# C&O 370/CM443 Deterministic OR Models – W11

## Project on Climate Forecasting

Due at the start (9:30-$\epsilon$AM) of class on the due date,
or under the door MC6065 by midnight-$\epsilon$ before the due date.

## Contents

# 1 Part 1 Data Collection and Linear Modelling —— Due Wed. Mar. 2, 2011

## 1.1 Collecting Data (Canadian)

The National Oceanic and Atmospheric Administration (NOAA) collects and archives weather data. [1]

There are many weather stations in Canada, e.g., in the readme file: starting with line 15744 [2] and ending with line 17310 [3]; as well there are many others throughout the file. Please pick a Canadian weather station, preferably in Ontario; and **email** the name of the weather station to the instructor. Download the annual (average daily temperatures) data files for at least 50 years.

## 1.2 Preliminary Regression Model

Let $T_d$ denote the average temperature on day $d$, and let $T = (T_d) \in \mathbb{R}^n$ be the data vector. Let $F_d(a)$ denote the model, where $a \in \mathbb{R}^m$ is a vector of parameters; with $F(a) = (F_d(a)) \in \mathbb{R}^n$. Find the best value for the parameters $a = (a_i)$ that solve the regression problem

$$(RP) \qquad \min_a \|F(a) - T\|_p,$$

where $\|\cdot\|_p$, denotes some norm. For example, with $p = 1$, we get the sum of the absolute deviations:

$$(LAD) \qquad \min_a \sum_d |F_d(a) - T_d| \,.$$

As a sample model, we could use the model composed of three terms:

$$F_d(a) = \begin{cases} (a_0 + a_1 d) & \text{affine/linear trend} \\ + & \\ a_2 \cos(2\pi d/365.25) + a_3 \sin(2\pi d/365.25) & \text{periodic seasons} \\ + & \\ a_4 \cos(a_6 2\pi d/(10.7 \times 365.25)) + a_5 \sin(a_6 2\pi d/(10.7 \times 365.25)) & \text{periodic solar cycle} \end{cases}$$

The three terms model the average temperature as: a constant plus a linear trend; a sinusoidal function with a one-year period representing seasonal changes; and a sinusoidal function with a period approximately equal to 10.7 years to represent the solar cycle.

[1] data format, instructions at: ftp://ftp.ncdc.noaa.gov/pub/data/gsod/readme.txt The list of 9000 weather stations is here: ftp://ftp.ncdc.noaa.gov/pub/data/gsod/ish-history.txt

[2] 710010 99999 CALLAGHAN VALLEY SK CN CA BC CVOW +50133 -123100 +09360 20070424 20100413

[3] 719994 99999 SEDCO 710 CN CA CWQJ +46500 -048500 +00000 19831104 19900626

## 1.3 Solve as a Linear Program

1. Use the (LAD) model and assume that the unknown parameter $a_6$ is fixed at 1, thus fixing the solar-cycle. Solve this problem using an LP model with AMPL.

2. Then allow $a_6$ to vary but give AMPL an initial value near 1, i.e. $\bar{a}_6 \cong 1$. Solve this nonlinear problem using AMPL.

## 1.4 Solve as a Quadratic Program

1. Use the (RP) model with $p = 2$ with the same assumptions as in Section 1.3, i.e. assume that the unknown parameter $a_6$ is fixed at 1, thus fixing the solar-cycle.

2. Then allow $a_6$ to vary but give AMPL an initial value near 1, i.e. $\bar{a}_6 \cong 1$. Solve this nonlinear problem using AMPL, if you can. Discuss the convexity and degree of nonlinearity.

## 1.5 Sensitivity Analysis

Your sensitivity analysis should include at least the following sensitivity to:

1. changing some/all of the current data to new data from a nearby station

2. changing the number of years and/or leaving out some of the years (e.g. the hot part around 2010 and/or the cold part around 1970)

3. perturbing individual data values $T_d$.