted. We checked to see how each of the features correlated to the known poured volume ( $y$ ) and the other 2 features.

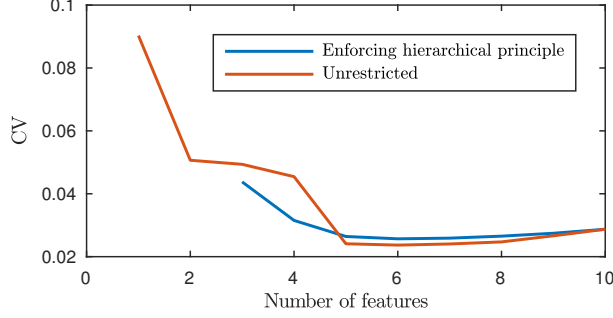


Figure 1: CV error from forward search with LOOCV as a function of feature numbers for an unrestricted search, and a search obeying the hierarchical principle

## 5 Machine learning

### 5.1 Ordinary least squares feature selection

Forward search was applied to identify relevant features for the ordinary least squares (OLS) fit. The original features ( $\bar{A}$ ,  $\bar{v}$ , and  $\bar{T}$ ), all 2nd order terms ( $\bar{A}^2$ ,  $\bar{v}^2$ , and  $\bar{T}^2$ ), and 4 interaction terms ( $\bar{T}\bar{v}$ ,  $\bar{T}\bar{A}$ ,  $\bar{A}\bar{v}$ , and  $\bar{T}\bar{A}\bar{v}$ ) were considered. As features were sequentially added to the model, leave-one-out cross validation (LOOCV) was applied to identify the feature at each step that reduced the generalize cross validation error ( $CV$ ) the most. Here,  $CV$  is defined as

$$CV = \frac{1}{m} \sum_{i=1}^m \left( y^{(i)} - h_{OLS}^{(i)} \right)^2, \quad (2)$$

where  $h_{OLS}^{(i)}$  is the model prediction of  $y^{(i)}$  at  $x^{(i)}$  from an OLS fit using all observation but  $x^{(i)}$ . The generalized cross validation error is shown as a function of the number of model features in Figure 1.

Forward search was applied in two forms. In the first, no restrictions were evoked on which features were contained within the model (red line). In the second, forward search was applied while enforcing the hierarchical principle (blue line), which states that if an interaction or high order term is added to

a model, then so should the main features involved (James et al. [2013]). For example, if  $\bar{T}\bar{A}\bar{v}$  is added to the model, then individual terms for  $\bar{T}$ ,  $\bar{A}$ , and  $\bar{v}$  must also be included. As shown in Figure 3, both forms of the forward search suggest model performance does not significantly increase using beyond five feature. However, the exact features selected depends on the restrictions applied. Enforcing the hierarchical principle selected  $\bar{T}$ ,  $\bar{A}$ ,  $\bar{v}$ ,  $\bar{T}\bar{A}\bar{v}$ , and  $\bar{A}^2$ , while the unrestricted search identified  $\bar{T}\bar{v}$ ,  $\bar{T}\bar{A}$ ,  $\bar{v}$ ,  $\bar{A}$ ,  $\bar{A}^2$  as the most relevant features. We chose the 5 features identified by the hierarchical search for our ordinary least squares model. This choice was made not only to obey hierarchical principle, but also because the selected model contains  $\bar{T}\bar{A}\bar{v}$  which has physical significance.

### 5.2 Physical prediction vs OLS prediction

As a baseline for comparison, volume predictions were also made assuming a physical based model

$$h_{phy}^{(i)} = \alpha T^{(i)} A^{(i)} v^{(i)}, \quad (3)$$

where  $\alpha = \sum_{j=1}^m y^{(j)} / (T^{(j)} A^{(j)} v^{(j)})$  and hence is a constant chosen to minimize the OLS fit for a model containing only  $\bar{T}\bar{A}\bar{v}$ . Physically,  $\alpha$  represents a correction factor that attempts to account for errors in the extracted features and encompasses the required constant for an area prediction (i.e. includes a factor of  $\pi/4$  if the cross-sectional area of the stream is truly a circle). We take  $h_{phy}^{(i)}$  as the baseline estimate for the volume prediction based on purely physics. LOOCV was performed and a  $CV$  error based on Equation (2) with  $h_{phy}^{(i)}$  as the hypothesis was computed as 0.173. Comparing this to the 0.026  $CV$  error from the OLS model with  $\bar{T}$ ,  $\bar{A}$ ,  $\bar{v}$ ,  $\bar{T}\bar{A}\bar{v}$ ,  $\bar{A}^2$  and bias term as features, the model identified by simple forward search reduces the  $CV$  error by more than 85%.

### 5.3 OLS Model evaluation

The 5-featured OLS model was further evaluated by investigating the dataset size affects on model performance. The original dataset was sub-sampled at

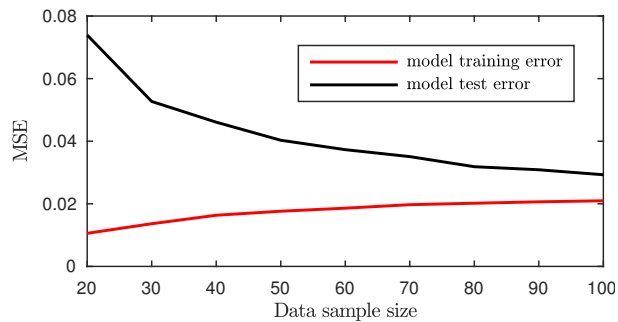


Figure 2: : Model and training error as a function of dataset size.

different sizes, then split into 70% training data, and 30% test data. The mean squared error (MSE) between the predicted and actual volume was then computed for both the test and training data, and is shown in Figure 2. The test and training MSE are converging, however there is currently a gap between the two. Additional data will likely improve the OLS model, and a more sophisticated model all together such as a random forest may reduce model bias and potentially reduce both the training and test MSE.