A Note on the Statistical Correction of Prediction Equations[1,2]

CHARLES E. SCHEMM[3] AND ALAN J. FALLER

Institute for Physical Science and Technology, University of Maryland, College Park 20742

29 March 1977 and 16 June 1977

ABSTRACT

Two modifications of the procedure for applying statistical corrections to the prediction equations discussed in Faller and Schemm (1977) are considered. In the first, the constant terms which normally appear in the regression equations are suppressed, leading to improved extended range STAT predictions. In the second, a noncentered advection term which gave exceptionally high correlations with one time step prediction errors is tested; however, we have been unable to use this type of term successfully in extended predictions. These results indicate the need for careful selection and testing of parametric terms for the proposed statistical corrections.

1. Introduction

In a recent paper (Faller and Schemm, 1977)[4] the authors discussed a statistical procedure for determining corrections to be applied to numerical prediction equations at the end of each time step. The statistical procedure is in essence an attempt, using multiple regression, to correlate errors after many one time step (OTS)[5] predictions with parametric terms (P terms) evaluated at the beginning of each time step and thereby to obtain coefficients for these P terms which may then be used to modify the model equations. The method was tested using a two-dimensional model system with equations similar to those discussed by Burgers (1950), and significant improvements in extended range predictions were obtained when the statistically modified (STAT) equations as opposed to the unmodified finite-difference (FIDI) equations were used. The STAT-II system of equations, where different correction coefficients were computed and used at each grid point, gave generally better results than did the spatially uniform STAT-I equations.

Unfortunately, however, all STAT predictions for the model in which the total energy was allowed to vary were characterized by excessive growth in the perturbation kinetic energy which, when unchecked, led eventually to computational instability. This problem was particularly acute in the case of STAT-II, where large variations between the STAT coefficients at adjacent grid points were thought to be a contributing factor.

With both the STAT-I and STAT-II systems of equations the best extended range (60 time step) predictions were always obtained when only the three "basic" terms, defined as those terms already in the FIDI equations, were used as parametric terms. As reported in FS, the inclusion in the multiple regressions of nonlinear viscosities, divergence terms and upwind advection terms always led to a higher percentage of explained variance for one time step. The corresponding STAT-I predictions were always superior through the first few time steps to comparable STAT-I predictions using only the basic parametric terms. After many time steps, however, the STAT-I predictions with additional parametric terms became generally inferior.

This note presents results from experiments in which the constant terms a_0 and b_0 were suppressed in the multiple regressions. These experiments are designated STAT-I-NC and STAT-II-NC (no constant terms). Also discussed are extensive tests of a special parametric term which correlated very well with the OTS errors but which, when used in the STAT-I or the STAT-I-NC forecasts, led inexplicably to poor cumulative predictions.

2. Effects of the constants

Unless the variates have zero mean, in which case the constant term is automatically zero, it is not common practice to omit the constant term in a linear regression equation. It is possible, however, to write a regression equation without the constant term and to determine the regression coefficients by the usual method of least squares. Recalculation of several experiments with the constant terms suppressed indicates that their inclusion was responsible in part for the undesirable growth of energy that was frequently observed in the STAT predictions of FS.

Table 1 shows the multiple correlation coefficients $\rho_{m,U}$ and $\rho_{m,V}$, the constants a_0 and b_0, and factors f_1-f_3 and g_1-g_3 which multiply the basic terms in the modified U and V prediction equations of STAT-I and

[1] This research has been supported in part by the National Science Foundation under Grant ATM-74-24132 A01. We also gratefully acknowledge the support of the Computer Science Center of the University of Maryland.

[2] Publication No. 77-164 of the Meteorology Program.

[3] Present affiliation: Applied Physics Laboratory, The Johns Hopkins University, Laurel, Md. 20810.

[4] Hereafter referred to as FS.

[5] Explanations of frequently used abbreviations may be found in the Appendix.

Unauthenticated | Downloaded 12/08/24 12:52 AM UTC

TABLE 1. Results of multiple regressions for the basic STAT-I and STAT-I-NC experiments using Reynolds number Re=225 and amplification coefficient $A=0.125$.

	STAT-I	STAT-I-NC
f_1	1.205	1.191
f_2	0.850	0.773
f_3	0.653	1.018
a_0	0.00117	0
$\rho_{m,U}$	0.396	0.387
g_1	1.189	1.188
g_2	0.980	0.979
g_3	0.318	0.319
b_0	0.00010	0
$\rho_{m,V}$	0.338	0.338

STAT-I-NC. As in FS the subscripts 1, 2 and 3 refer to advective, diffusive and drag terms, respectively. The major change when a_0 and b_0 were constrained to be zero was in the coefficient f_3 which multiplies the $-kU$ term. Results from 60 time step integrations of the STAT-I-NC equations are shown in Fig. 1 along with comparable FIDI and STAT-I results. Predictability here is judged by the same four measures utilized in FS: E_{rms}, the root-mean-square error of prediction; ρ_{pc}, the correlation between predicted and correct velocities; $R_{\bar{U}}$, the ratio of predicted to correct mean flow; and $R_{K'}$, the ratio of predicted to correct average perturbation kinetic energy. Of particular note is the improvement in all measures except $R_{\bar{U}}$. This grew more rapidly in STAT-I-NC than in STAT-I.

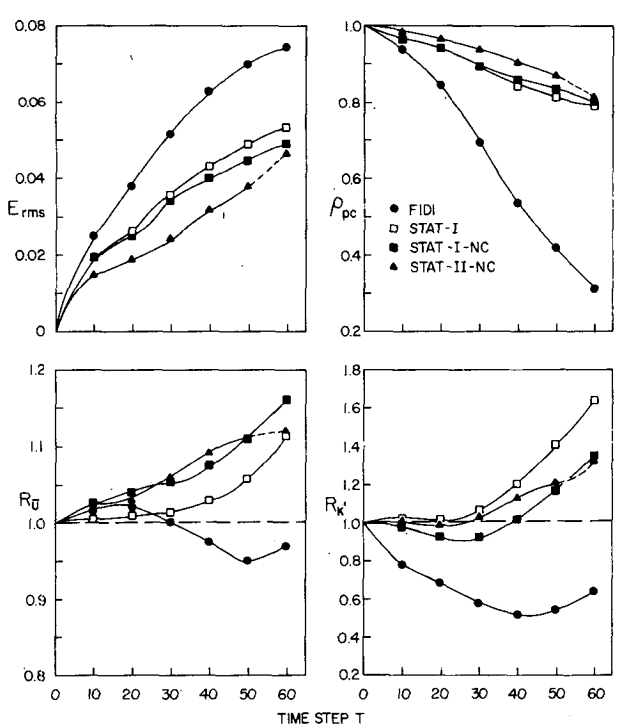

FIG. 1. Average values of E_{rms}, ρ_{pc}, $R_{\bar{U}}$ and $R_{K'}$ for FIDI, STAT-I, STAT-I-NC and STAT-II-NC predictions with Re=225 and $A=0.125$. Averages are generally for six integrations starting at $M=104$, 143, 163, 203, 223 and 263. Because one of the STAT-II-NC integrations became unstable between $T=50$ and $T=60$, five-point averages are shown at $T=60$ for this experiment.

The relative effects upon $R_{\bar{U}}$ and $R_{K'}$ of retaining or omitting the constant terms can be partially understood by analyzing the tendency equation

$$\frac{\partial U}{\partial t}=-f_3kU-\frac{a_0}{\Delta t}, \tag{1}$$

where Δt is the constant time step interval. This equation is a simplified differential form of Eq. (12) of FS. It includes the constant term $a_0/\Delta t$ and the modified drag term $-f_3kU$ but omits the advection and diffusion terms. Writing $U=\bar{U}+U'$ in (1) and averaging over all grid points yields the tendency equation for $\bar{U}$, i.e.,

$$\frac{\partial \bar{U}}{\partial t}=-[f_3k+(a_0/\bar{U}\Delta t)]\bar{U}. \tag{2}$$

The tendency equation for $K'=\overline{U'^2}/2$ is

$$\frac{\partial K'}{\partial t}=-2f_3kK' \tag{3}$$

which does not involve a_0.

From (2) and (3) the fractional rates of decay $-\bar{U}^{-1}\partial\bar{U}/\partial t$ and $-K'^{-1}\partial K'/\partial t$ are $(f_3k+a_0/\bar{U}\Delta t)$ and $2f_3k$, respectively. Using $\Delta t=0.06$, $\bar{U}=0.13$ and values of a_0 and f_3 from Table 1, the decay rates are as given in Table 2. Since the FIDI results ($a_0=0$) would give fractional decay rates of f_3k and $2f_3k$ for $\bar{U}$ and K', respectively, Table 2 indicates that inclusion of the constant term in STAT-I reduced the rate of decay of K' significantly—a result consistent with the excessive growth of perturbation kinetic energy noted in the STAT-I experiments of FS. The fractional rates of decay for STAT-I-NC, however, were both close to 0.40; and hence the improved results for $R_{K'}$. The excessive growth of $\bar{U}$, however, is not explained by this analysis.

It was found in FS that further improvements in predictability could be obtained when spatially variable regression coefficients were used, and we have tested omission of the constant terms in these cases as well. Although one of the six STAT-II-NC integrations became computationally unstable between $T=50$ and $T=60$, these predictions were significantly more

TABLE 2. Fractional rates of decay for $\bar{U}$ and K' for simplified FIDI, STAT-I and STAT-I-NC tendency equations.

	FIDI	STAT-I	STAT-I-NC
$-\bar{U}^{-1}\partial\bar{U}/\partial t$	0.40	0.4112	0.4072
$-K'^{-1}\partial K'/\partial t$	0.40	0.2612	0.4072

Unauthenticated | Downloaded 12/08/24 12:52 AM UTC

accurate than the STAT-I-NC forecasts through 50 time steps (see Fig. 1); and they were less susceptible to the spurious energy growth which rendered the comparable STAT-II forecasts (with nonzero constants) useless after about 30 time steps.

When the regression equations were obtained by the STAT-II-S (smoothed) method all integrations were stable through 60 time steps. Table 3 compares results of the STAT-II-S and STAT-II-S-NC experiments. Based on average values of ρ_{pc}, the predictions made without the constant terms were at first slightly inferior, but by $T=40$ they surpassed the forecasts with the constant terms included. This improvement probably was related to an improved prediction of the kinetic energy as indicated by $R_{K'}$.[6]

TABLE 3. Average values of ρ_{pc} and $R_{K'}$ for comparable STAT-II-S and STAT-II-S-NC experiments with Re=400 and $A=0.114$.

T	STAT-II-S	STAT-II-S-NC
10	0.976	0.974
20	0.961	0.957
30	0.928	0.926
40	0.892	0.898
50	0.855	0.877
60	0.820	0.840
10	0.985	0.945
20	1.001	0.895
30	1.097	0.920
40	1.224	1.005
50	1.390	1.152
60	1.547	1.278

3. A special parametric term

As reviewed in the Introduction, several parametric terms of possible interest gave generally poor results in extended predictions. In an attempt to understand this behavior we have extensively tested a particular form of noncentered advection that gives exceptionally high correlations with the OTS errors in a concerted attempt to improve the predictability using this type of term.

The representative advection term UU_x may be written in finite-difference form as

$$UU_x \approx (\tfrac{1}{2}-\alpha)U_{I,J}\left(\frac{U_{I+1,J}-U_{I,J}}{\Delta x}\right) + (\tfrac{1}{2}+\alpha)U_{I,J}\left(\frac{U_{I,J}-U_{I-1,J}}{\Delta x}\right), \quad (4)$$

where I and J are indices for grid points in the x and y directions, respectively. If we take $\alpha=\frac{1}{2}$ when $U_{I,J}>0$ and $\alpha=-\frac{1}{2}$ when $U_{I,J}<0$, we have the common upwind difference representation which was tested in FS. The value $\alpha=0$ gives the usual centered advection scheme.

To provide a smoothly varying upwind advection we take α proportional to the speed of the flow. With $\alpha=c_1U_{I,J}\Delta t/\Delta x$, where c_1 is a nondimensional constant (to be determined later in a multiple regression), Eq. (4) becomes

$$UU_x \approx U_{I,J}\left(\frac{U_{I+1,J}-U_{I-1,J}}{2\Delta x}\right) - c_1\Delta t U_{I,J}^2\left(\frac{U_{I+1,J}-2U_{I,J}+U_{I-1,J}}{\Delta x^2}\right). \quad (5)$$

The first term on the right of (5) represents centered advection. The second term is in a form reminiscent of a higher order correction to a first-order time-stepping method. For example, using the method of analysis devised by Hirt (1968), Caponi (1976) included artificial viscous terms of the form $\frac{1}{2}U^2\Delta tU_{xx}$ to correct for false dissipative effects introduced by forward time-differencing. One might not expect such a term to be significant in the present model because Adams-Bashforth time differencing, used in these calculations, is accurate to second order in Δt.[7] Nevertheless, the use of parametric terms of this type have yielded multiple correlation coefficients ρ_m typically in excess of 0.6. Multiple correlations without these new terms were generally less than 0.4 (FS, Table 1).

Although the parametric terms in question were designed as modifications of the normal advection terms, in form they more closely resemble nonlinear viscous terms. Hence, we have been ambivalent as to how we should refer to them: as advective terms or as dissipative terms. Because of the need to frequently refer to these terms by name, we have denoted them simply as C terms. In the present study four C terms arise, these being proportional to U^2U_{xx} and V^2U_{yy} in the U momentum equation and to U^2V_{xx} and V^2V_{yy} in the V equation.

A wide variety of trial computations using the C terms included the following variations:

1) The straightforward C terms can have nonzero spatial averages that could lead to changes of the mean flow. To avoid this possibility, spatial averages were subtracted from the point values to give parametric terms of the form $(U^2U_{xx}-\overline{U^2U_{xx}})$.

2) The C terms were calculated using both the total U and its fluctuating component U', with terms of the form $(U'^2U_{xx}-\overline{U'^2U_{xx}})$ being designated as the "modified" C terms.

3) The regression equations were calculated by two different approaches, here referred to as the "multiple" and the "stagewise" methods. In the former a single regression equation which included the C terms was calculated for each of the component equations. In the

[6] A further discussion of these results and of additional comparative experiments may be found in Schemm and Faller (1976).

[7] In Lilly (1965) the accuracy of the Adams-Bashforth scheme is considered in comparison with that of other methods.

Unauthenticated | Downloaded 12/08/24 12:52 AM UTC

"stagewise" method the usual STAT-I-NC regression equations were obtained, new one-time-step predictions were carried out, and the new OTS errors were correlated with the C terms alone, thus providing a second regression equation.

4) Damping of the regression coefficients with time—a technique used successfully in FS—was applied here to the basic terms and to the C terms as well.

5) The mean flow $\bar{U}$ and/or the fluctuating kinetic energy K' were adjusted by keeping $R_{\bar{U}}$ and $R_{K'}$ equal to unity or equal to the values obtained from the FIDI or STAT-I-NC predictions as appropriate. To correct $\bar{U}$ to a specified time variation, a spatially uniform value was subtracted from each $U_{i,j}$ after each time step. To adjust the perturbation kinetic energy, U' and V' were multiplied by the constant necessary to give the selected value of K' at each step.

Many combinations of the above were tried in experiments with the C terms in attempts to obtain predictions better through 60 time steps than those of the basic STAT-I-NC. Examples of the high "multiple" and "stagewise" correlations are shown in Table 4. Predictions using the C terms were correspondingly excellent for one or more time steps, but after about $T=10$ they generally became worse than the basic STAT-I-NC forecasts. Among the best results were those computed using the modified C terms with coefficients determined by the "stagewise" regression procedure. When K' (but not $\bar{U}$) was constrained to equal that of the basic STAT-I-NC predictions, this system gave the best 60 time step correlations that we were able to obtain using the C terms, but these results still were not the equal of the basic STAT-I-NC predictions. Damping some or all of the regression coefficients was not found to be helpful. The trends of ρ_{pc} and $R_{K'}$ for three relevant experiments are shown in Table 5.

4. Conclusions

The reason for the poor performance of the C terms is not at all clear. If parametric terms could be found that produced a perfect correlation with the OTS errors (and if the values of $R_{K'}$ and $R_{\bar{U}}$ were kept at unity), it is apparent that predictions starting with the "correct" data would be accurate for all time. Accordingly, one might assume that among various possible prediction equations those that produced the highest correlations with "correct" data after one time step would also be the preferred equations for longer range predictions. Unfortunately, these attempts to make extended predictions with the C terms clearly indicate that this would be an unwarranted assumption. Results presented in Section 2, however, have shown that distinct improvements in predictability can be gained when the constant terms which normally appear as output from the multiple regressions are suppressed.

TABLE 4. Correlation coefficients ρ_m for the "multiple" and "stagewise" regressions using the modified C terms. In the stagewise case only the basic terms are included in the first regressions and only the C terms are included in the second regressions.

	Multiple regression	
	$\rho_{m,U}$	$\rho_{m,V}$
All terms	0.668	0.549
	Stagewise regressions	
	$\rho_{m,U}$	$\rho_{m,V}$
First regression	0.387	0.338
Second regression	0.353	0.260

TABLE 5. Results of STAT-I-NC integrations using the modified C terms with coefficients computed by the "stagewise" method. Column 1 presents results obtained with no energy adjustment, and column 2 shows results obtained when K' is adjusted after each time step to that of the basic STAT-I-NC predictions. The basic STAT-I-NC results are shown for comparison in column 3.

		(1)	(2)	(3)
		C terms included		Basic terms
	T	(no adjustment)	(K' adjusted)	only
ρ_{pc}	5	0.975	0.978	0.973
	10	0.958	0.967	0.966
	20	0.904	0.926	0.939
	30	0.818	0.858	0.892
	40	0.742	0.803	0.859
	50	0.702	0.769	0.824
	60	0.660	0.740	0.800
$R_{K'}$	5	0.877	0.999	0.999
	10	0.822	0.984	0.984
	20	0.748	0.928	0.928
	30	0.718	0.924	0.924
	40	0.756	1.023	1.023
	50	0.925	1.173	1.173
	60	1.059	1.352	1.352

APPENDIX

Frequently Used Notation

OTS	One time step
P terms	Parametric terms whose coefficients are determined by multiple regression
FIDI	Finite-difference equations before addition of the P terms
STAT-I	Statistically modified equations with spatially constant regression coefficients
STAT-II	Same as STAT-I but with spatially variable regression coefficients
STAT-II-S	Smoothed STAT-II equations for which regression coefficients were calculated from data over a 3×3 array of mesh points
STAT-I-NC	No constant terms included in the regression equations used to form the STAT–I equations

Unauthenticated | Downloaded 12/08/24 12:52 AM UTC

E_{rms} The root-mean-square error of prediction

ρ_{pc} Correlation between predicted and "correct" velocities

$R_{\overline{U}}$ Ratio of predicted to "correct" mean flow

$R_{K'}$ Ratio of predicted to "correct" fluctuating energy

ρ_m Multiple correlation coefficient of the OTS prediction errors with the parametric terms.

REFERENCES

Burgers, J. M., 1950: The formation of vortex sheets in a simplified type of turbulent motion. *Proc. Roy. Neth. Acad. Sci.*, **53**, 122–133.

Caponi, E. A., 1976: A three-dimensional model for the numerical simulation of estuaries. *Advances in Geophysics*, Vol. 19, Academic Press, 189–310.

Faller, A. J., and C. E. Schemm, 1977: Statistical corrections to numerical prediction equations. II. *Mon. Wea. Rev.*, **105**, 37–56.

Hirt, C. W., 1968: Heuristic stability theory for finite difference equations. *J. Comput. Phys.*, **2**, 339–355.

Lilly, D. K., 1965: On the computation stability of numerical solutions of time-dependent nonlinear geophysical fluid dynamics problems. *Mon. Wea. Rev.*, **93**, 11–26.

Schemm, C. E., and A. J. Faller, 1976: On the effect of the constant in multiple regressions used for correcting numerical forecast equations. Tech. Note BN 844, Inst. Phys. Sci. and Tech., University of Maryland, 22 pp.

U.S. POSTAL SERVICE

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

(Required by 39 U.S.C. 3685)

1. TITLE OF PUBLICATION		2. DATE OF FILING
MONTHLY WEATHER REVIEW		9/28/77
3. FREQUENCY OF ISSUE	A. NO. OF ISSUES PUBLISHED ANNUALLY	B. ANNUAL SUBSCRIPTION PRICE
MONTHLY	12	$60. & $20. members

4. LOCATION OF KNOWN OFFICE OF PUBLICATION *(Street, City, County, State and ZIP Code) (Not printers)*
45 Beacon St., Boston, Suffolk, Mass. 02108

5. LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*
45 Beacon St., Boston, Mass. 02108

6. NAMES AND COMPLETE ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR

PUBLISHER *(Name and Address)*
American Meteorological Society, 45 Beacon St., Boston, Mass. 02108

EDITOR *(Name and Address)*
Dr. Chester W. Newton, NCAR, P.O. Box 3000, Boulder, Co. 80307

MANAGING EDITOR *(Name and Address)*
Dr. Kenneth C. Spengler, American Meteorological Society, 45 Beacon St., Boston, Ma.

7. OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given.)*

NAME	ADDRESS
American Meteorological Society	45 Beacon St., Boston, Ma. 02108

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

NAME	ADDRESS
None	

9. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 132.122, PSM)*
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

☒ HAVE NOT CHANGED DURING PRECEDING 12 MONTHS ☐ HAVE CHANGED DURING PRECEDING 12 MONTHS *(If changed, publisher must submit explanation of change with this statement.)*

10. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES PRINTED *(Net Press Run)*	3,124	3,152
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	13	10
2. MAIL SUBSCRIPTIONS	2,152	2,730
C. TOTAL PAID CIRCULATION *(Sum of 10B1 and 10B2)*	2,165	2,740
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	27	21
E. TOTAL DISTRIBUTION *(Sum of C and D)*	2,192	2,761
F. COPIES NOT DISTRIBUTED 1. OFFICE USE, LEFT OVER, UNACCOUNTED, SPOILED AFTER PRINTING	932	391
2. RETURNS FROM NEWS AGENTS	---	---
G. TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	3,124	3,152

11. I certify that the statements made by me above are correct and complete. SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER: Kenneth C. Spengler, Executive Director

12. FOR COMPLETION BY PUBLISHERS MAILING AT THE REGULAR RATES *(Section 132.121, Postal Service Manual)*

39 U. S. C. 3626 provides in pertinent part: "No person who would have been entitled to mail matter under former section 4359 of this title shall mail such matter at the rates provided under this subsection unless he files annually with the Postal Service a written request for permission to mail matter at such rates."

In accordance with the provisions of this statute, I hereby request permission to mail the publication named in Item 1 at the phased postage rates presently authorized by 39 U. S. C. 3626.

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER: Kenneth C. Spengler, Executive Director

PS Form 3526 Apr. 1976 *(Page 1)* (See instructions on reverse)

Unauthenticated | Downloaded 12/08/24 12:52 AM UTC