STUDY OF THE MAXIMA OF PARTIAL SAMPLES OF
OXIDANT CONCENTRATIONS IN THE BAY AREA

by

YASHASWINI MITTAL¹
Stanford University

WORKING PAPER #1

This draft is intended for limited circulation.

September 1, 1976

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Energy Research and Development Administration, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

There is no objection from the
point of view to the publication,
dissemination or the documentation
listed in this letter.

BROOKHAVEN PATENT GROUP

12/8 1976 By CLP

¹This research was supported by Grant from SIMS

MASTER

Prepared For
THE U.S. ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION
UNDER CONTRACT NO. EY-76-S-02-2874.*000

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED *leg*

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Suppose $\{X_i, i \geq 1\}$ is a mean zero variance one stationary Gaussian sequence and $M_n = \max_{1 \leq i \leq n} X_i$. Let G_n be a subset of $\{1, 2, \dots, n\}$ of size $\lfloor \frac{n}{k} \rfloor$ (e.g. every k^{th} observation), $[x]$ being the integral part of x . Define $m_{n,k} = \max_{i \in G_n} X_i$. The difference of the two maxima $M_n - m_{n,k}$ was studied in [1]. It is shown there that

$$(1) \quad P(c_n(M_n - m_{n,k}) \leq x) \rightarrow \frac{1}{1 + (k-1)e^{-x}}$$

as $n \rightarrow \infty$ where $c_n = (2 \ln n)^{\frac{1}{2}}$. The above result holds under the following conditions on the covariance $r_n \equiv EX_n X_0$ of the process, i.e., either

$$(2) \quad r_n \ln n \rightarrow 0 \text{ as } n \rightarrow \infty$$

or

$$(3) \quad r_n \text{ convex } r_n \rightarrow 0 \text{ such that } r_n \ln n \text{ increases to } \infty \text{ monotonically.}$$

This report is an attempt to apply the above result to the data on oxidant concentrations from nine Bay Area stations during January 1971-December 1973.

We consider here daily averages of the oxidant concentrations in pphm. We will treat this as a Gaussian sample and apply the above results. The choice of daily averages instead of hourly averages was made in view of the Gaussian assumption. Our aim is twofold. The first aim will be to verify if assumptions of Gaussian sequence and (2) or (3) above for oxidant concentrations will be consistent with the observed behavior of the maximum in the past. The second aim will be to forecast the behavior of the maximum assuming that the oxidant concentrations satisfied required assumptions.

For each fixed station we will treat the oxidant observations (daily averages) as being stationary Gaussian with fixed mean m and variance σ^2 . The values of m and σ^2 will be computed from three years of data 1971-73. For nine stations in the Bay Area we have virtually complete data on oxidants in 1971-73. These are chosen for our study. As a step towards our first aim we will treat all the observations of 1971 as a full sample. For example from January 1, 1971 to December 31, 1971, Livermore has $n = 359$ daily observations on oxidants. We will treat this as the full sample $x_1 \dots x_{359}$. Its maximum M_n is observed to be 7.1 pphm. Next we will create various partial samples by observing, say, every other (or third, or fourth, etc.) x_i . The observed maximum of this partial sample will be $m_{n,2}$ (or $m_{n,3}$, $m_{n,4}$, etc.). The observed difference $M_n - m_{n,k}$ is independent of the constant mean but should be normalized by the s.d. 1.09. We compute
$$x_0 = \left(\frac{M_n - m_{n,k}}{\text{s.d.}} \right) c_n \quad \text{and} \quad P_{x_0} = \frac{(k-1)e^{-x_0}}{1+(k-1)e^{-x_0}}.$$
 For $k = 10$ in Livermore observations P_{x_0} is 0.055. Thus if the assumptions made on the oxidants were valid the probability of observing the difference (7-1-5.5) or bigger in the two maxima for $k = 10$ will be 5.5%. We can interpret this as evidence in support of our assumptions at 5% level of significance.

Next we treat the 1971 data as a partial sample and 71-73 data as the full sample. Based on the partial sample we forecast the value of the maximum in the full sample and compare it with the observed value. Thus in Pittsburg we forecast the value 6.8 at 5% and 6.6 at 10% while as the observed max over the three years 71-73 is 6.8. The interpretation of 5% level here is as follows. The observed max of the partial sample is 5.8. If the

assumptions were valid then the probability of observing difference of at least 1 pphm will be 0.05 and hence the 5% critical value for max of 71-73 will be 5.8 ± 1 .

Remarks:

(1) It seems like the data is in fair agreement with assumptions in case of Pleasanthill and Pittsburg stations. The rest of the stations show a wide disparity.

(2) The assumptions of the result (1) do not require choosing alternate values of X_i for a partial sample with $k = 2$. In practice we know that oxidants have a seasonal trend. We try to minimize its effect by choosing alternate values rather than any $\frac{n}{2}$ values.

(3) It's also apparent that oxidants have a yearly trend. The year 1971 must have been a bad year for oxidants. Eight out of nine stations considered achieved their three-year maximum in 1971. Thus 1971 was a particularly unfortunate choice for the study.

(4) The calculations above are of course very naive and are basically done to illustrate the method. If the basic objective was studying the oxidant data then more care should have been taken to adjust for seasonality and the weather dependency of the data.

The next three pages give the results for the nine stations observed. The confidence intervals computed for Pleasanthill and Pittsburg stations have following interpretations. The lower limit (L.L) for fixed k is such that $P\{m_{n,k} < \text{L.L.}\} = 0.1$ and the upper limit (U.L.) is such that $P\{m_{n,k} > \text{U.L.}\} = 0.1 + \frac{1}{k}$ if the assumptions on the process were valid.

Thus e.g. the probability that the maximum $m_{n,4}$ of the partial sample for $k=4$ will be less than 5.05 is 0.1 and bigger than 5.95 is 0.35.

References

- [1] Mittal, Yashaswini, Maxima of partial samples in Gaussian sequences.
Submitted to Ann. Prob.

Daily Averages of Oxidants (pphm) for 1971

1. Pleasanthill: Valid Observations (n) = 344
Observed 1971 maximum = 6.1

k	Confidence interval		Obs. Max	P_{x_0}
	L.L.	U.L.		
2	5.40	5.97	6.1	0.5000
3	5.18	5.99	5.6	0.2943
4	5.05	5.95	6.1	0.7500
6	4.89	5.92	4.9	0.1040
10	4.70	5.8	5.1	0.2812

Three-year forecasting:

5% critical value: 7.26

10% critical value: 7.02

Observed three year max

6.1

2. Pittsburg: Valid observations (n) = 364
Observed 1971 maximum = 5.8

k	Confidence interval		Obs. Max	P_{x_0}
	L.L.	U.L.		
2	5.14	5.68	5.8	0.5000
3	4.93	5.70	5.2	0.2130
4	4.81	5.66	4.7	0.0710
6	4.66	5.63	4.8	0.1510
10	4.48	5.52	5.8	0.9000

Three-year forecasting:

5% critical value: 6.8

10% critical value: 6.6

Observed three-year max

6.8

Daily Averages of Oxidants (pphm) for 1971

San Rafael: n = 365

Observed 1971 max: 4.8
Observed 71-73 max: 4.8

k	Obs. Max	P_{x_0}
2	4.8	0.500
3	4.4	0.214
4	4.8	0.750
6	4.4	0.406
10	4.1	0.216

Forecasting:

5% critical value: 5.47
10% critical value: 5.33

San Leandro: n = 358

Observed 1971 max: 7.5
Observed 71-73 max: 7.5

k	Obs. Max	P_{x_0}
2	4.6	0
3	7.5	0.670
4	4.1	0
6	3.8	0
10	4.5	0

Forecasting:

5% critical value: 8.47
10% critical value: 8.27

Redwood City: n = 365

Observed 1971 max: 7.7
Observed 71-73 max: 7.7

k	Obs. Max	P_{x_0}
2	7.7	0.500
3	5.0	0
4	7.7	0.750
6	5.0	0
10	5.4	0

Forecasting:

5% critical value: 8.51
10% critical value: 8.34

Oakland:

Observed 1971 max: 8.9
Observed 71-73 max: 8.9

k	Obs. Max	P_{x_0}
2	6.8	0
3	4.4	0
4	6.8	0
6	4.4	0
10	2.9	0

Forecasting:

5% critical value: 9.25
10% critical value: 9.10

Daily Averages of Oxidants (pphm) for 1971

Livermore: n = 359

Observed 1971 max: 7.1

Observed 71-73 max: 7.1

k	Obs. Max	P_{x_0}
2	7.1	0.500
3	5.3	0.006
4	5.5	0.019
6	5.3	0.017
10	5.5	0.055

Forecasting:

5% critical value: 8.16

10% critical value: 7.94

Fremont: n = 357

Observed 1971 max: 8.5

Observed 71-73 max: 8.5

k	Obs. Max	P_{x_0}
2	8.5	0.500
3	7.1	0.019
4	8.5	0.750
6	4.9	0
10	6.8	0.023

Forecasting:

5% critical value: 9.51

10% critical value: 9.30

San Francisco: n = 361

Observed 1971 max: 5.8

Observed 71-73 max: 5.8

k	Obs. Max	P_{x_0}
2	2.0	0
3	5.0	0.020
4	2.0	0
6	5.0	0.049
10	2.3	0

Forecasting:

5% critical value: 6.38

10% critical value: 6.26