AFFDL-TR-68-76

PILOT RATING TECHNIQUES FOR THE ESTIMATION AND EVALUATION OF HANDLING QUALITIES

JOHN D. McDONNELL

Systems Technology, Inc.

TECHNICAL REPORT AFFDL-TR-68-76

DECEMBER 1968

This document has been approved for public release and sale; its distribution is unlimited.

AIR FORCE FLIGHT DYNAMICS LABORATORY
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

PILOT RATING TECHNIQUES FOR THE ESTIMATION AND EVALUATION OF HANDLING QUALITIES

JOHN D. McDONNELL

FOREWORD

This report represents an effort to improve the validity and usefulness of the pilot rating in estimating and evaluating vehicle handling qualities. The reported research was sponsored by the Air Force Flight Dynamics Laboratory, Air Force Systems Command, under Contract AF 33(615)-3960. The work was performed at the Systems Technology, Inc., Hawthorne, California, facility. The FDL project monitor was Mr. Paul E. Pietrzak. The principal investigator and STI project engineer, John D. McDonnell, planned the approach and is listed as the author.

The author wishes to express his acknowledgment and thanks to D. T. McRuer and I. L. Ashkenas, STI technical directors, for their contributions, both general and detailed, and for their continuous encouragement. Dr. R. O. Besco of American Airlines contributed many ideas detailing the psychometric methodology, and Major Donald Kutyna, USAF, was a willing and cheerful pilot in the sometimes tedious simulation experiments. Thanks are due Miguel deVirgilio, Wallace Tope, and Paul Sutton for their outstanding help in the simulation experiments and digital programming. Finally, thanks are extended to those 63 professionals who took the time to complete the rating scale survey. Their contribution was invaluable.

This technical report has been reviewed and is approved.

C. B. WESTBROOK
Chief, Control Criteria Branch
Flight Control Division
Air Force Flight Dynamics Laboratory

ABSTRACT

Although rating scales of varied forms have been widely used to estimate and evaluate handling qualities over the past decade, a number of deficiencies in both method and data base have been apparent. This investigation was aimed at overcoming many of these deficiencies by attempting to resolve the difficulties experienced with rating scales themselves, and by extending and adding to already existing relationships between ratings and pilot/vehicle system parameters.

Rating scales have come under increasing criticism for problems related to wording ambiguity, the dual mission character of some scales, the nonuniformity in the distribution of descriptors across the scale, and the misuse of scales which has occurred when ratings have been averaged. Psychometric methods provide an approach to these problems, and in this study were used to scale several phrases descriptive of vehicle handling qualities. Thus, quantitative characteristics were derived for contemporary scales through the use of a scaling technique known as the "Method of Successive Intervals," where data for the method were obtained from a survey experiment.

An experiment was conducted which added to available data relating Cooper ratings and pilot/vehicle parameters, and which also tested some potential alternate scale candidates. The correlation results indicate that ratings are probably based on performance and the degree of difficulty experienced in maintaining the performance. The difficulty is most easily represented by the pilot equalization required and the vehicle stick characteristics.

CONTENTS

FIGURES

Page

Page

TABLES

SYMBOLS

a_y	Lateral acceleration; generally measured at a distance l_x from the c.g., $a_y \equiv a_{y_{c.g.}} + l_x\dot{r}\cos\alpha_o + l_x\dot{p}\sin\alpha_o$
a_z	Vertical acceleration
b	Wing span
c	Mean aerodynamic chord
c(t)	Operator output time function, limb position
C_n	Yawing moment coefficient, (Yawing moment)/(1/2)$\rho U_o^2 Sb$
C_{n_p}	$\partial C_n/\partial(pb/2U_o)$
C_{n_δ}	Yaw control effectiveness, $\partial C_n/\partial\delta$
dB	Decibel $\equiv 20 \log_{10} \lvert\ \rvert = \lvert\ \rvert_{dB}$
e_i	Fourier coefficient of error at ith frequency
e(t)	Error time function
F(t)	Limb-applied force
g	Acceleration due to gravity
h	Altitude
i(t)	Forcing function time function
I_x, I_y, I_z	Moments of inertia about the X, Y, and Z axis, respectively
I_{xz}	Product of inertia in XZ plane
jnd	Just noticeable difference, the difference in stimulus magnitude which is detected by a subject 50 percent of the time
$j\omega$	Imaginary part of the complex variable, $s = \sigma \pm j\omega$
K_B	Best controlled element gain
K_c	Controlled element gain
K_p	Human pilot gain
l_x	Distance along the fuselage longitudinal reference axis from the c.g., positive forward

L	1/2 (Slope of $\|Y_p\|_{dB/decade})_{\omega=\omega_c}$
m	Mass
m(t)	System output time function
N_i'	$[N_i + (I_{xz}/I_z)L_i]/[1-(I_{xz}^2/I_xI_z)]$, $i = p, r, \beta$, etc.
NM	Neuromuscular tension of triceps
N_p	$\rho SU_o b^2 C_{n_p}/4I_z$
N_r	$\rho SU_o b^2 C_{n_r}/4I_z$
N_δ	$\rho SU_o^2 bC_{n_\delta}/2I_z$
p	Roll rate, angular velocity about the X axis, positive right wing going down
q	Generalized output of the system; or pitch rate, angular velocity about the Y axis, positive nose going up
r	Yaw rate, angular velocity about the Z axis, positive nose going right
R	Cooper or Cornell rating
R_o	Observed score
R_T	Transformed score, true rating
s	Complex variable, $s = \sigma \pm j\omega$; Laplace transform variable
s	Discriminal dispersion, i.e., standard deviation of responses on ψ continuum
S	Wing area
t	Time
t'	Computed student's t statistic
t_c	Sample student's t statistic
t_r	Rise time, time required for response to a step input to reach 90 percent of its final value
T	Time constant, particularized by subscript
T_I	General lag time constant of human pilot describing function
T_L	General lead time constant of human pilot describing function

T_R	Run length, roll subsidence time constant
u	Output motion quantity (linear perturbed velocity along the X axis)
U_o	Linear steady-state velocity along the X axis
v	Linear perturbed velocity along the Y axis
w	Output motion quantity (linear perturbed velocity along the Z axis
W	Weight
y	Lateral flight path displacement
$Y_c(j\omega)$	Controlled element (machine and display) transfer function
Y_p	Pilot describing function
α	Low frequency phase approximation parameter
α	w/U_o, perturbed angle of attack under no-wind condition
β	v/U_o, sideslip angle under no-wind condition
γ	Vertical flight path angle
δ	Control deflections, particularized by subscript
δ_{e_i}	Fourier coefficient of elevator output at ith frequency
$\Delta(s)$	Denominator of airframe transfer functions; characteristic equation when set equal to zero
ζ	Damping ratio of linear second-order transfer function quantity, particularized by subscript
θ	Pitch angle
θ_i	Fourier coefficient of pitch at ith frequency
λ	Unstable root position of a particular Y_c
λ_c	Critical score of loading task
λ_s	Secondary loading task score
ρ	Mass density of air
ρ_{ac}	Relative remnant at pilot's output, $\sqrt{1-\overline{n^2}/\overline{c^2}}$
σ	Standard deviation
σ	The real portion of the complex variable $s = \sigma \pm j\omega$

σ_i rms value of the forcing function

τ_e Effective time delay

φ Roll angle

φ_m Phase margin, $\varphi_m = 180^\circ - \measuredangle Y_{\text{open loop}}$ where $|Y_{O.L.}| = 1$

Φ_{cc} Power spectra of pilot output

Φ_{ee} Power spectra of the error

Φ_{ii} Power spectra of the input

Φ_{nn} Closed-loop remnant power spectra

ψ The psychological continuum, aircraft heading angle

ω Undamped natural frequency of a second-order mode, particularized by subscript

ω_b System bandwidth, i.e., frequency at which $|Y_pY_c|/|Y_pY_c|_{\omega=0} = -3\,\text{dB}$

ω_c System crossover frequency, i.e., frequency at which $|Y_pY_c| = 1$

ω_i Forcing function bandwidth

ω_n Frequency of forcing function sinusoidal component

ω_r Frequency of component of input

Subscripts

a Aileron, aileron axis transfer functions

c Command; crossover; controlled element

d Dutch roll

e Elevator; system error

e Effective

o Initial value

p Phugoid

r Rudder; yaw axis transfer functions

R Roll subsidence

sp	Short period
T	Throttle
θ	Pitch transfer functions
φ	Roll transfer functions

SECTION I

INTRODUCTION

The suitability of a manually controlled vehicle to serve its intended purpose is ultimately assessed by a series of judgments. Perhaps the most difficult portion of such an assessment is the evaluation of the vehicle's handling qualities, which play such a key role in the overall suitability of the vehicle, and yet have in the past been perplexing even to define satisfactorily. Cooper (1) originally proposed a handling qualities rating scale which found wide acceptance. Subsequently, modifications and variations were proposed and used in special applications [for example, Harper (2)]. As experience with rating scales accumulated, the amount of information desired from them also increased, resulting in inconsistencies and confusion from the interpretation and use of the ratings. The problem was further compounded when the engineer, who was charged with producing a suitable vehicle, faced the task of extrapolating the rating data to increasingly complex vehicle systems.

The purpose of this study is to attempt to overcome some of the rating scale difficulties encountered in the decade of experience with the scales, to structure the evaluation problem in terms that can be applied to future pilot/vehicle systems, and to extend our knowledge of the causal factors of pilot ratings, i.e., the relationship between ratings and pilot/vehicle system parameters.

The study naturally divides itself into two parts. Many of the problems with contemporary scales are independent of a specific rating situation, and are related to the semantics, definitions, and structure of the scale itself. These problems are investigated in Section II, where it is attempted to clearly define handling qualities, and in Section III, where psychological measurement techniques are used to evaluate the utility of rating scales in general, and to obtain numerical data for specific scale terminology.

The second part of this study is concerned with the search for the physical causes of a pilot's opinion of a vehicle. Section IV describes a simulation experiment in compensatory tracking, where ratings were taken at the same time that parameters of interest were measured. Section V presents and discusses the results of the experiment, and Section VI reiterates the major findings and conclusions of the study, and makes several recommendations regarding the use and future of pilot ratings.

SECTION II

RATING SCALE BACKGROUND AND TASK ELEMENTS DEFINITION

In the process of measuring and evaluating pilot/vehicle performance, it is necessary, as one facet of the investigation, to measure operator opinion. These subjective measures are in fact the ultimate evaluation of the system and consequently are foremost in the designer's mind throughout vehicle development. Unfortunately, the current connections between pilot ratings, pilot behavior, and vehicle characteristics are, at best, highly qualitative. This situation has not improved as vehicles and associated pilot/vehicle handling qualities considerations have steadily increased in complexity, for then the difficulties with existing rating scales and subjective measures become still more obscure. As introductory background to existing rating scales, the difficulties providing much of the motivation for the current work will be outlined.

A. DIFFICULTIES WITH EXISTING RATING SCALES

Several scales for use in handling quality ratings exist, the most recent and widely used containing ten probably unequal divisions. Primacy among these can be claimed by a scheme proposed by Cooper (1) and extensively used by the NACA and NASA. The scale is shown in Table I. In spite of its ten subdivisions, it is probably fair to say that the Cooper scale deliberately emphasizes three handling qualities categories. The category boundaries are between satisfactory for normal operation and acceptable for emergency operation (a numerical 3.5), and between the emergency operation category and unacceptable (a numerical 6.5). Cornell Aeronautical Laboratory (Harper, 2) has evolved a scale primarily for use with the many configurations possible with variable-stability aircraft. This scale is shown in Table II. Their scale is not intended to emphasize any particular levels. Others have used variants of these two scales, modified to emphasize particular types of flying operations such as tracking tasks.

The two scales of Cooper and CAL are not directly comparable point by point. However, the opinion has been ventured that they are probably

TABLE I

THE ORIGINAL COOPER SCALE (1)

COOPER					PR
DESCRIPTION	ADJECTIVE RATING	MISSION	PRIMARY MISSION ACCOMPLISHED?	CAN BE LANDED?	
Excellent, includes optimum			Yes	Yes	1
Good, pleasant to fly	Satisfactory	Normal operation	Yes	Yes	2
Satisfactory, but with some mildly unpleasant characteristics			Yes	Yes	3
Acceptable, but with unpleasant characteristics			Yes	Yes	4
Unacceptable for normal operation	Unsatisfactory	Emergency operation	Doubtful	Yes	5
Acceptable for emergency operation (stab. aug. failure) only			Doubtful	Yes	6
Unacceptable even for emergency condition (stab. aug. failure)			No	Doubtful	7
Unacceptable – dangerous	Unacceptable	No operation	No	No	8
Unacceptable – uncontrollable			No	No	9
$@≠*! Did not get back to report	Unprintable	What mission?			10

TABLE II

THE CORNELL AERONAUTICAL LABORATORY SCALE (HARPER, 2)

MISSION SUITABILITY (CAL'S "CATEGORY")		PILOT ATTENTION OR EFFORT REQUIRED	ADJECTIVE DESCRIPTION WITHIN CATEGORY	PR
FLYING QUALITIES	AIRCRAFT ACCEPTABILITY			
SATISFACTORY Criterion: Mission performance is not seriously affected by any flying quality deficiencies which may be present Definition: "Seriously affected" = pilot would ask that the deficient characteristics be improved	ACCEPTABLE		Excellent	1
			Good	2
			Fair	3
UNSATISFACTORY Criterion: Mission performance is sufficiently affected by flying quality deficiencies that pilot asks that characteristics be fixed	"RELUCTANTLY" ACCEPTABLE Criterion: Mission performance deficiencies cannot be improved without a serious compromise of the other factors which influence the mission capability of the airplane		Fair	4
			Poor	5
			Bad	6
	UNACCEPTABLE	Requires major portion of pilot's attention	Bad	7
		Controllable only with a minimum of cockpit duties	Very bad	8
		Aircraft just controllable with complete attention	Dangerous	9
	UNFLYABLE	Control will be lost sometime during mission	Unflyable	10

most similar at about the 3.5 level (see Section V) and obviously much parallelism exists. From a detailed examination and consideration of the scales, it is plain that difficulties, if not deficiencies, are inherent in both. Some of these are listed below:

General:

1. The scales are ordinal, and of such nature as to have practically no chance of having equal intervals on some hypothetical underlying interval scale.
2. The definitions of qualities, tasks, and rating descriptors are sometimes vague.
3. As performance measures, ratings are incomplete. They usually are not directly connected with specific measurable parameters, so comments and detailed analyses are usually needed to discover underlying reasons for a given rating.

Cornell:

1. Very poor category delineation (e.g., "Unsatisfactory" flying qualities seem to be properties of a "Reluctantly Acceptable" aircraft; there are apparently no flying quality characteristics below "Unsatisfactory," etc.).
2. Double-duty adjective descriptors (e.g., bad and fair).
3. Incompatible adjectives, i.e., degrees of "goodness" (excellent, good, fair, poor, bad, very bad) mixed with degrees of "safety" (dangerous) and degrees of "controllability" (unflyable).

Cooper:

1. Mixes tasks (normal and emergency conditions).
2. Mixes mission phases (whatever phases are involved in the tests and some hypothetical landing operation).
3. Confusing nomenclature (e.g., "Unsatisfactory" is satisfactory for emergency operation).
4. Incompatible adjective descriptors.

Recently, Cooper and Harper (3) took into account some of the deficiencies of existing scales and published a revised scale (see Table III). Some experience has been gained with the new scale, and it appears that some of the difficulties may have been resolved. For example, the revised Cooper-Harper scale has vastly improved the compatibility of adjective descriptors. However, the scale is still ordinal

TABLE III

THE REVISED COOPER-HARPER SCALE (3)

CONTROLLABLE CAPABLE OF BEING CONTROLLED OR MANAGED IN CONTEXT OF MISSION, WITH AVAILABLE PILOT ATTENTION	ACCEPTABLE MAY HAVE DEFICIENCIES WHICH WARRANT IMPROVEMENT, BUT ADEQUATE FOR MISSION. PILOT COMPENSATION, IF REQUIRED TO ACHIEVE ACCEPTABLE PERFORMANCE, IS FEASIBLE.	SATISFACTORY MEETS ALL REQUIREMENTS AND EXPECTATIONS, GOOD ENOUGH WITHOUT IMPROVEMENT CLEARLY ADEQUATE FOR MISSION.	EXCELLENT, HIGHLY DESIRABLE	A1
			GOOD, PLEASANT, WELL BEHAVED	A2
			FAIR. SOME MILDLY UNPLEASANT CHARACTERISTICS. GOOD ENOUGH FOR MISSION WITHOUT IMPROVEMENT.	A3
		UNSATISFACTORY RELUCTANTLY ACCEPTABLE. DEFICIENCIES WHICH WARRANT IMPROVEMENT. PERFORMANCE ADEQUATE FOR MISSION WITH FEASIBLE PILOT COMPENSATION.	SOME MINOR BUT ANNOYING DEFICIENCIES. IMPROVEMENT IS REQUESTED. EFFECT ON PERFORMANCE IS EASILY COMPENSATED FOR BY PILOT.	A4
			MODERATELY OBJECTIONABLE DEFICIENCIES. IMPROVEMENT IS NEEDED. REASONABLE PERFORMANCE REQUIRES CONSIDERABLE PILOT COMPENSATION.	A5
			VERY OBJECTIONABLE DEFICIENCIES. MAJOR IMPROVEMENTS ARE NEEDED. REQUIRES BEST AVAILABLE PILOT COMPENSATION TO ACHIEVE ACCEPTABLE PERFORMANCE.	A6
	UNACCEPTABLE DEFICIENCIES WHICH REQUIRE MANDATORY IMPROVEMENT. INADEQUATE PERFORMANCE FOR MISSION EVEN WITH MAXIMUM FEASIBLE PILOT COMPENSATION.		MAJOR DEFICIENCIES WHICH REQUIRE MANDATORY IMPROVEMENT FOR ACCEPTANCE. CONTROLLABLE. PERFORMANCE INADEQUATE FOR MISSION, OR PILOT COMPENSATION REQUIRED FOR MINIMUM ACCEPTABLE PERFORMANCE IN MISSION IS TOO HIGH.	U7
			CONTROLLABLE WITH DIFFICULTY. REQUIRES SUBSTANTIAL PILOT SKILL AND ATTENTION TO RETAIN CONTROL AND CONTINUE MISSION.	U8
			MARGINALLY CONTROLLABLE IN MISSION. REQUIRES MAXIMUM AVAILABLE PILOT SKILL AND ATTENTION TO RETAIN CONTROL.	U9
UNCONTROLLABLE CONTROL WILL BE LOST DURING SOME PORTION OF MISSION.			UNCONTROLLABLE IN MISSION.	10

and the question of its quantitative character is as unanswered as with the previous scales.

Difficulties experienced with the use of scales can be divided into four convenient categories which include the problems just mentioned:

1. **Extrapolation of the simulated task to the real flight situation.** The necessity of using simple simulations gives rise to the problem of extrapolating the simulation to the actual flight situation. Interpretation of the display, and agreement between the experimenter and pilot on the objectives of the evaluation, are the important factors here.

2. **The alternate mission character.** Some scales allow for a change in mission should the pilot be unable to carry out the primary mission (landing the aircraft in the event of stability augmenter failure is an example). This is perhaps a tenable concept for actual flight testing, but becomes increasingly difficult to structure as the simulation is simplified.

3. **Verbal descriptions and phrases.** Incomplete and ambiguous scale category descriptors result in an undesirable arbitrariness in the calibration between real and simulated flight, thereby causing evaluation to be nearly a "black art" and lacking in good repeatability and consistency across the subject population.

4. **A scale's qualitative character.** Data relating subjective measures with vehicle and operator parameters are far from complete. Additionally, past experience has shown that the pilot is occasionally unable to articulate the primary causes of his discontent. It is not surprising, then, that existing scales do not solicit opinion expressed in terms of the quantities to which the operator is sensitive.

The difficulties of items 1 and 2 can be at least partially alleviated by carefully defining the conditions under which ratings are taken. Discussions follow in Section II.B which delineate those areas requiring special attention from the experimenter. Various alternatives to the language problem noted in item 3 will be discussed in Section II.C. Section III will then explore the possibility of a quantitative scale underlying the contemporary scales. We will then be in a position to investigate the connections between pilot ratings and system measures, as noted in item 4. This will be carried out in Sections IV and V.

B. CLARIFICATION AND REFINEMENT OF TASK

An adequate delineation of the types of assessments (and therefore the corresponding tasks or subtasks) that an operator will be required to make is a necessary preliminary to any evaluation problem using pilot opinion as a tool. It is not surprising that some pilots have been unable to rate a simulated configuration simply because of the inadequacy of instructions and statement of purposes. If a scheme of evaluation is to be universally useful, we must improve our understanding of the task situation and our ability to define it.

1. Mission and Task Elements

A "mission" is the composite of pilot/vehicle functions that must be performed to fulfill operational requirements. The pilot/vehicle functions, or mission elements, are properly called "tasks," and are defined by specifying (a) the control activities required, (b) the environment affecting the control situation (e.g., random disturbance levels), and (c) the performance specifications for the pilot/vehicle system. (Note that by these definitions, the task is redefined when, for example, the disturbance level is changed; thus a mission could have several parallel task alternatives which are dependent on environmental conditions.) These "task elements" will be discussed briefly below.

a. Control. When an aircraft is flown manually the pilot is concerned chiefly either with maintaining the aircraft in a steady condition of flight or with changing the aircraft from one steady condition to another. Control is the means to accomplish these ends and is defined in the Handbook of Astronautical Engineering (58), Sect. 27.5, p. 35, as:

> "The development, and application to a vehicle, of appropriate forces which (1) establish some operating equilibrium state of vehicle motion (operating-point control), and (2) restore a disturbed vehicle to its equilibrium state and/or regulate, within desired limits, its departure from operating-point conditions (stabilization)."

Control implies the imposition of commands upon the system and the suppression of the effects of disturbances. Disturbance suppression is conventional closed-loop regulation when the pilot is active. Also, some disturbance suppression capacity is inherent in the craft even when it is operating unattended. Thus both closed- and open-loop pilot/vehicle systems are involved in suppressing the effect of disturbances on the aircraft. Pilot inputs to the craft may be pure commands, which are functions of time alone, or may depend on some vehicle deviation from a desired state of motion; so command operations are also both open- and closed-loop in nature. Therefore control activities in piloted flight have four aspects:

- Command maneuvers, open-loop
- Command maneuvers, closed-loop
- Regulation
- Unattended operation (open-loop regulation)

Closed-loop features are dominant in the first three aspects; explicitly for the middle two; and implicitly for open-loop command maneuvers because these end in closed-loop operations unless the maneuvers are flawlessly performed. Although the open-loop characteristics can have a large influence on pilot workload, ratings tend to depend on the closed-loop control characteristics because most deficiencies will appear only under the difficult and demanding higher gain conditions. Thus, handling qualities studies have historically concentrated on closed-loop tracking as the primary evaluation task and will probably continue to do so for some time to come.

b. **System input.** Environmental factors influencing the pilot/vehicle system characteristics and/or response must be specified to completely define the mission. These factors are most commonly of a system input nature (either disturbance or command) and since the mission is comprised of tasks, system inputs will be included in the task specification also. This breakdown is somewhat arbitrary, but useful. An input catalog can be constructed to show typical command and disturbance inputs [gust and terrain inputs, ILS spectra, precision approach radar noise, etc., e.g., Ref. 4] so that the task may be defined in terms of the input to a high degree of accuracy.

c. **Performance specifications.** To complete the task definition, performance specifications must be stated. It is here that mission effects become apparent. With the definition of "mission" as "required operations," and with a catalog of generic tasks, mission effects become a matter of degree rather than of kind, e.g., scaling of amplitudes, response times, regulation accuracy, time duration of task, etc. For example, Table IV summarizes many of the common flying tasks of a command maneuver nature and could be considered the beginning of a generic task catalog. The variables are shown in Fig. 1.

The inclusion of environmental and performance specifications will enable us to avoid the embarrassing conflict that apparently exists when two vehicles of similar kind are given widely different pilot ratings. Thus a stability-augmented hovering helicopter is often rated "poor" in gusty air, while a lunar landing vehicle, which has essentially the same dynamics, is rated "good." The difference is obviously the disturbance input. Although the state-of-the-art is not advanced enough at this time, sufficient data will no doubt exist sometime in the future to enable an analytical tie to be established between pilot ratings for two different tasks, where the vehicle dynamics are the same and the task differences are entirely due to input level and performance requirements. Our ability to find the tie, however, will depend heavily on record keeping and instructions to the rater. The rater must evaluate in the context of the mission.

With the mission phase or task completely specified, the designer is in a position to solicit an evaluation of a specific vehicle, or a comparison between vehicles. Note that without a complete definition, only a comparison can be made, and it will be based on some nonspecified performance characteristics which (1) may preclude close agreement between evaluators, (2) does not really help the designer in determining the suitability of the vehicle to perform its reason for being, and (3) makes it impossible to pass along any useful information to other experimenters.

2. Bases for Rating—Handling Qualities

With the approach to task definition established, the general factors influencing pilot opinion of a given task can be discussed. The purpose here is to indicate some classifications of these factors which will be helpful in establishing better communication between the engineer and test pilot.

TABLE IV

COMMAND MANEUVERS IN GENERAL FLYING TASKS

	MANEUVER	DIRECT CONTROL		COORDINATING CONTROL			EXTREME CONDITIONS
		Surface	Feedbacks	Surface	Feedbacks	Coordinating Control Function	
LONGITUDINAL CONTROL	Establish equilibrium flight, i.e., U_o, γ_o: Climb, dive, or horizontal flight	δ_e	α_c, u_c, θ	δ_T, δ_B δ_a, δ_r	$f(\gamma)$, M φ, a_y	Keep wings level, sideslip zero, during power or speed changes	Low q Large a_x Longitudinal asymmetries
	Pullup, pushover	δ_e	α_c, θ_c or γ_c, h_c or a_{z_c}	δ_a, δ_r	φ, a_y	Keep wings level (i.e., counter engine angular momentum)	
LATERAL CONTROL	Establish roll rate	δ_a	p_c	δ_r δ_e	δ_a, P φ	Maintain $a_y' = 0$ by offsetting N_{δ_a}', N_p' Provide desired lift increment	
	Establish bank	δ_a	φ_c	δ_r	a_y', r	Maintain $a_y' = 0$, offset N_r'	
	Banked turn heading change	δ_a	ψ_c		As above		
	Heading change	δ_r	ψ_c	δ_a	φ		Decrab
	Maintain ground track	δ_r	γ_e	δ_a	φ		Crosswind
	Maintain heading	δ_r	ψ_c				Lateral asymmetry, including engines out

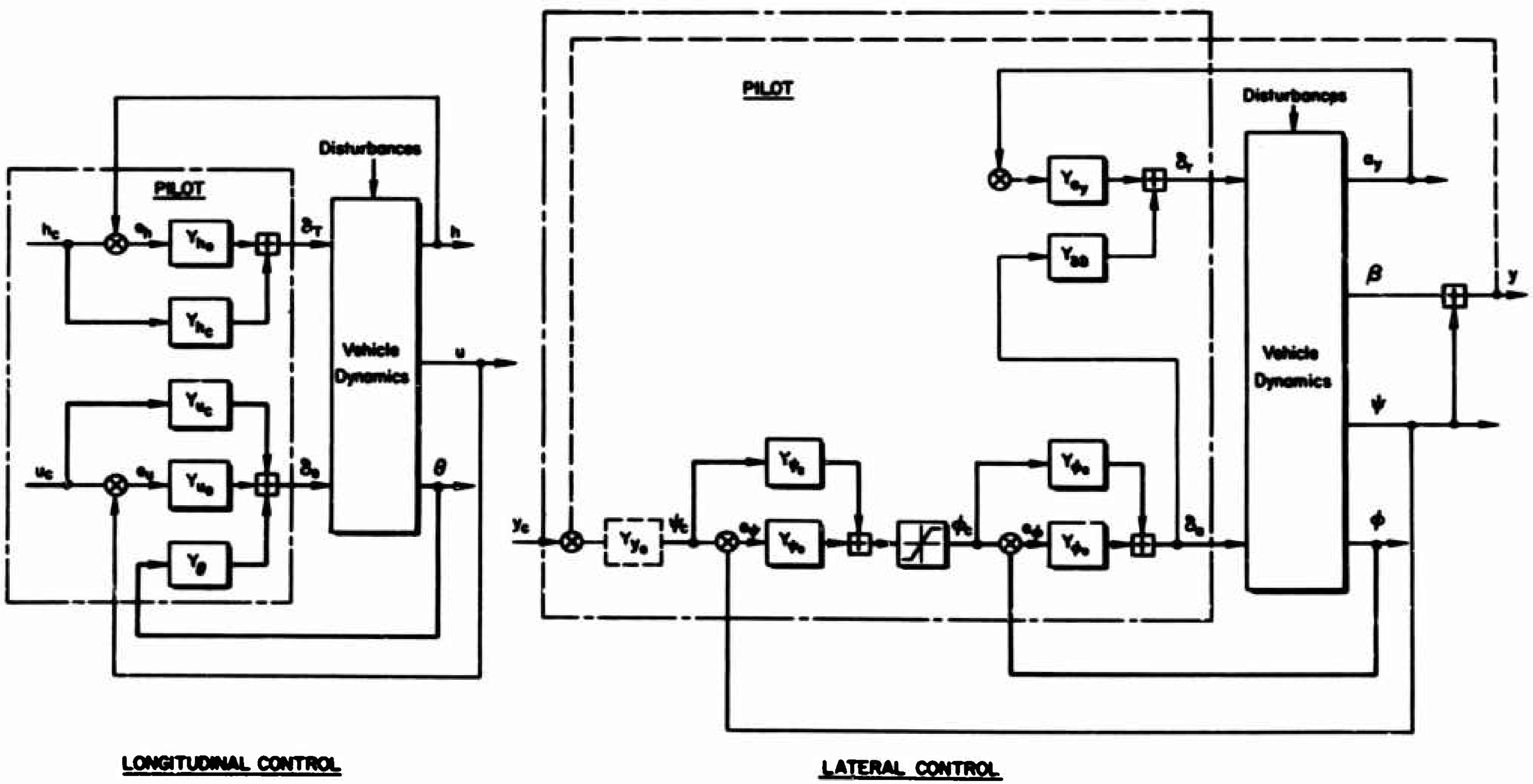


Figure 1. Block Diagrams Showing Loop Closures Used in Conventional Aircraft Control

Handling qualities may be defined as those characteristics which determine the control nature and behavior of pilot/vehicle systems. In this context a handling quality is, therefore, any property of the pilot/vehicle system which relates to open- or closed-loop command or regulation. Handling qualities thus include any properties or attributes of the vehicle and the pilot as they interact, either actively or passively, in the pilot/vehicle system. Vehicle characteristics associated with

- changing the equilibrium flight condition
- controlled maintenance of equilibrium
- unattended maintenance of equilibrium
- changes in behavior of the equilibrium

are clearly such properties. Pilot behavior characteristics necessary for control are also handling qualities. These include

- open-loop command insertion
- kinds of control loops closed (airframe motion quantities sensed by the pilot)
- the type of control effort required within each control loop (e.g., the necessary pilot equalization, as discussed in Section V.A) to achieve crossovers compatible with adequate pilot/vehicle system stability and response

Less direct pilot-connected handling qualities are the attention and skill (i.e., training and experience) levels needed to generate the pilot behavior qualities listed above.

Properties of the pilot/vehicle system as an entity are a third kind of handling quality factor. Examples would include closed-loop characteristics such as

- bandwidths (loop closures or crossover frequencies) of control loops closed
- average system performance, such as rms errors, in the presence of representative commands or disturbances

The total pilot/vehicle system characteristics, as a class, reflect only those pilot and vehicle dynamic interactions which cannot be expressed just as well by either pilot or vehicle characteristics. This category is especially sensitive to the external environment as the source of disturbances.

A comprehensive list of handling qualities could be developed by extending the above vehicle, pilot, and pilot/vehicle system characteristics. In such a list, however, the vehicle and pilot properties are not as obviously interconnected as, in fact, they are forced to be by pilot adaptability. An alternative scheme is to generalize on those attributes possessed by the vehicle for which corresponding, or associated, pilot capabilities exist. Such a classification is given in Table V, where pilot and vehicle properties are expressed in terms of the dynamic parameters of manual vehicle control (see Section V for a more thorough discussion of the parameters).

3. Abstraction of Tasks

With an understanding of what handling qualities are, the abstraction of real tasks to simplified simulations can be made using the criterion that qualities of Table V be observable in the simplified abstraction. Table VI shows some idealized vehicle configurations for the simple controlled elements used in the McRuer, et al, (5) investigation and for which significant data exist. Combinations of the simplified characteristics are appropriate for general flying tasks involving closed-loop and some open-loop control. In particular, they are reasonable idealized handling qualities of the "maneuverability" and "command-ability" nature. The longitudinal cases involving short period dynamics may require the addition of a stiffening term to approach an idealized situation for "trimmability."

The idealized configurations of Table VI are quite compatible with simplified displays. In fact, when a rating scale is accompanied by a task definition, the key factor regarding the display is that the experimenter and subject agree on interpretation, since the display abstraction becomes essentially a "mission effect." The experimenter must be explicit about his objectives. He may very well wish to evaluate multiple handling qualities, e.g., "controllability" (a function of the controlled element), "trackability" (a function primarily of the system input), and open-loop characteristics. Making such an evaluation could very well require two or three ratings from a subject, and a suitable scale would of course need to exhibit a flexibility capable of handling such requirements.

TABLE V

SOME HANDLING QUALITIES

QUALITY	VEHICLE-CENTERED CORRELATES		PILOT-CENTERED CORRELATES
	Vehicle Properties	Involved in Unattended Operation	
Trimability Resetability	Low-frequency* vehicle characteristics which are connected with the establishment and maintenance of equilibrium.	Yes	Low-frequency equalization, i.e., lag or reset requirement; Bode gain.
Maneuverability	Mid-frequency vehicle characteristics** associated with the establishment of a new equilibrium.	Yes	Mid-frequency (crossover) characteristics. Anticipation requirements.
Open-loop commandability	General form of low and mid-frequency vehicle characteristics associated with changes in equilibrium.		Form required for Y_p; i.e., simplicity of pilot response pattern.

*"Low-frequency characteristics" are "low" relative to crossover frequencies of the pilot/vehicle system.

**"Mid-frequency characteristics" are those in the general region of pilot/vehicle system crossover.

TABLE VI

IDEALIZED VEHICLE CONFIGURATIONS

Y_c	CORRESPONDING IDEALIZED VEHICLES			
	Longitudinal		Lateral	
K_c	$a_{z_c} \rightarrow \delta_e$ $\alpha \rightarrow \delta_e$	for very large maneuver margin	$\beta_c \rightarrow \delta_r$ $\varphi_c \rightarrow \delta_a$	for very large directional stability in steers-like-a-car control mode
K_c/s	$\theta_c \rightarrow \delta_e$	approximation for ideal ω_{sp}, ζ_{sp}, $1/T_{\theta_2}$	$\varphi_c \rightarrow \delta_a$	approximation for ideal $1/T_R$
K_c/s^2	$h_c \rightarrow \delta_e$, $\theta_c \rightarrow \delta_e$,	formation flight for very small maneuver margin and large $1/T_{\theta_2}$	$\varphi_c \rightarrow \delta_a$	for small $1/T_R$

C. SCALE LANGUAGE ALTERNATIVES

The discussion of the previous section (II-B) essentially defined the problems related to task, mission, and simulation, and evolved some alternate ways to consider handling qualities. The language used to solicit responses from subjects will be discussed here.

As noted by Cooper and Harper (3), the pilot evaluation is intended to meet two objectives: (1) to provide an overal assessment of the suitability of the vehicle in its intended use (called a "global" rating by some) and (2) to provide information pertaining to the specific deficiencies which interfere with the intended use.

The first objective requires that the rater be able to express his subjective impression of the handling qualities of the vehicle in performing the required maneuvers. This "impression" is the sum total of all of the sundry physical factors which contribute to the handling qualities of the vehicle. Since there is no common physical measure which integrates all of the factors, a scale must be, in part at least, in subjective terms.

The second objective requires that the rater be able to provide information on specific problem areas to aid the experimenter or designer in solving the problems. Thus, a language is required which is valid and unambiguous to as large a population as possible to minimize training requirements and to maximize repeatability.

What handling qualities are was discussed briefly in the previous section — let us use that information to present some alternative language possibilities for a scale. Table VII shows various handling quality related measures and parameters grouped arbitrarily by what might be called "disciplines." Thus, reponses could be solicited from pilots in each of these groups. For example, raters could be trained in the engineering language of pilot parameters (column 1). The rater would

TABLE VII

HANDLING QUALITY RELATED MEASURES OR PARAMETERS IN TERMS OF:*

PILOT (1)	VEHICLE (2)	SYSTEM (3)	FREQUENCY DOMAIN (See Table V for definitions) (4)	PERFORMANCE (5)	SUBJECTIVE (6)
K_p	K_c	ω_c	Trimability	$\overline{e^2}/\overline{i^2}$	Sensitivity
T_I	ζ_i , ω_i	ω_b	Maneuverability	$\overline{c^2}$	Controllability
T_L	T_i	φ_m	Open-loop Commandability	$\overline{F^2}$	Precision
τ_e		K_m		Accident Rates	Effort
α		t_r		Gunnery Scores	Range of Tasks
		over-shoot		Probable Error	(Attention)
		σ_i			Safety
		ω_i			Comfort
					Trustworthiness
					Regulation

*See List of Symbols for definitions.

then be telling the experimenter what it was about his own responses that he disliked. The disadvantage of an engineering, or scientific, language is the high level of training required for selective and repeatable ratings. Of those questioned, unanimous agreement among pilots and almost unanimous agreement among engineers was obtained on this point.* Similar results were obtained on the definitions of handling qualities based on frequency domain (column 4) characteristics. It was concluded that a pilot would indeed have a difficult time remembering and interpreting the distinction between the frequency domain terms. The ability to assess vehicle and system parameters (columns 2 and 3) depends heavily on training and also requires a variety of maneuvers to be performed. Even then it is doubtful that a pilot could consistently determine the state of sundry frequency and time response parameters. Past work has shown that performance is very often not correlated with the pilot's opinion (Refs. 21, 26, 52), so the performance measures of column 5 are unlikely to yield useful results, even if the pilot could estimate them.

Column 6 represents an attempt to define subjective piloting problems or problem areas. The list could be extended indefinitely, but in Table VII they have been arranged in what is felt to be an order of decreasing validity. The table does not imply that safety, for example, is unimportant, only that a rater would have difficulty comparing vehicles based on the ambiguous quality "safety."

The criteria used in selecting possibilities for a scale were that:

(1) The language be as natural and unambiguous to the rater as possible so that little analysis by the pilot is required during the rating situation.

(2) The language be as descriptive of piloting problems or problem areas as possible.

*Informal discussion on scale language possibilities were held with several persons, including six STI handling qualities engineers, one Air Force test pilot, and three NASA test pilots.

From the discussion above, it is apparent that the subjective words of column 6, Table VII, are most likely to be suitable. By "suitable," it is meant that the descriptors should be unambiguous semantically, and universally valid in the rating situation. The semantic problem can be tested by a simple survey and is discussed in Section III, while the validity question can then be considered through actual rating experiments, which are described in Section IV.

D. SUMMARY

The conclusions to be drawn from the discussion to this point are that:

- The experimenter/designer should draw from his catalog of common maneuvers to construct a series of tasks representative of the mission. Similar tasks can be grouped so that the differences between them become scaling problems.
- The tasks can then be abstracted, if desired, to simplified control situations capable of being easily simulated.
- The pertinent variables of the evaluation should be set down in writing. These will include the task definition, performance requirements, time duration of task, interpretation of display, disturbance, and any other information necessary to establish agreement between the experimenter and the pilot on the purposes and objectives of the evaluation.
- A scale (or scales) is most likely to be universally applicable and valid if constructed from subjective descriptions of handling qualities.

The problem of quantizing scale descriptor candidates is quite complex; consequently, the entire following section will be devoted to an application of psychophysical measurement techniques to rating scales.

SECTION III

DETERMINATION OF THE QUANTITATIVE NATURE OF RATING SCALES

A. INTRODUCTION

As discussed earlier in this report, a major objective of this study is to evolve a rating scale which has some underlying functional structure so that certain mathematical operations may be performed with pilot rating data. Our approach to the problem will draw heavily on the methods of psychometrics. Briefly, we will select a group of phrases which are possible candidates for a rating scale language. We will then construct an experiment (in the form of a survey) to gather data on the proposed phrases. The data will then be reduced using notions and techniques evolved from the theory and methods of psychometrics. Some of the concepts are quite involved; hence this entire section will be devoted to the scaling problem. Since most handling qualities engineers are not familiar with the field of psychometrics, let us review the fundamentals of the techniques we will be using before we construct the experiment.

B. REVIEW OF MEASUREMENT CONCEPTS

1. Types of Scales

If a measurement is made on a physical object with an instrument (nonhuman) of some sort, the measure is an objective one and the resulting data lie along a physical continuum. When an observer estimates a measure, it is a subjective judgment and the estimates lie along a psychological continuum. The relationship between the objective and subjective scales have been studied for many years for certain stimuli (such as estimation of weight, loudness, pitch, etc.) and is an area of endeavor called psychophysics.

There are several degrees of sophistication of psychophysical scales. Table VIII repeats the measurement scale classification as found in Rosenblith, et al (6). As will be noted in the table, in order for means to be legitimately taken, the rating scale must be an interval scale as a minimum. But the examples of scales in Table VIII are all

TABLE VIII

CLASSIFICATION OF MEASUREMENT SCALES
(From Rosenblith, Ref. 6. Reproduced by permission of John Wiley & Sons.)

Scale	Basic empirical operations	Mathematical group-structure	Permissible statistics	Typical examples
Nominal	Determination of equality	Permutation group $x' = f(x)$ where $f(x)$ means any one-to-one substitution	Number of cases Mode "Information" measures Contingency correlation	"Numbering" of football players Assignment of type or model numbers to classes
Ordinal	Determination of greater or less	Isotonic group $x' = f(x)$ where $f(x)$ means any increasing monotonic function	Median Percentiles	Hardness of minerals Grades of leather, lumber, wool, and so forth
Interval	Determination of the equality of intervals or of differences	Linear group $x' = ax + b$ $a > 0$	Mean Standard deviation	Temperature (Fahrenheit and Celsius) Position on a line Calendar time Potential energy
Ratio	Determination of the equality of ratios	Similarity group $x' = cx$ $c > 0$	Geometric mean Harmonic mean Per cent variation	Length, density, numerosity, time intervals, work, and so forth Temperature (Kelvin)

of physically measurable quantities. What about psychological quantities such as vehicle handling qualities where no physical parallel exists? Can an "interval scale" of a purely subjective quantity be constructed? The work of psychologists in the field of psychometrics indicates that it is indeed possible. [The excellent works of Guilford (7) and Torgenson (8) would provide the reader with a thorough background in the field should he desire to delve further into the details of the subject.] Applications to problems somewhat akin to the problem being considered here have been made by, for example, Uhrbrock (9), where scale values were determined for a large number of rating scale statements regarding an employee's suitability to be employed as a foreman. Other examples are readily found in the literature [see, for example, Ferguson (10), Thurstone (11), or Uhrbrock (12)].

The techniques often used in scaling problems of the type we have here have been derived from notions about the distributions of estimations, particularly the concepts associated with discrimination thresholds. Thus, we will have to review some additional measurement concepts. A class of methods introduced by Fechner [see, for example, Guilford (7)] measures just noticeable differences (jnd) along the physical continuum and uses these measures of resolving power as equal units on a scale of sensation. By assuming (1) that the jnd is proportional to the stimulus magnitude (Weber's law) and (2) that each jnd represents a constant increment in sensation, Fechner derived his logarithmic law.

Thus, if s is the stimulus and R is the response, or sensation, then the difference in stimulus magnitude corresponding to a jnd is

$$\Delta s = s_2 - s_1 = k_1 s \quad \text{(Weber)} \tag{1}$$

Also,

$$\Delta R = R_2 - R_1 = k_2 \quad \text{(Fechner)} \tag{2}$$

Combining the two expressions yields Fechner's logarithmic law:

$$\Delta R = k \frac{\Delta s}{s} \tag{3}$$

or

$$R = k \log s \tag{4}$$

The accuracy of these assumptions has been given considerable attention subsequently by those interested in measurement, and they can be shown to be not quite true for some stimuli. A pertinent distinction has to be made between types of stimuli. If a sensation is produced by adding to a stimulus, i.e., by increasing its magnitude, such as would be the case in weight, brightness, or loudness estimation, the nature of the continuum is quantitative and is called "prothetic." The class of continua including qualitative and positional aspects of things, such as pitch and length, are called "metathetic." For this class, a change in sensation seems to be a result of substituting stimuli rather than adding them. The main point of this distinction is that in the metathetic domain, the jnd are subjectively equal over the continuum; whereas in the prothetic domain, the jnd grow rapidly in subjective size as we go up the scale of the continuum.

The significance of making the metathetic/prothetic distinction is the following: Using Fechner's assumption (Eq. 2) we would expect that summing up a number of jnd's above the absolute threshold (say 50) would yield twice the response as summing up half that number (25), since each jnd is supposed to yield equal sensation increments. But this appears to be only true for metathetic stimuli (pitch, color, position, etc.) and not for prothetic stimuli (weight, loudness, etc.). Thus, if three different scaling methods are used to estimate the psychophysical scale of a prothetic stimulus, "apparent duration" [Churchman (13)], each procedure yields a different scale. Stevens (14) is thus forced to conclude that scaling methods employing the assumption of subjectively equal jnd's or discriminal dispersions, or equally often noticed differences, probably do not result in interval scales for prothetic stimuli.

Since the "handling qualities" of a vehicle are obviously qualitative characteristics, we would expect the continuum to be metathetic. We could quite reasonably make the assumption that it is, which in effect would be <u>defining</u> the desired psychological continuum as being one on which a subject has a constant sensitivity, or discriminability, across the entire scale. Rather than make the assumption, however, we shall use a scaling method which yields the subjective size of the sensitivity, so that the question of metathetic or prothetic is empirically determined. Let us say only that evidence indicating a metathetic continuum would be most welcome, since the resultant scales produced by different scaling methods tend to be more consistent with one another than is the case for prothetic continua. The notions which lead to the scaling method to be used in this study will be discussed next.

2. An Intuitive Example of the Scaling Method

Before writing down the formal equations for the method to be used, called the "Method of Successive Categories," it would be instructive to consider an example from a field more closely associated with psychometric methods — the evaluation of people.

Let us suppose that we have a collection of descriptions of various traits of people. Our problem is to discover how suitable a person would be for a foreman's job by soliciting the appropriate description of him from persons who know him. If the descriptions have somehow been previously scaled, a direct numerical indication of foreman suitability will be available. Uhrbrock (9) solved the scaling problem by applying the "Method of Successive Categories" (also called the "Method of Successive Intervals") as discussed briefly below.

Several descriptive phrases (called "items") of foremen were collected. The descriptions covered the entire spectrum of foreman suitability, from the best to the worst. Each item was then typed on a small card, and the resulting stack of cards was given to each participant (called a "rater"). The rater was placed before a row of boxes (say 11) and asked to sort the cards into the appropriate boxes using the following rules: The box at one end was considered to represent an "extremely poor foreman," while the box at the other end represented an "extremely good foreman." The boxes between the two end boxes represented foremen between the two extremes. The rater could recheck his card placement as often as necessary to satisfy himself that he had ordered the cards correctly. After many raters had sorted the cards, a histogram could be drawn for each item, showing its frequency of placement in each box. Although Uhrbrock did not publish his raw data, let us hypothesize that four of the items had distributions as shown in Fig. 2a, where the histograms have been approximated by continuous curves.

It can be seen in the figure that, for example, most of the raters put phrase A in box 2, while phrase D was distributed between boxes 8, 9, 10, and 11. After noticing the locations of the means of the phrases, one might be tempted to say that the amount that A was better than B was the same as C was better than D, or that A-B = C-D. That is clearly not the case, because for A and B there was very little confusion about which was the better phrase, while considerable confusion existed when C and D were evaluated, as exhibited by the overlap in the distributions.

The effect of applying the Method of Successive Intervals to the data is shown sketched in Fig. 2b. In this example, the method (which will be shown in more detail in the following section) "stretches out" the scale where the dispersions are small and "squeezes up" the portion of the scale where dispersions are large until all the dispersions are approximately equal.

The effect of the manipulations on the scale values of the items is obvious. On the psychological continuum, labeled ψ, the means reflect our earlier feelings that there was indeed a larger separation between A and B than between C and D. It is the application of this method to handling qualities descriptors that we shall work toward in the subsequent evolution of a rating scale.

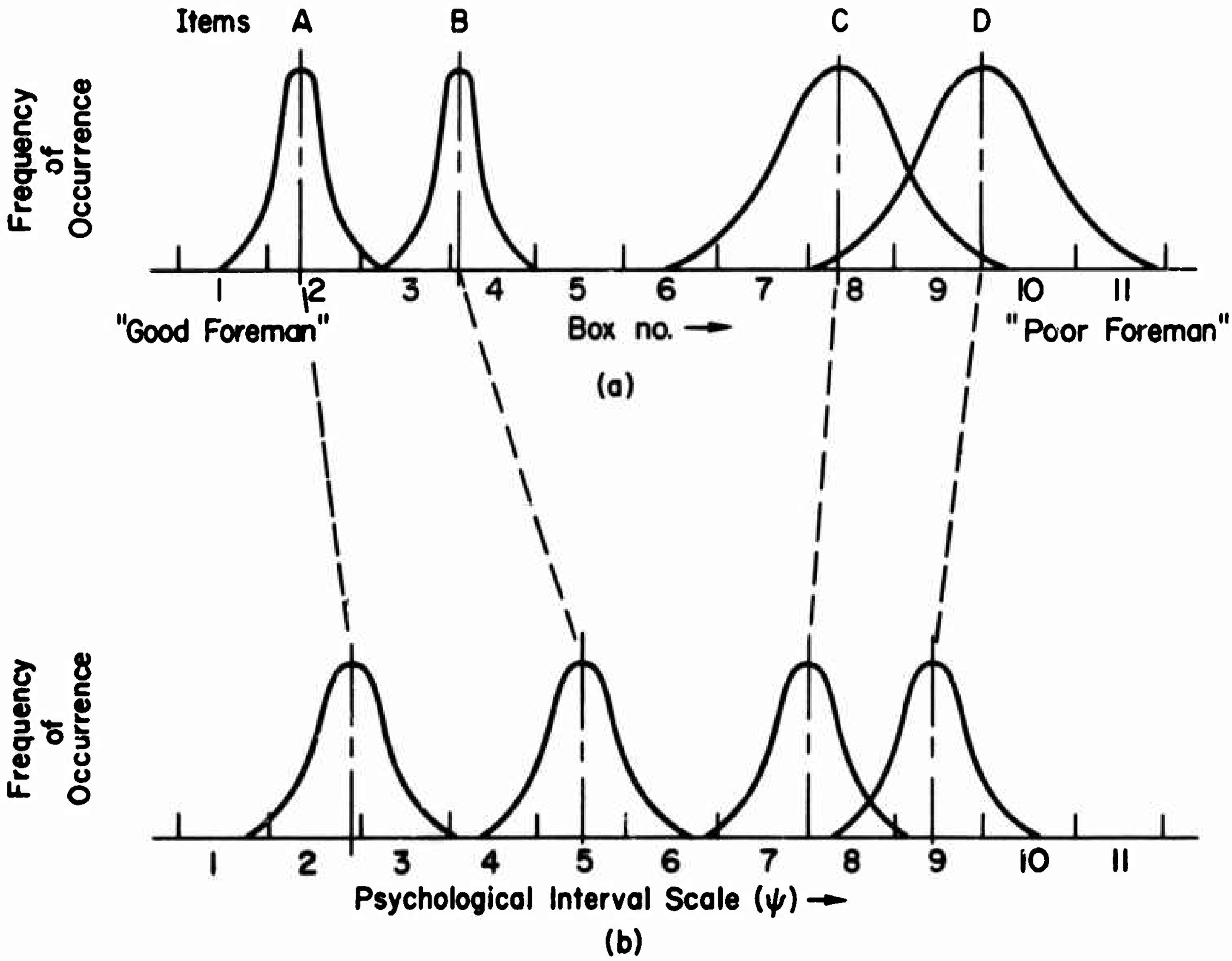


Figure 2. Hypothetical Results of Uhrbrock's (9) Rating Scale Results

C. SCALE VALUES AS DETERMINED BY THE METHOD OF SUCCESSIVE INTERVALS

1. Selection of Items to be Scaled

Regardless of the scale form finally selected, it will doubtless contain descriptions of one or more traits, each scaled in several "degrees of goodness." The fact that there are not many distinct "degrees" which are couched in simple terms requires that a careful selection of the candidates be made. So, for example, what are ten (or so) degrees of handling qualities? "Excellent" would probably be fairly specific to most, but what are some others?

To get at this problem a series of phrases were assembled from various sources (including rating scales currently in use) which expressed subjective traits in which a rater might wish to reply in a rating situation. Degrees of the first five traits of column 6, Table VII, were included and were considered to include the majority of problem areas to which a rater would respond. An attempt was made to include a fairly even distribution across the continuum from "best" to "worst." Table IX shows the distributions for the traits considered. The traits are shown vertically, while degrees of goodness of the traits are shown horizontally. The columns do not imply that all traits in a specific column have the same psychological weight or value. The procedure to be followed should show, however, that the degrees are in the correct order.

A form of a graphic rating scale [Guilford (7)] was used to gather the necessary data for the Successive Interval Method. The graphic scale, which serves the same purpose as the "boxes" of the foreman rating experiment of Section III-B-2, is similar to a technique used by Lefritz (15) to scale 200 adverb-adjective combinations. Unfortunately, Lefritz's items were not directly suitable for a rating scale.

Briefly, in our survey the rater was instructed to read over a list of phrases arranged in random order. Then each phrase was presented one at a time. The rater was to imagine he were reading a handling qualities report where the test pilot has used the presented phrase in describing a vehicle. The rater was then instructed to indicate his impression of

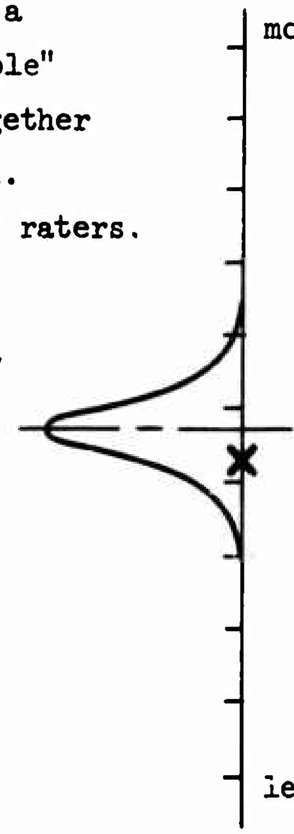
TABLE IX

DEGREES OF GOODNESS OF VARIOUS HANDLING QUALITIES DESCRIPTORS FROM TABLE VII

TRAITS*	← "BEST"				DEGREES OF GOODNESS OF THE TRAIT				"WORST" →	
Handling Qualities	Excellent	Highly desirable	Good	Pleasant	Fair	Poor	Very poor	Bad	Nearly uncontrollable	
Control	Extremely easy	Very easy	Easy	Controllable	Controllable	Controllable	Controllable	Barely Controllable	Nearly uncontrollable	Uncontrollable
Precision	Excellent	Good	Good	Fair, somewhat inadequate	Definitely inadequate	Poor	Very imprecise			
Range of Tasks and Maneuvers	All	All	All	Somewhat restricted	Restricted	Very restricted	Very restricted	Very restricted	Extremely restricted	
Response	Excellent, pure	Good, relatively pure	Fair, somewhat impure							
Sensitive Sluggish Uncomfortable			Mildly	Marginally, but disturbingly	Definitely	Quite	Very	Extremely	Much too	
Demands on Pilot	Completely Undemanding	Largely Undemanding	Mildly demanding	Somewhat undesirably demanding	Definitely demanding	Demanding	Very demanding	Extremely demanding	Completely demanding	

*From Column 6, Table VII; p. 18

the vehicle, as gained from the phrase, on a graphic scale with end points "most favorable" and "least favorable." The survey form together with the raw data is included in Appendix A. Tables VII and IX were not available to the raters.



For example, the phrase "controllable with definitely inadequate precision" might be responded to by a rater as shown by the x in the sketch. The distribution of all of the raters surveyed might appear as shown by the bell-shaped curve in the sketch. A total of 63 persons contributed their time in scoring 64 phrases, thus providing adequate data for the subsequent processing.

2. The Method of Successive Intervals

The particular method we shall use to reduce the survey data is called the Method of Successive Intervals. This particular method is based upon the Law of Categorical Judgment, which in turn is derived from Thurstone's general judgment model [see, for example, Guilford (7), p. 35, and chap. 10].

Consider an observer comparing two stimuli and evaluating their relative values with respect to some attribute. Thurstone's model for such a process is given by

$$m_g - m_i = z_{ig}(\sigma_i^2 + \sigma_g^2 - 2r_{ig}\sigma_i\sigma_g)^{1/2} \qquad (5)$$

where m_g, m_i are the scale values of the i and g stimuli along the psychological (ψ) continuum

z_{ig} is the normal deviate, or the proportion of times that g was judged greater than i

σ_i, σ_g are the discriminal dispersions of i and g, i.e., the standard deviations of the distribution of responses to i and g on the ψ continuum

r_{ig} is the correlation between i and g

The assumptions made in constructing the judgment model are:

a. Each stimulus gives rise to a "discriminal process" which has some value on the ψ continuum.

b. When presented with the stimulus a large number of times, the observer, or rater, responds with a distribution of processes because of fluctuations within the observer.

c. The resulting distribution on the psychological continuum is normal, with a mean called the scale value and a standard deviation called the discriminal dispersion.

In the derivation of the successive interval notions, some additional assumptions are made:

d. The psychological continuum can be divided into categories, and the category boundaries exhibit a fluctuating value along the continuum similar to stimuli. The category boundaries can then be treated as a stimulus.

e. The dispersions associated with the boundaries are assumed to be constant across the continuum.

f. The correlation between momentary positions of two stimuli is zero ($r_{ig} = 0$).

These assumptions reduce Eq. 5 to

$$t_g = m_i + s_i z_{ig} \tag{6}$$

where now a boundary scale value, t_g, has been substituted for the stimulus scale value, m_g. We now have in Eq. 6

t_g = upper boundary of the gth category

m_i = the scale value of item i

s_i = the discriminal dispersion for item i

z_{ig} = the normal deviate corresponding to the cumulative proportion of the gth category for item i

Thus the Law of Categorical Judgment is reduced to the notion that the differences in scale values between a stimulus (phrase, in our case) and the category boundary is equal to the proportion of times that the boundary is judged greater than the phrase (z), times the measure of central tendency (s) of the phrase.

A particular application of these notions was made by Diederich (16), where a procedure was derived to minimize the mean-square error between the model (Eq. 6) and the actual data. Cumrey (17) computerized the procedure so that a routine is available to minimize the error expression,

$$E = \sum_{i=1}^{n} \sum_{g=1}^{k} (m_i + s_i z_{ig} - t_g)^2 \tag{7}$$

where n = the number of items or phrases
k = the number of categories

The routine then uses the normal deviates (z_{ig}) obtained from the survey as the actual data, and through an iterative procedure determines the values of m_i, s_i, and t_g which minimize the E of Eq. 7.

Certain additional restrictions and conditions are made in the procedure, which can be found in the paper (17). This procedure differs slightly from the example cited earlier (the suitability for foreman problem) in that here the dispersions are not assumed constant but are subject to empirical test. The determination of the "best fit" dispersions, then, will provide a check on our earlier feelings that opinions of handling qualities are metathetic in nature.

3. Results of the Experiment

In addition to subjecting the survey data to the successive interval program, some rather simple manipulations were also made to yield the means and variances of the raw scores, as well as the means and variances of "transformed" scores, where the end points of the scale were fixed for all raters by making a simple linear transformation to the raw data. Since the raw and transformed scores lend considerable insight into the nature of the rating scale problem, those results will be discussed first.

a. Means and Variances of the Raw Scores. To determine the nature of the responses from the 63 raters, a computer routine was used to compute the mean and variance of each item, then print out the items rank ordered according to their means. The results are shown in Appendix B, Table B-I, columns 1 and 2. The "most favorable," or top of the axis in the survey, was arbitrarily labeled zero, while the bottom was labeled ten. An indication of the semantic ambiguity of the ratings is obtained by plotting the variance of the item against the item position along the scale, and is shown in Fig. 3a. As can be seen, the items become increasingly ambiguous in the middle part of the scale, where standard deviations as high as 1.5 occur. The curve also shows a definite skew toward the bad end of the scale. The relative ambiguity of descriptors can be assessed by carefully studying columns 1 and 2 of Table B-I. Notice, for example, that "very poor handling qualities" and "bad handling qualities" (items 48 and 57, p. B-4) convey the same meanings to raters based on their means. An attempt was made to reduce the dispersions of Fig. 3a by constraining all raters to abide by the same rules. To do this, a simple transformation routine was developed.

b. Means and Variances of the Transformed Scores. Let us assume that a rating scale in its final form will have numerals associated with it, and, further, that there are two points along the scale to which we could insist that everyone rate in common. Ideally, the two points would demonstrate a low variability semantically. Two such points are available, one at each end of the scale. At the good end is "excellent handling qualities," while "uncontrollable" is universally agreed upon to fall at the bad end. Both of these phrases have very low variances associated with them. By insisting that all raters should have placed these two phrases at the same two spots along the scale, we can make a linear transformation of all of the scores. Thus, if a rater had a tendency to bunch all of his ratings in the middle of the scale, the transformation would stretch them out. If a rater had a bias toward one end of the scale, the transformation would remove it. The justification for applying such a routine is that in the final scale, the words

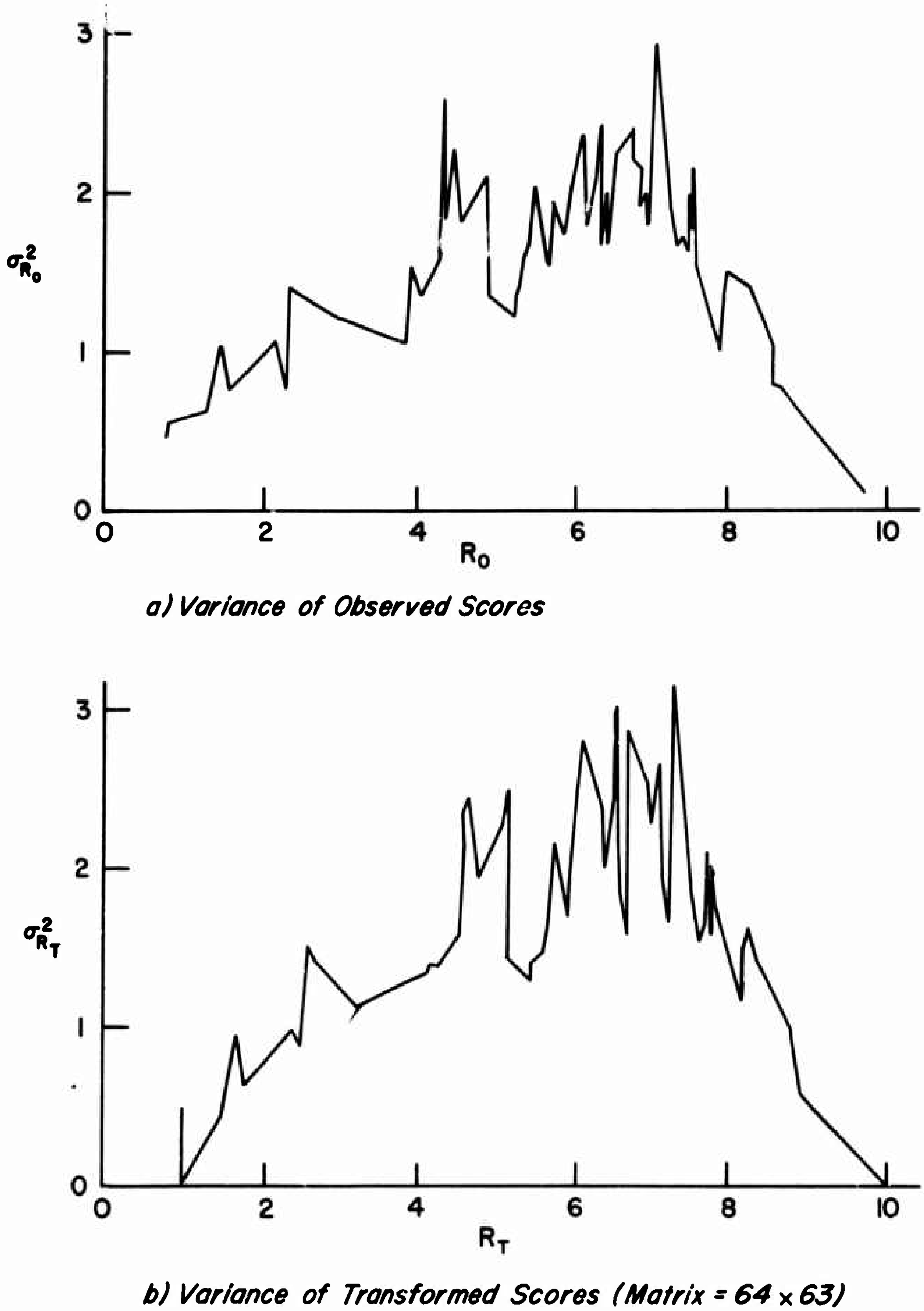


Figure 3. Variances of Semantic Judgments of Handling Qualities Phrases

along the scale will be fixed, i.e., all raters will have common end points. Typical raw scores and transformed scores might appear as in Fig. 4.

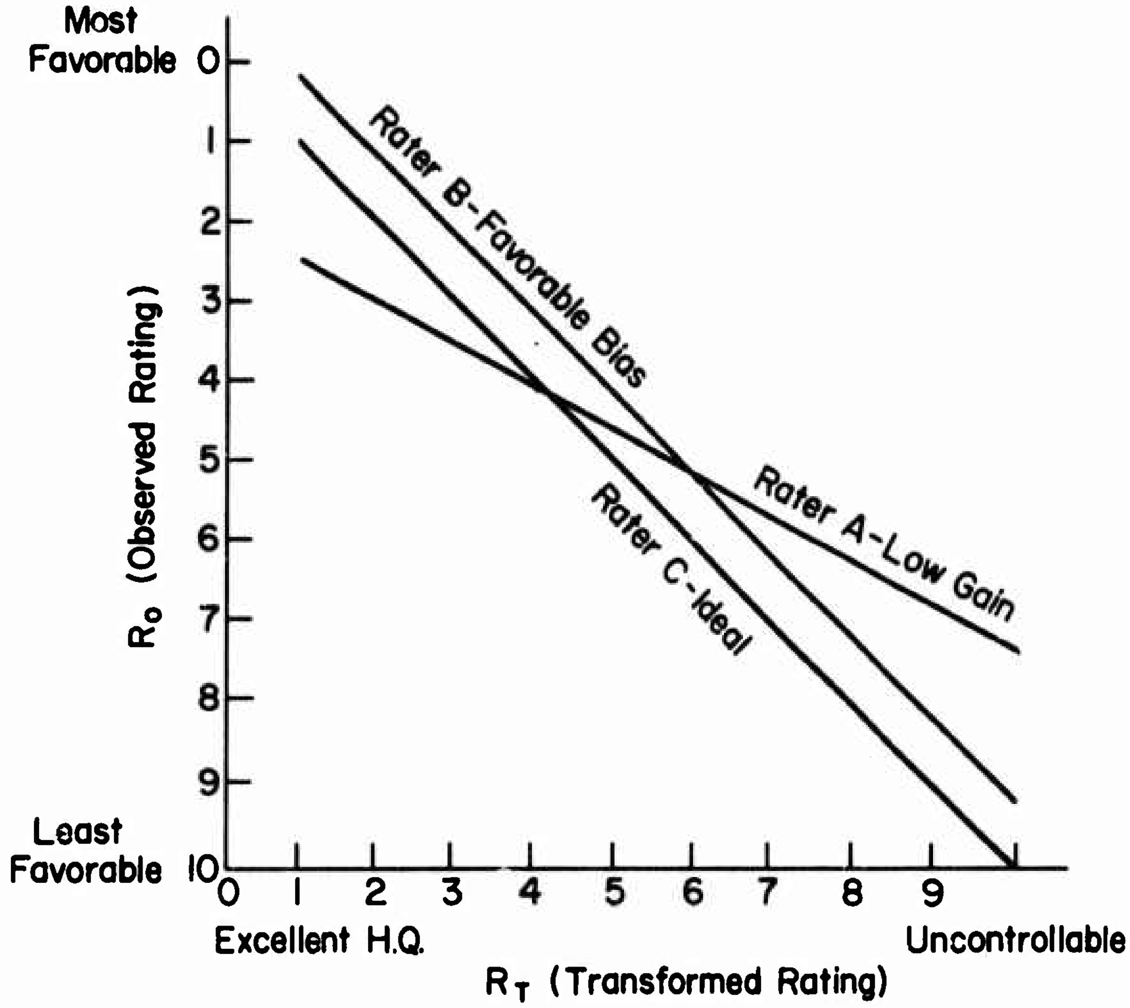


Figure 4. Possible Effects of a Linear Transformation on the Observed Scores

The results of obtaining the transformed ratings are shown in Fig. 3b, where the variances of the transformed scores are shown plotted against the scores themselves. A comparison of the two sets of variances, those from the raw scores and transformed scores, shows that the agreement between raters is made worse by the transformation, if anything. The conclusion is, then, that rater "bias" and rater "gain" are not significant factors causing the noted dispersions, and that no advantage will be gained in further manipulation of transformed scores.

c. **Successive Interval Results.** The questionnaire data was put through the successive interval routine of Section III-C-2 twice. The first time all phrases* were scaled. The phrases were then culled on the basis of the variance of the raw scores, and the high variability items (semantically ambiguous) were removed from the list. The remaining phrases, shown in Table B-II, Appendix B, were then put through the scaling routine again. This second set of values is a good approximation to those which would have been obtained if only the unambiguous phrases had been included in the questionnaire initially. The results of both runs are tabulated in Appendix B, Table B-I and will be discussed in the following subsection. The scale values have been arbitrarily adjusted to a nine point scale, with "excellent handling qualities" defined as 1.0, and "nearly uncontrollable" defined as 9.0. In a final scale form, 10.0 could be reserved for "uncontrollable," although it would then be inappropriate to include the 10.0 in any data processing.

The scale values obtained through the Successive Interval Method are by far the most interesting and important results of the survey. Before discussing their significance, however, it would be appropriate to point out that the dispersions of all the items are approximately equal on the ψ continuum (see column 6, Table B-I, Appendix B). Since it was not necessary to assume equal dispersions with the particular mean-square routine used, we have empirically shown that we are dealing with a metathetic continuum (see Section III-B-1), and we would expect the results obtained here to be entirely consistent with any obtained through other approaches.

D. DISCUSSION OF RESULTS

Armed now with legitimate numerical scale values for myriad descriptive handling quality phrases, we can now assess the numerical character of contemporary scales. First, to get an idea of what the

*With the exception of no. 28, uncontrollable. When all responses are in the first or last category, the routine will not converge, which reflects that no. 28 is an absolute end point and does not properly deserve a scale value in an interval scale.

scale values mean, let us look at the distributions of ψ values for the degrees of goodness of handling qualities. Then we can consider the connections between the ψ values and the Cooper ratings. Finally, we can estimate the errors of past analyses which were introduced through unjustified processing of data.

1. Scale Values for Degrees of Goodness of Handling Qualities

The adjectives modifying handling qualities are repeated in Table X from Appendix B, Table B-I, column 5, and are plotted in Fig. 5.

Several characteristics of rating scales can be inferred from the figure. Notice that the "neutral" area (that point on the scale which is neither "more favorable" or "less favorable") is in the vicinity of "fair handling qualities." When the questionnaire was developed, normal practice dictated that the midpoint be labeled "neutral," but since a neutral vehicle was difficult to envision, only two tie points (the end points) were labeled. We now know what "neutral handling qualities" are — they are "fair."

TABLE X

DEGREES OF GOODNESS OF HANDLING QUALITIES

Adjective	Scale Values, ψ
Excellent	1.00
Highly desirable	2.25
Good	3.70
Pleasant	3.71
Fair	5.34
Poor	7.39
Very poor	7.87
Bad	7.97
Very bad	8.33
Nearly uncontrollable (for ref.)	9.00

A considerable amount of interest in the Cooper "boundaries" is exhibited by most experimenters. A careful comparison of the words shown in Fig. 5 with the Cooper and Cooper-Harper scales of Tables I and III, pp. 4 and 7, establishes the probable intersection of these boundaries with the ψ scale. These "probable areas" are shown in Fig. 5 as the crosshatched bands.

Perhaps a key observation about the scale is that in terms of discrimination ability, the words at the good end of the scale are

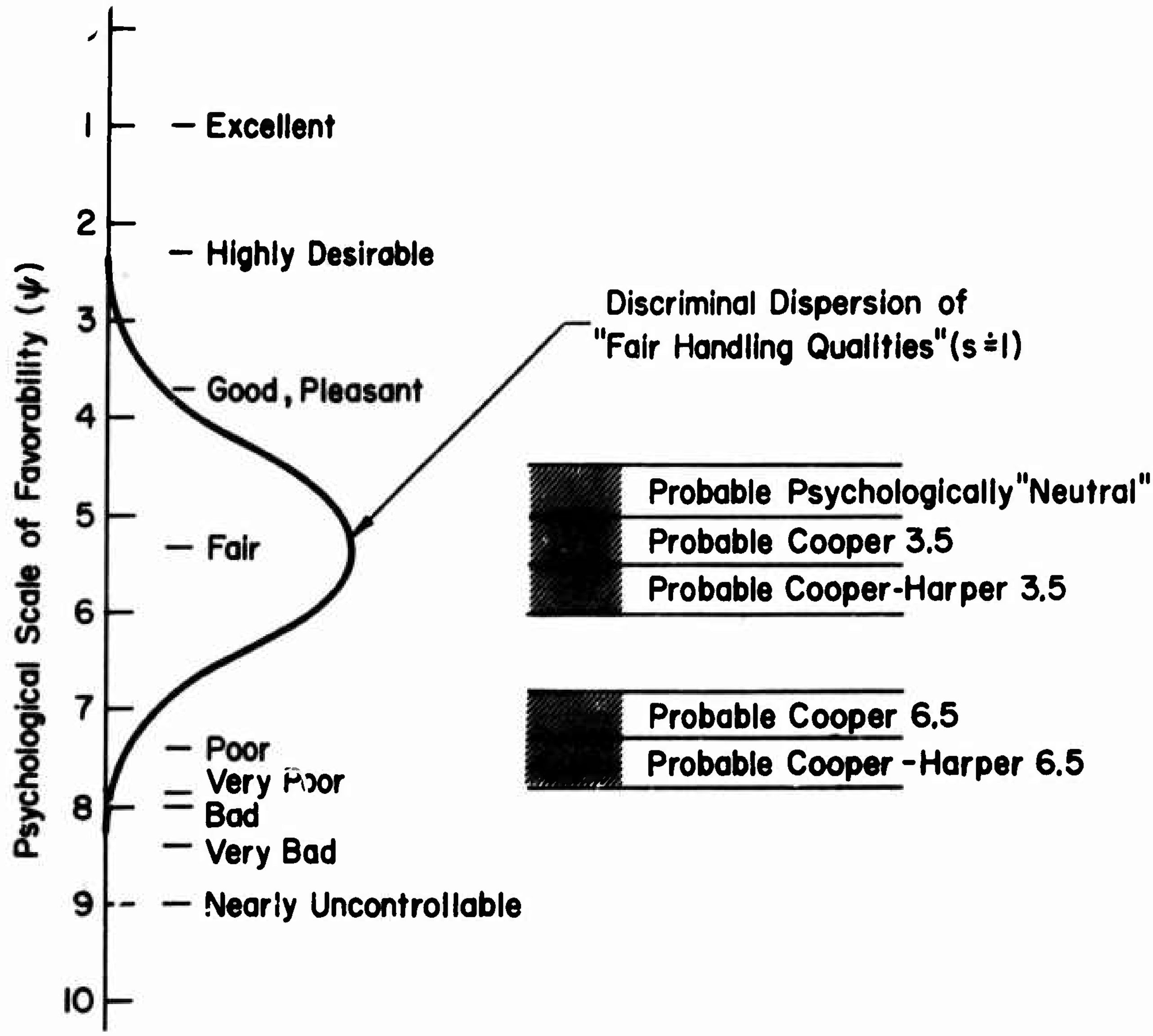


Figure 5. Distribution of the Degrees of Goodness of Handling Qualities Along the ψ-Scale

much more distinct to a rater than at the bad end. The discriminal dispersion is sketched on the figure ($s \doteq 1$) at the "fair" rating. Recalling that dispersions along the ψ-scale are nearly constant, it can be seen that a considerable amount of overlap (demonstrating confusion, or ambiguity) in dispersions exist for words at the bad end of the scale.

Thus, if a rater were to have as a tool the words shown in Fig. 5, we would expect to observe considerably more scatter in the ratings of a bad vehicle. Evidence supporting this contention is sparse, due to experimenters' habits of averaging data before publication, but some support exists in works by Jex (18) and Durand (19). Figure 5 does lead us to suspect that we are fooling ourselves when we place great weight on the fine distinctions made by raters near the "bad" end of current rating scales. Let us look more closely at the connections between the ψ scale and contemporary rating scales.

2. Connections Between the ψ, Cooper, and Cooper-Harper Scales

The list of phrases presented to the 63 participants in the survey contained a large proportion of the individual statements describing Cooper ratings (1) and Cooper-Harper (C-H) ratings (3). A plot of these statements is shown in Fig. 6. Space is not available to label each point by its phrase, but Table XI shows the scale values in superscript following each phrase of the C-H scale. These statements and values were culled from Appendix B to reconstruct the scale and are the data correlations at the most fundamental level. It is clear that some of the phrases describing the poorer ratings are not even ordinal, as shown by the nonmonotonic nature of the data at the ratings of 7 and 8.

A smoothed-over view of the data is obtained by the curve fit shown in Fig. 6. The curve fits reasonably well and is given by

$$\psi = 1 + 8 \log R \qquad (8)$$

A fit which is only slightly less accurate, but which would be easier to handle mathematically in some cases is

$$R = a\psi^2 + b \qquad (9)$$

and is also shown in Fig. 6. Figure 7 shows a logarithmic plot of the same data.

A precise view of the data would take into account the flat spots in the Cooper, Cooper-Harper ratings around 3–4, and 6–8 which indicate that the adjectives used to describe differences in these regions are inadequate for discrimination.

TABLE XI

THE COOPER-HARPER SCALE (3)

CONTROLLABLE CAPABLE OF BEING CONTROLLED OR MANAGED IN CONTEXT OF MISSION, WITH AVAILABLE PILOT ATTENTION	ACCEPTABLE MAY HAVE DEFICIENCIES WHICH WARRANT IMPROVEMENT, BUT ADEQUATE FOR MISSION. PILOT COMPENSATION, IF REQUIRED TO ACHIEVE ACCEPTABLE PERFORMANCE, IS FEASIBLE.	SATISFACTORY MEETS ALL REQUIREMENTS AND EXPECTATIONS, GOOD ENOUGH WITHOUT IMPROVEMENT CLEARLY ADEQUATE FOR MISSION.	EXCELLENT,$^{1.0}$ HIGHLY DESIRABLE$^{2.25}$	A1
			GOOD;$^{3.7}$ PLEASANT;$^{3.7}$ WELL BEHAVED	A2
			FAIR.$^{5.3}$ SOME MILDLY UNPLEASANT CHARACTERISTICS.$^{5.7}$ GOOD ENOUGH FOR MISSION WITHOUT IMPROVEMENT.$^{5.7}$	A3
		UNSATISFACTORY RELUCTANTLY ACCEPTABLE. DEFICIENCIES WHICH WARRANT IMPROVEMENT. PERFORMANCE ADEQUATE FOR MISSION WITH FEASIBLE PILOT COMPENSATION.	SOME MINOR BUT ANNOYING DEFICIENCIES.$^{5.6}$ IMPROVEMENT IS REQUESTED.$^{6.0}$ EFFECT ON PERFORMANCE IS EASILY COMPENSATED FOR BY PILOT.$^{5.3}$	A4
			MODERATELY OBJECTIONABLE DEFICIENCIES.$^{6.5}$ IMPROVEMENT IS NEEDED.$^{6.6}$ REASONABLE PERFORMANCE REQUIRES CONSIDERABLE PILOT COMPENSATION.$^{6.8}$	A5
			VERY OBJECTIONABLE DEFICIENCIES.$^{7.9}$ MAJOR IMPROVEMENTS ARE NEEDED.$^{7.6}$ REQUIRES BEST AVAILABLE PILOT COMPENSATION TO ACHIEVE ACCEPTABLE PERFORMANCE.$^{7.7}$	A6
	UNACCEPTABLE DEFICIENCIES WHICH REQUIRE MANDATORY IMPROVEMENT. INADEQUATE PERFORMANCE FOR MISSION EVEN WITH MAXIMUM FEASIBLE PILOT COMPENSATION.		MAJOR DEFICIENCIES$^{7.9}$ WHICH REQUIRE MANDATORY IMPROVEMENT$^{8.0}$ FOR ACCEPTANCE. CONTROLLABLE. PERFORMANCE INADEQUATE FOR MISSION, OR PILOT COMPENSATION REQUIRED FOR MINIMUM ACCEPTABLE PERFORMANCE IN MISSION IS TOO HIGH.$^{7.1}$	U7
			CONTROLLABLE WITH DIFFICULTY.$^{7.6}$ REQUIRES SUBSTANTIAL PILOT SKILL AND ATTENTION TO RETAIN CONTROL AND CONTINUE MISSION.$^{7.5}$	U8
			MARGINALLY CONTROLLABLE IN MISSION.$^{8.3}$ REQUIRES MAXIMUM AVAILABLE PILOT SKILL AND ATTENTION TO RETAIN CONTROL.$^{8.9}$	U9
UNCONTROLLABLE CONTROL WILL BE LOST DURING SOME PORTION OF MISSION.			UNCONTROLLABLE IN MISSION.	10

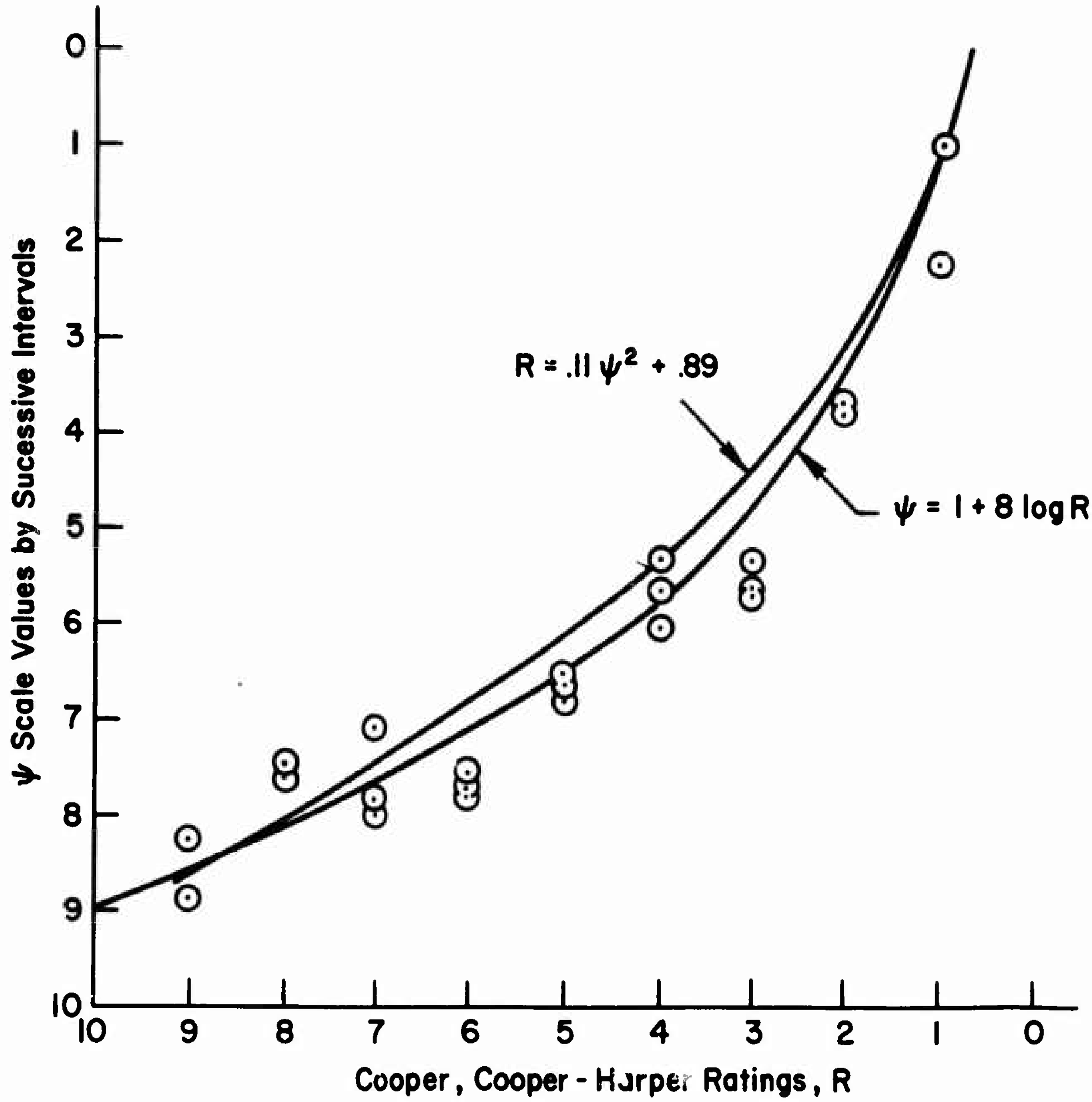


Figure 6. A Comparison of Cooper and Cooper-Harper Ratings with Corresponding ψ-Scale Values (Linear Scale)

Either the data points themselves or the "smoothed" relationships of Eqs. 8 and 9 provide a means to average data obtained from contemporary scales. As will be recalled, to obtain the best estimate of the true value of a measured quantity, the data to be averaged should come from an instrument with constant sensitivity along its scale. Thus, although it could be argued that since ψ and R are functionally related either could be averaged; the desired quantity is ψ. We have argued earlier that from other considerations (i.e., prothetic versus metathetic) the ψ scale data, by virtue of its linearity with subjective magnitude,

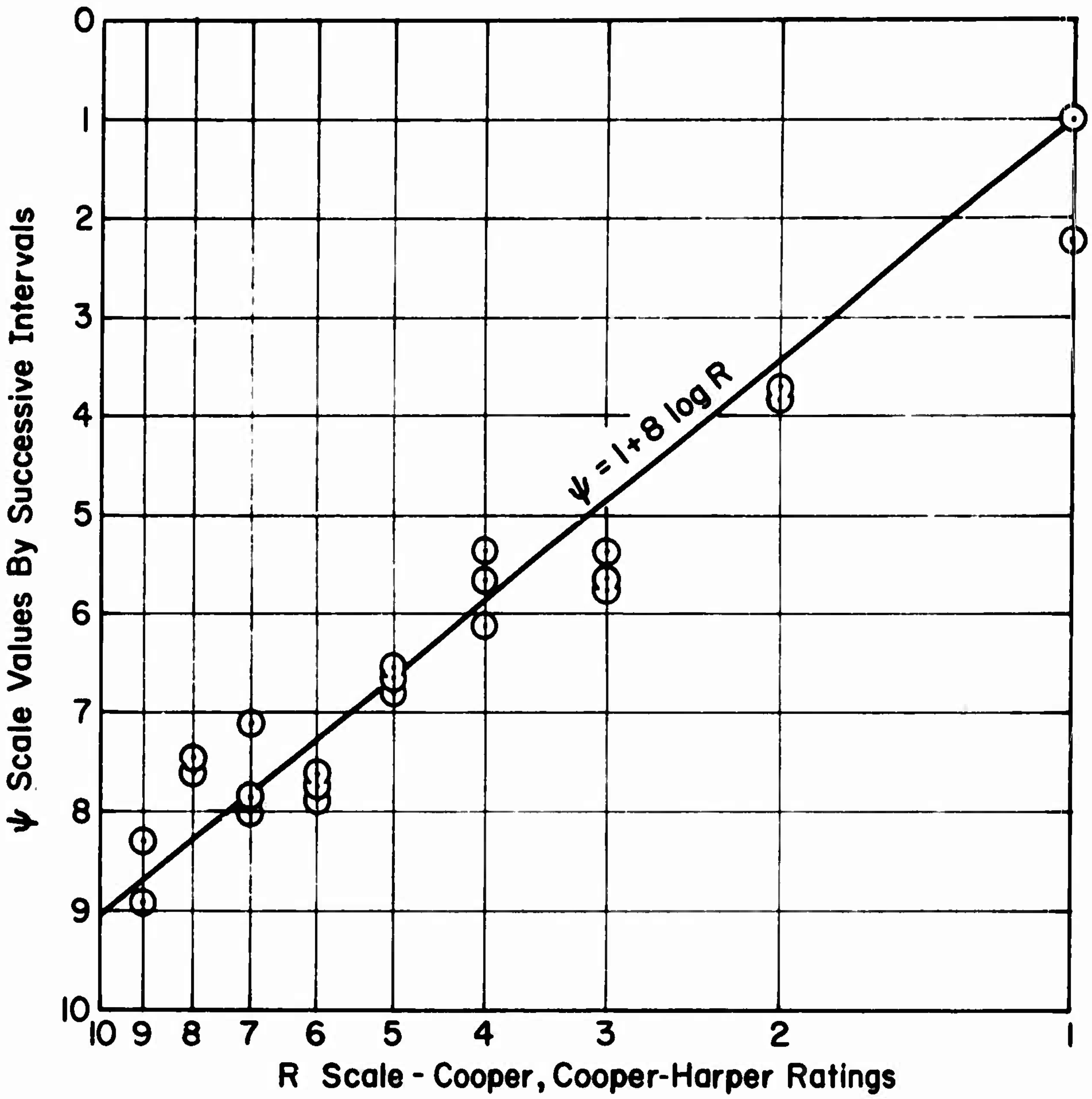


Figure 7. A Comparison of Cooper and Cooper-Harper Ratings with Corresponding ψ-Scale Values (Log Scale)

should be the quantitites which are averaged. The constant discriminal dispersion and the linearity of subjective magnitude are just two ways to reach the same conclusion; that is, the best estimate of a rater's subjective opinion is obtained by averaging the ψ data.

3. Error Introduced by Averaging Cooper-like Ratings

Let us try to estimate the error which would be introduced by averaging Cooper ratings directly instead of using the ψ transformation of Fig. 6. The assumption is that a large number of ratings, R, are available for one vehicle, so that there is no question of there being a difference in means. Based on our earlier discussion, the true mean is obtained by averaging the ψ values for a set of observations. For convenience in analytic treatment, the "smoothed" fit of Eq. 9 will be used.

Let us call this true mean along the R axis $\bar{R}_T$. On the other hand, if our habit has been to average the R values directly, the mean would be $\bar{R}$. Let us define an error, e, which is the difference between averaging the ratings, R, directly and averaging the ψ values obtained by transforming R via Eq. 9 and then converting back to R. Let

$$e = \bar{R} - \bar{R}_T \tag{10}$$

Then

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i \tag{11}$$

and

$$\bar{R}_T = a\bar{\psi}^2 + b \tag{12}$$

Recalling that the variance of ψ is given by

$$\sigma_\psi^2 = \frac{1}{n}\sum(\psi_i - \bar{\psi})^2 = \frac{1}{n}\sum\psi_i^2 - \bar{\psi}^2 \tag{13}$$

We can solve for $\bar{\psi}^2$ and substitute from Eq. 9,

$$\bar{\psi}^2 = \frac{1}{n}\sum\psi_i^2 - \sigma_\psi^2 = \frac{1}{n}\sum\left(\frac{R_i - b}{a}\right) - \sigma_\psi^2$$

$$= \frac{1}{a}\bar{R} - \frac{b}{a} - \sigma_\psi^2 \tag{14}$$

So, from Eq. 12,

$$\bar{R}_T = \bar{R} - a\sigma_\psi^2 \tag{15}$$

Finally, from Eqs. 10 and 15,

$$e = a\sigma_\psi^2 \tag{16}$$

From Fig. 6 it is seen that the values $a = 0.11$, $b = 0.89$ give a good fit, and it will be recalled that σ_ψ is the discriminal dispersion which we found from the successive interval method (see Fig. 5 and Appendix B) and which is approximately constant with a value of unity along the ψ axis. Thus, the errors obtained by averaging Cooper-Harper ratings are given by Eq. 16 as $0.11 \times 1^2 \doteq 0.1$ Cooper unit. This is an optimistic calculation since it made the assumption that the only errors in the rating were due to the nature of the scale itself. Other errors are likely in the rating situation (i.e., bias between raters due to training, experience, etc.) so this would be the limiting best case.

4. Determination of Necessary Trial Size

Although we have demonstrated that very little error is introduced by averaging Cooper ratings directly, the assumption was made that an adequate quantity of data were available to give a high level of confidence. Let us see what sample size requirements are. It should be obvious from Fig. 5 that more reliable data, in terms of Cooper ratings, are obtained at the "good" end of the Cooper scale. Let us consider the case where an experimenter is trying to compare two slightly different (he thinks) vehicles. In the past, experimenters have liked to distinguish between vehicles differing by one Cooper unit. Let us hypothesize that case, and compute the number of trials that should have been made to achieve a confidence of 95 percent.

The t-test will be used, which requires that the variance of the data be known and be approximately the same for the two independent samples. This requirement is reasonably met by the conditions here, since the variance along the R scale changes very little in one Cooper unit. We shall have to calculate its magnitude, however, since we only know that the variance is constant on the ψ scale at this time.

a. **Computation of Variance Along the R Scale.** From Eq. 9 we know that

$$R_i = a\psi_i^2 + b$$

so that

$$R_i - \sigma_{R_i} = a(\psi_i - \sigma_{\psi_i})^2 + b \qquad (17)$$

Solving for σ_{R_i}, we obtain

$$\sigma_{R_i} = a\sigma_{\psi_i}(2\psi_i - \sigma_{\psi_i}) \qquad (18)$$

Since $2\psi_i >> \sigma_{\psi_i}$,

$$\sigma_{R_i} \doteq 2a\sigma_{\psi_i}\psi_i \qquad (19)$$

In terms of the R continuum,

$$\sigma_{R_i}^2 \doteq 4a^2\sigma_{\psi_i}^2\left(\frac{R_i - b}{a}\right) = 4a\sigma_{\psi_i}^2(R_i - b) \qquad (20)$$

Since the t-test requires that the variances of both samples be equal, we shall calculate $\sigma_{R_i}^2$ at the $R = m + 1/2$ points and let those values approximate the variances at m and m + 1.

b. **The t-Test for Difference of Means.** The minimum trial size can be simply determined from the t-test. Given two sets of independent observations, form the sample statistic

$$t_c = \frac{\bar{R}_1 - \bar{R}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (21)$$

with sample variance σ^2 and with $n_1 + n_2 - 2$ degrees of freedom [Hald, p. 391 (20)]. We will specify the minimum difference of means which we want to detect as $|\bar{R}_1 - \bar{R}_2| = 1.0$ Cooper unit. For an equal number of observations in each set, the sample statistic becomes

$$t_c = \frac{1}{\sigma}\sqrt{\frac{n}{2}} \qquad (22)$$

Substituting Eq. 20 for σ,

$$t_c = \frac{\sqrt{n}}{2.82\sqrt{0.11R - 0.098}} \qquad (23)$$

which is plotted in Fig. 8 for several values of R.

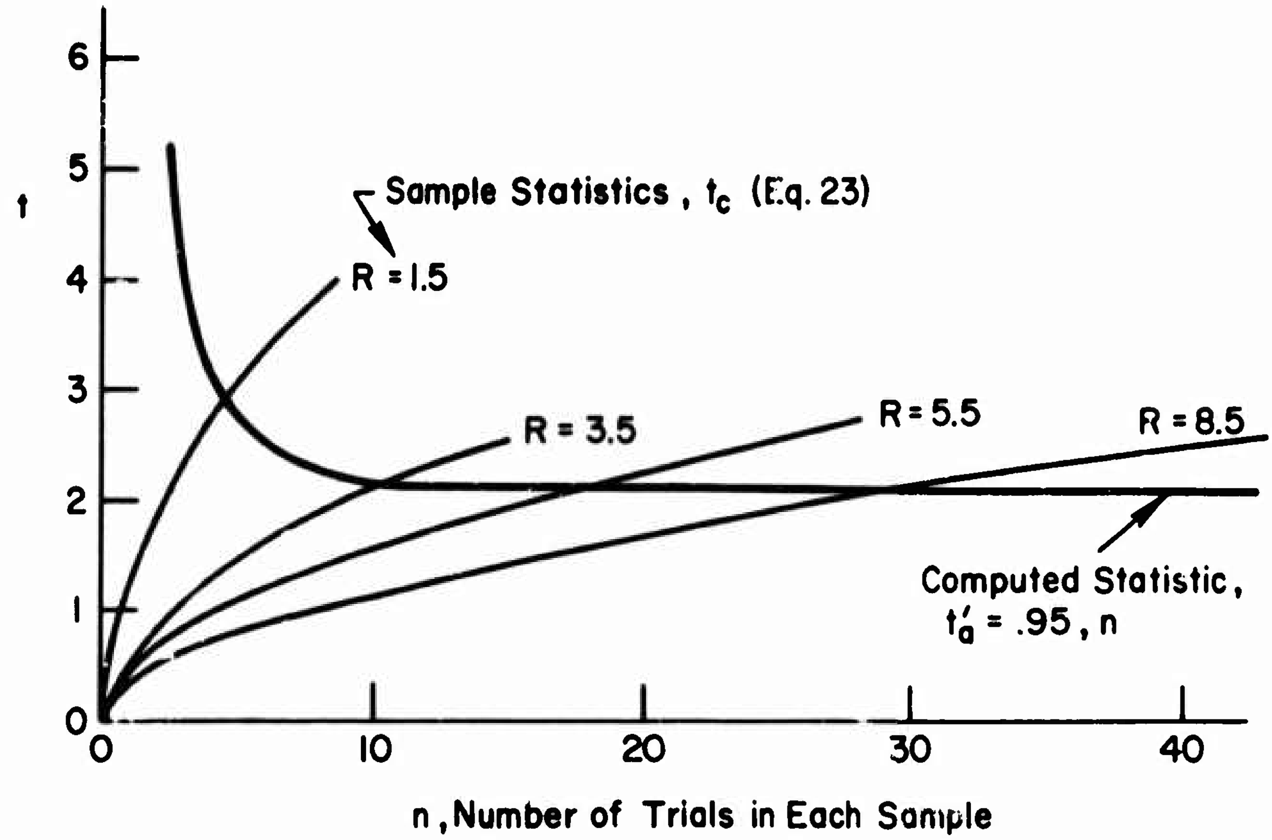


Figure 8. Trial Size Determination from the t-Test

The sample statistic, t_c, is to be compared with the computed statistic, $t'_{a,n}$ based on tables of the t-distribution. The tables give $t'_{a,n}$ at the a-level of confidence and for $n-1$ degrees of freedom. The table values of t' are also plotted in Fig. 8 for $a = 95$ percent. The condition indicating a significant difference in means of 1.0 at the 95 percent confidence level requires that

$$t' < t_c \tag{24}$$

It can be seen that the number of trials is a function of location along the R scale, as we originally expected. If the locus of points where $t' = t_c$ is plotted, the number of trials as a function of R will be available. This has been done in Fig. 9, where it can be seen that

$$n = 3.5R \tag{25}$$

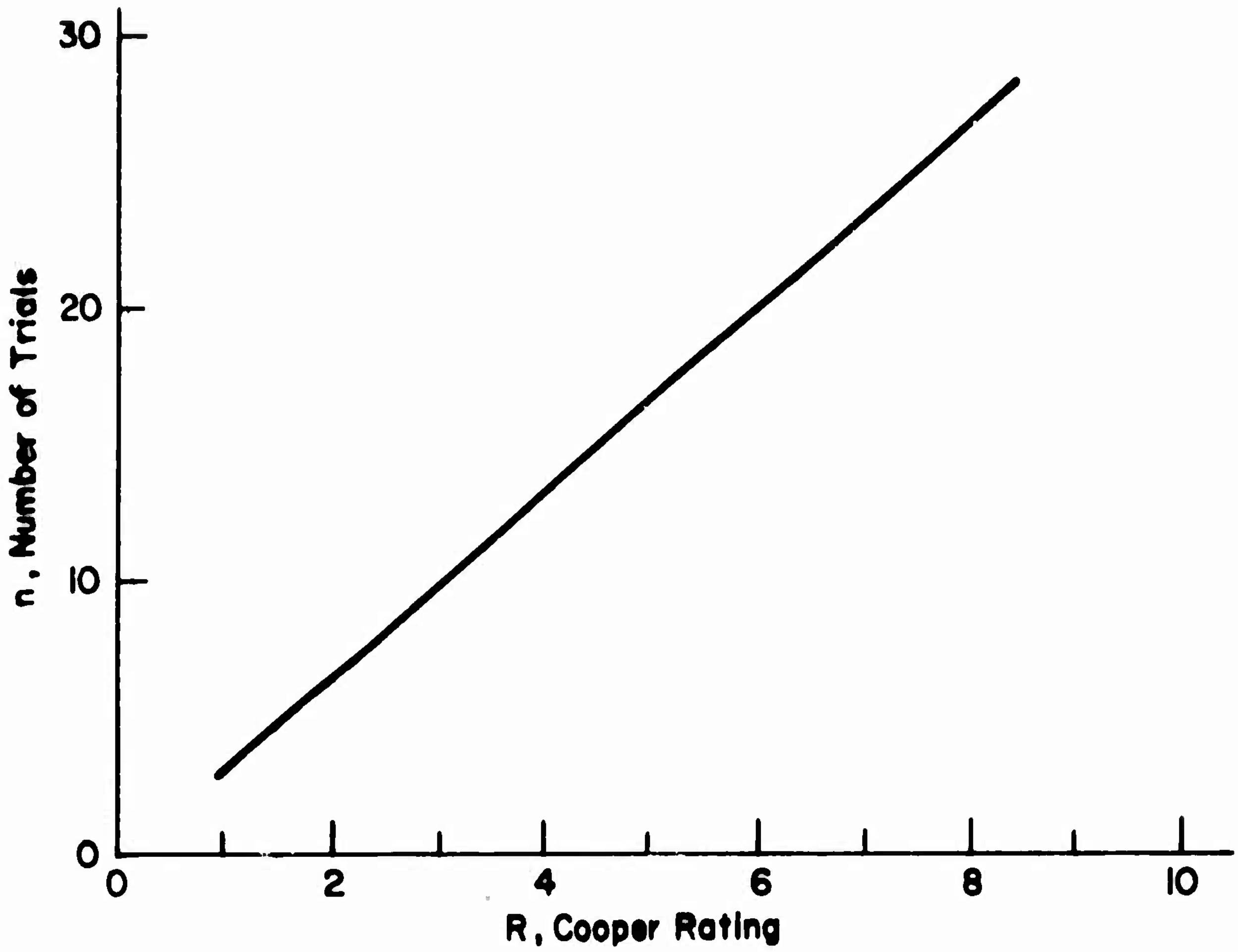


Figure 9. Minimum Number of Trials Required in Each of Two Independent Samples to Determine that the Sample Means are Different by One Cooper Unit with 95 Percent Confidence

These results are somewhat surprising. For a vehicle which is near the Cooper "Acceptable/Unacceptable" boundary, Fig. 9 indicates that approximately 20 trials would be needed for high confidence. Remember, too, that these calculations are optimistic, i.e., sources of variability other than semantic ambiguity have not been considered. We have used the "average rater," one who has the rating characteristics shown in Fig. 6.

It is highly doubtful that trial sizes on the order of 20 have been obtained in practice, which means that the level of confidence is lower than 95 percent in the measures. Here is another reason to keep careful records and publish the raw data. In any event, Eq. 25 shows that for any given confidence level, the number of observations made for "bad" vehicles should be increased an order of magnitude over "good" ratings if the Cooper scale is used.

E. SUMMARY AND CONCLUSIONS

In this section we have established a rationale for quantitative handling qualities ratings using psychological measurement techniques. In addition to determining numerical values for 63 descriptions (see Appendix B, Table B-I), which should be useful in constructing any scale, we have shown that contemporary scales (i.e., Cooper) are very nearly functionally related to the underlying quantitative scale. The smooth appearance of the function (for example, Fig. 6) demonstrates that a very large amount of thought and wisdom went into these original scales, and also demonstrates why subsequent improvement has been so difficult. The data shows that very little error is introduced by averaging Cooper ratings directly rather than transforming to the quantitative ψ scale. However, in order to obtain adequate data for averaging, and to place any weight on differences of one or two Cooper units, a large number of trials will have to be made, particularly when the vehicle is "bad" (see Fig. 9).

With an underlying quantitative scale now established, the next step will be to construct several scales, then test them with some actual rating experiments. In the next section, the experiment will be described together with the rating scales which were used and the measurements which were taken.

SECTION IV

DESCRIPTION OF THE EXPERIMENT

A. OBJECTIVES

With the more complete understanding of rating scales which has been obtained in the previous sections, we are now in a position to conduct rating experiments. The general objectives of the experimental program are to determine the factors which influence pilot opinion and to determine if a modified scale (or scales) would be an improvement over present scales.

B. EXPERIMENTAL PLAN AND SETUP

1. Single-Loop Experiments

a. **Simulation.** A fixed-base simulator with a CRT display and fighter-aircraft-type center stick was used for the experiment. Compensatory tracking in pitch was used for the primary rating task, with the dynamics being simulated on a GEDA analog computer and displayed with a horizon bar-like line on the CRT. A roll axis was also mechanized to enable a secondary tracking task, so that the CRT horizon bar could both pitch and roll. The sensing was inside-out, i.e., as in a conventional aircraft artificial horizon, and is sketched in Fig. 10. At the distance that the pilot sat from the CRT (about 46 cm), a one cm displacement in θ subtended an angle of 1.25 deg at the pilot's eye. The spring gradients of the stick were 13 N/cm (7.5 lb/in.) for the elevator (δ_e) and 3.5 N/cm (2 lb/in.) for the ailerons (δ_a).

A random-appearing sum of twelve sinusoids was used as the command input to the pitch axis. Three bandwidths (1.88, 2.89, and 4.77 r/s) and three amplitudes (0.5, 1.0, and 1.5 cm rms) of input were available. The frequencies of the input were selected to be suitable for a 100 sec run length, and are given in Table XII together with the number of cycles in a run length. The sinudoids making up the shelf, i.e., the frequencies beyond the bandwidth, have an amplitude 14 dB down from the main rectangular portion of the input. A sketch of the spectral characteristics for the 1.88 r/s, 0.5 cm rms input is shown in Fig. 11, and is labeled B6"-1.88-0.5 in accordance with the convention used by McRuer (5).

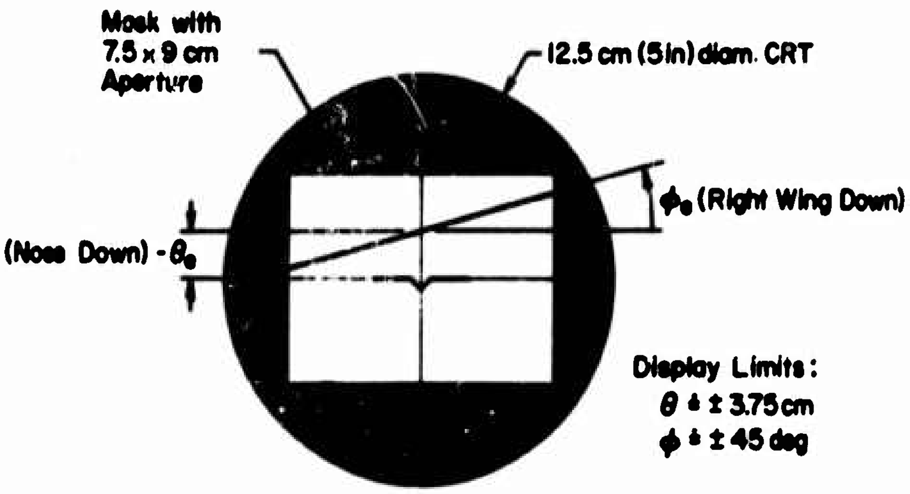


Figure 10. CRT Display for Single-Loop Plus Secondary Tasks

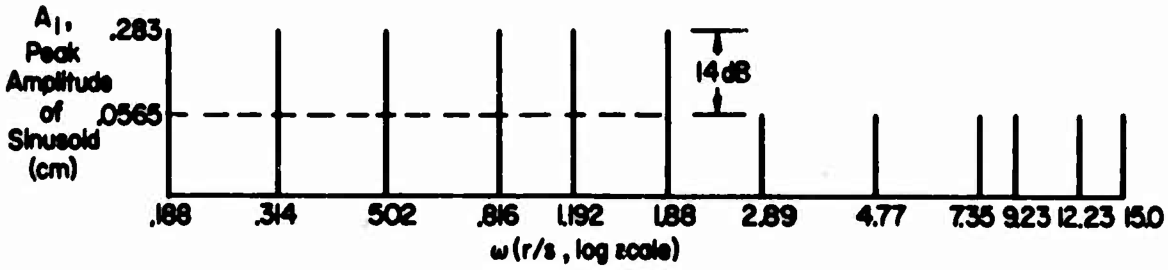


Figure 11. B6"-1.88-0.5 Input Spectrum

TABLE XII

INPUT FREQUENCY COMPONENTS

COMPONENT NO.	COMPONENT FREQUENCY, ω (rad/sec)	NO. OF CYCLES/100 SEC, n = 100ω/2π
1	0.188	3
2	0.314	5
3	0.502	8
4	0.816	13
5	1.192	19
6	1.88	30
7	2.89	46
8	4.77	76
9	7.35	117
10	9.23	147
11	12.23	195
12	15.00	239

b. **Controlled Elements.** As discussed earlier (Section II.B.3), single-loop compensatory tracking with a few simple controlled elements will adequately describe a variety of vehicle/task configuration. Thus, in the main rating experiments we shall use the idealizations studied by McRuer (5) together with some additional simplified elements necessary to obviate the requirement for certain inferred correlations related to pilot lead. A matrix of controlled elements, together with the gains and the command inputs used with each is shown in Table XIII (the key to the inputs of Table XIII is given in Table XIV).

As will be noted from Table XIII, the possible number of configurations is considerable if all of the inputs were applied to each controlled element and gain. In order to make the experiment feasible, the experimental design had to yield less than approximately fifty configurations, and at the same time obtain enough data to allow the testing of trends across the many variables (both explicit and implicit) of interest. A detailed look at the finally selected configurations of Table XIII will yield the following:

1. Excellent tests of consistency for K/s and K/s^2 at the $K/K_B = 1$ points of the matrix would be expected in accordance with findings in previous studies of system and operator parameters [McRuer (5)].

2. Trends with gain are provided for six Y_c's, three controlled element forms have five gain levels, and the remaining three Y_c's fill in between the extremes of equalization required with three gain levels. Thus an adequate range of element forms exists for a dynamic gain range of 100.

3. Variation of parameters with input bandwidth and amplitude can be extrapolated to all of the the forms from the $.1K_B/s$, K_B/s, $10K_B/s$, $.1K_B/s^2$, K_B/s^2 and $10K_B/s^2$ points.

4. A good range of each of the myriad system and operator parameters is obtained.

The configurations of Table XIII were thus considered to adequately represent the single-loop tasks of interest.

TABLE XIII

INPUT MATRIX FOR CONFIGURATIONS
(See Table XIV for Input Key)

CONTROLLED ELEMENT, Y_c	CONTROLLED ELEMENT GAIN, K/K_B					K_B $\left(\frac{cm, \theta}{cm\text{-}sec^n, \delta_e}\right)$ *
	0.1	0.5	1	5	10	
K/s	a,c	a	a,a,b,c,c d,e,f,g	a	a,b,c	0.586
$K/s(s+4)$	a		a		a	2.15
$K/s(s+2)$			a			2.15
$K/s(s+1)$	a,a		a,a		a	2.15
K/s^2	a,b	a	a,a,b,c d,e,f,g	a	a,b,c	1.17
$K/s(s-1)$			a			1.075
$K/(s-2)$			a			3.45
$K/[s^2+2(0.7)7.8s+7.8^2]$	a		a		a	8.38
$K/[s^2+2(0.7)16s+16^2]$	a	a	a	a	a	35.2

*n = exponent of free s in denominator of Y_c. $K_B = K_{BEST}$ as determined in an independent set of trials.

TABLE XIV

KEY TO INPUTS OF TABLE XIII

CODE FROM TABLE XIII	INPUT BANDWIDTH, ω_i (r/s)	INPUT AMPLITUDE, σ_i (cm rms)
a	1.88	1
b	1.88	0.5
c	1.88	1.5
d	2.89	0.5
e	2.89	1
f	2.89	1.5
g	4.77	1

c. **Secondary Task (Lateral)**. In an attempt to find a good correlate with pilot opinion, a workload measure of some sort was considered extremely desirable, primarily because a vehicle evaluator invariably expresses some subjective impressions regarding "attention," etc., when in a given rating situation. Experimenters have made numerous workload related measures with secondary tasks such as the extinguishing of lights, mental exercises, tracking tasks, etc., [e.g., Gaul (57)] and have met with varying degrees of success. A number of difficulties are apparent:

1. The scores obtained from secondary tasks (such as number of lights turned out, etc.) are difficult to relate to any measurable characteristics of the system because most are discrete in nature. Those tasks which are continuous have no analytical tie with system parameters.

2. The scores are quite variable since they depend highly upon the subject's motivation and the performance requirements of the task.

3. If it is attempted to force the operator to his capacity via a technique which paces the difficulty of the secondary task, the primary task generally is neglected in favor of the secondary task.

To overcome these difficulties, an unstable tracking task, called the "critical task" [see Jex, et al (22)] was used as a secondary loading task and was mechanized such that it could not become the primary task when the operator was near capacity. The use of the critical task offered the advantages of having an easily adjustable unstable root which is somewhat proportional to task difficulty and is related directly to the operator's time delay while tracking. Thus, although it was not the objective of this program, a workload theory involving system parameters could be evolved at a suitable time. Here we wanted to find an objective measure which was sensitive to handling qualities and thus could be correlated with pilot opinion.

The mechanization scheme used was similar to that proposed by Kelly (23). The difficulty of the secondary task was made proportional to primary task performance. Thus, when the operator was keeping primary system error (performance) less than a criterion value, the secondary difficulty increased. When the operator was so busy with the secondary task that primary error was larger than the criterion value, the secondary difficulty decreased.

The final level of difficulty was determined by the sensitivity of the primary task performance to loading by the secondary task. The results of the experiment will show if this "sensitivity" is a determining factor of pilot opinion.

The secondary task was prevented from becoming the primary task by giving the following instructions to the subject: "Your objective is to get the highest secondary task score you can. To get a high score you must keep the primary task error very small. If you allow the primary error to get large, your score will decrease. The problem will stop if either primary or secondary tasks are allowed to exceed the display limits."

A block diagram of both primary and secondary tasks is shown in Fig. 12. To avoid any confusion over the definition of workload (i.e., is it physiological or psychomotor workload?), the parameter λ shown in the figure was assumed to be related to the "attention level" required of the operator.

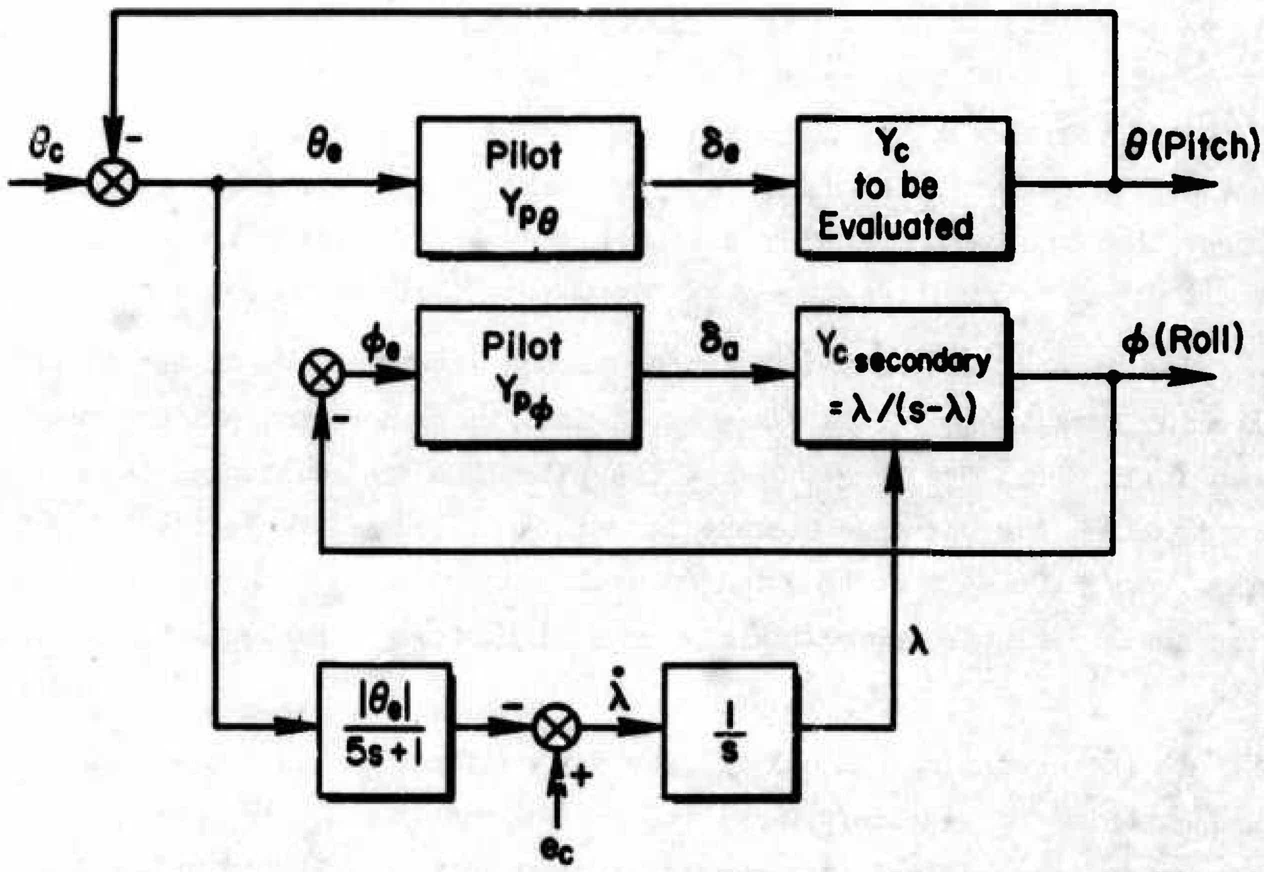


Figure 12. Single-Loop Primary Task
with Secondary Cross-Coupled Loading Task

2. Procedure

The configurations were presented in a randomized sequence to minimize halo effects (recall that a halo effect is the tendency of a subjective response to be influenced by the preceding stimulus-response pair). Repeats were placed carefully in the sequence to balance out time of day and halo effects. A detailed run log is given in Appendix C (Table C-I). The time required for each configuration was eight minutes. The pilot actually had to perform two tasks sequentially. First he was asked to track longitudinally to minimize the pitch error. During this tracking period, which lasted 120 seconds, he was asked to formulate his opinion of the configuration based on the task performance criterion. If more time was required to form an opinion, it could be taken after the 120-second recorded run. Approximately 15 seconds were allowed before each run for the subject to reach steady-state tracking so that the first 100-second portion of the 120-second run would be suitable to use for describing function computations and performance measures. At the completion of the 120-second run, the pilot was asked to write down the deserved ratings on his clipboard forms. He was not allowed to "play" with the configuration because his ratings would then be based on characteristics other than those specified in the task definition. The rating scales used and the task definitions are given in the next subsection.

The second task of the sequence was the determination of the secondary loading task score. This was a multi-axis task where the primary task was still pitch tracking, but now the pilot had the additional task of controlling the unstable element in roll as discussed in Section IV.B.1.c. This very difficult combination generally consumed about 2 minutes. Thus the total 8-minute (approximately) run might follow the sequence shown in Fig. 13.

At the beginning and end of each day calibration runs were made, and a series of secondary task (lateral) alone trials were made to determine the critical (maximum) secondary score attainable under no-load conditions.

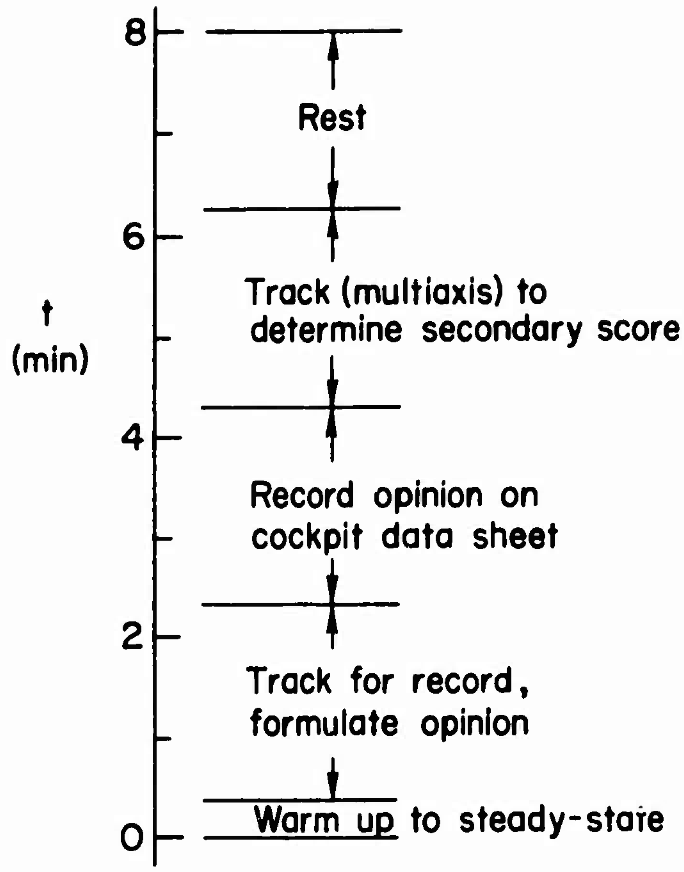


Figure 13. Typical Run Sequence for Each Experimental Configuration

a. **Measures Obtained During Each Run**. In addition to ratings obtained from the pilot, a large amount of objective data were taken. Some data were tape recorded for later use in describing function calculations. Strip chart recordings were made to determine various performance levels attained while tracking and to possibly contribute clues to the causes of the resulting ratings. A digital voltmeter was used to sequentially read out numerous performance measures. The variables recorded in the trials are given in Table XV. One of the variables given, the EMG signal, is perhaps not self-explanatory. The EMG, or electromyograph, was utilized in the experiments to obtain an indication of neuromuscular effort, which could then be correlated with pilot rating. Probes were attached to the pilot's right triceps and were monitored continuously during the experiments. The pre-experiment calibrations included an EMG calibration (stick force versus EMG amplifier output).

TABLE XV

RECORDED MEASURES

		CHANNEL	VARIABLE
HONEYWELL FM TAPE RECORDER	BRUSH STRIP CHART	1	θ_c, Pitch Command Input
HONEYWELL FM TAPE RECORDER	BRUSH STRIP CHART	2	θ_e, Primary Task Error
HONEYWELL FM TAPE RECORDER	BRUSH STRIP CHART	3	δ_e, Pilot's Stick Output
HONEYWELL FM TAPE RECORDER	BRUSH STRIP CHART	4	θ, Pitch Output
HONEYWELL FM TAPE RECORDER	BRUSH STRIP CHART	5	λ_s, Secondary Task Score
HONEYWELL FM TAPE RECORDER	BRUSH STRIP CHART	6	$\int e^2 dt$, Error Performance
HONEYWELL FM TAPE RECORDER	BRUSH STRIP CHART	7	$\dot{\lambda}_s$, Secondary Score Rate (see Fig. 12)
HONEYWELL FM TAPE RECORDER	BRUSH STRIP CHART	8	$EMG/(0.05s+1)$, Muscle Tension
HONEYWELL FM TAPE RECORDER		9	Timing Signal, Master reference timing signal
HONEYWELL FM TAPE RECORDER		11	EMG, Unfiltered myograph signal
HONEYWELL FM TAPE RECORDER		12	Step, Identifies 120-second portion of run
DIGITAL VOLTMETER		1	$\int \lvert\theta_c\rvert dt$, Performance measure
DIGITAL VOLTMETER		2	$\int \lvert\theta_e\rvert dt$, Performance measure
DIGITAL VOLTMETER		3	$\int \lvert\delta_e\rvert dt$, Performance measure
DIGITAL VOLTMETER		4	$\int \lvert\theta\rvert dt$, Performance measure
DIGITAL VOLTMETER		5	$\int EMG/(0.05s+1)dt$, Performance measure

C. RATING SCALES

Using the phrases for which scale values were determined in Section III, scales were constructed to solicit opinion from the pilot during the experiment. A "global" scale was constructed using the degrees of goodness of handling qualities. Opinion was also solicited about the specific traits of "Response Characteristics," "Control," "Demands on the Pilot," and "Effects of Deficiencies." To enable a comparison with already existing scales, Cooper ratings (Ref. 1) and Cooper-Harper ratings (Ref. 3) were also obtained. The number of ratings required of the pilot were thus considerable, but it was found that the 3" X 5" cards containing the scales could be flipped through quickly when the pilot became familiar with them (with the exception of the Cooper-Harper scale, which was presented on 8-1/2" X 11" paper, as shown). The scales are shown in Table XVI, where the number in the upper right-hand corner of each box represents its position in the sequence of presentation.

Two pilots participated in the experiments. One was an engineer-pilot, the other a pilot from Aerospace Test Pilots' School at Edwards AFB. The instructions to the pilots are repeated in Appendix C.

TABLE XVI

RATING SCALES USED IN EXPERIMENTS

SINGLE-LOOP

Task Specification: Simulation of a tail-chase condition in pitch (longitudinal only) for gunnery run. Lead aircraft taking evasive action. Condition might last 3-10 min in real life. View is through a gun sight.

Maneuver: Compensatory tracking in pitch-minimize error.

Inputs:

Command: Random input simulating evasive action of lead plane.

Gust: None

Duration of Task: 2 min

Performance Specification: Best gunnery results will probably be obtained if error is kept less than .75 cm.

Peripheral Loads (i.e., a second axis to control, tuning of radios, etc.): None

①

PILOT OPINION RATING SCHEDULE

	Adjective rating	Numerical rating	Description	Primary mission accomplished	Can be landed
Normal operation	Satisfactory	1	Excellent, includes optimum	Yes	Yes
		2	Good, pleasant to fly	Yes	Yes
		3	Satisfactory, but with some mildly unpleasant characteristics	Yes	Yes
Emergency operation	Unsatisfactory	4	Acceptable, but with unpleasant characteristics	Yes	Yes
		5	Unacceptable for normal operation	Doubtful	Yes
		6	Acceptable for emergency condition only[1]	Doubtful	Yes
No operation	Unacceptable	7	Unacceptable even for emergency condition[1]	No	Doubtful
		8	Unacceptable - dangerous	No	No
		9	Unacceptable - uncontrollable	No	No

TABLE XVI (Continued)

②

Answer the following questions in order:

	Yes	No
1. Is the vehicle controllable during the task?	☐	☐
2. Is the vehicle acceptable for the task? (May have deficiencies which warrant improvement, but is adequate for the task.)	☐	☐
3. Is the vehicle satisfactory for the task? (i.e., adequate for the task without improvement.)	☐	☐

HANDLING QUALITIES ③

Rating	
0	
1	-Excellent
2	-Highly desirable
3	
4	-Good, pleasant
5	-Fair
6	
7	
8	-Bad
	-Very bad
9	-Nearly uncontrollable
10 ☐	Uncontrollable

TABLE XVI (Continued)

RESPONSE CHARACTERISTICS ④

0
- Excellent, pure (i.e., no accidental excitation) primary and secondary response characteristics

1

2

3
- Good, relatively pure, primary and secondary response characteristics

4

5
- Fair, somewhat impure, primary or secondary response characteristics

6
- Quite sensitive, sluggish, or uncomfortable in primary or secondary responses

7
- Extremely sensitive, sluggish, or uncomfortable in primary or secondary responses

8

9 - Nearly uncontrollable

10 ☐ Uncontrollable

☐ Not applicable

CONTROL ⑤

0

1 - Extremely easy to control with excellent precision

2
- Very easy to control with good precision

3

4
- Easy to control with fair precision

5

6
- Controllable with somewhat inadequate precision

7
- Controllable, but only very imprecisely
- Difficult to control

8 - Very difficult to control

9 - Nearly uncontrollable

10 ☐ Uncontrollable

☐ Not applicable

TABLE XVI (Continued)

DEMANDS ON PILOT ⑥

0
1
2
- Completely undemanding, very relaxed and comfortable

3
- Largely undemanding, relaxed

4
5
- Mildly demanding of pilot attention, skill, or effort

6
- Demanding of pilot attention, skill, or effort

7
- Very demanding of pilot attention, skill, or effort

8
- Completely demanding of pilot attention, skill, or effort

9 - Nearly uncontrollable

10 ☐ Uncontrollable

☐ Not applicable

EFFECTS OF DEFICIENCIES ⑦

0
1
2
3
4
5
- Effects of deficiencies on performance is easily compensated for by pilot

6
- Moderately objectionable deficiencies

7
- Major, very objectionable deficiencies

8
9 - Nearly uncontrollable

10 ☐ Uncontrollable

☐ Not applicable

TABLE XVI (Concluded)

⑧

CONTROLLABLE CAPABLE OF BEING CONTROLLED OR MANAGED IN CONTEXT OF MISSION, WITH AVAILABLE PILOT ATTENTION	ACCEPTABLE MAY HAVE DEFICIENCIES WHICH WARRANT IMPROVEMENT, BUT ADEQUATE FOR MISSION. PILOT COMPENSATION, IF REQUIRED TO ACHIEVE ACCEPTABLE PERFORMANCE, IS FEASIBLE.	SATISFACTORY MEETS ALL REQUIREMENTS AND EXPECTATIONS, GOOD ENOUGH WITHOUT IMPROVEMENT CLEARLY ADEQUATE FOR MISSION.	EXCELLENT, HIGHLY DESIRABLE	A1
			GOOD, PLEASANT, WELL BEHAVED	A2
			FAIR. SOME MILDLY UNPLEASANT CHARACTERISTICS. GOOD ENOUGH FOR MISSION WITHOUT IMPROVEMENT.	A3
		UNSATISFACTORY RELUCTANTLY ACCEPTABLE. DEFICIENCIES WHICH WARRANT IMPROVEMENT. PERFORMANCE ADEQUATE FOR MISSION WITH FEASIBLE PILOT COMPENSATION.	SOME MINOR BUT ANNOYING DEFICIENCIES. IMPROVEMENT IS REQUESTED. EFFECT ON PERFORMANCE IS EASILY COMPENSATED FOR BY PILOT.	A4
			MODERATELY OBJECTIONABLE DEFICIENCIES. IMPROVEMENT IS NEEDED. REASONABLE PERFORMANCE REQUIRES CONSIDERABLE PILOT COMPENSATION.	A5
			VERY OBJECTIONABLE DEFICIENCIES. MAJOR IMPROVEMENTS ARE NEEDED. REQUIRES BEST AVAILABLE PILOT COMPENSATION TO ACHIEVE ACCEPTABLE PERFORMANCE.	A6
	UNACCEPTABLE DEFICIENCIES WHICH REQUIRE MANDATORY IMPROVEMENT. INADEQUATE PERFORMANCE FOR MISSION EVEN WITH MAXIMUM FEASIBLE PILOT COMPENSATION.		MAJOR DEFICIENCIES WHICH REQUIRE MANDATORY IMPROVEMENT FOR ACCEPTANCE. CONTROLLABLE. PERFORMANCE INADEQUATE FOR MISSION, OR PILOT COMPENSATION REQUIRED FOR MINIMUM ACCEPTABLE PERFORMANCE IN MISSION IS TOO HIGH.	U7
			CONTROLLABLE WITH DIFFICULTY. REQUIRES SUBSTANTIAL PILOT SKILL AND ATTENTION TO RETAIN CONTROL AND CONTINUE MISSION.	U8
			MARGINALLY CONTROLLABLE IN MISSION. REQUIRES MAXIMUM AVAILABLE PILOT SKILL AND ATTENTION TO RETAIN CONTROL.	U9
UNCONTROLLABLE CONTROL WILL BE LOST DURING SOME PORTION OF MISSION.			UNCONTROLLABLE IN MISSION.	10

SECTION V

ANALYSIS OF THE DATA

A. INTRODUCTION TO THE PILOT MODEL

The correlations which will be made in this section will include parameters of the pilot model, so a very brief summary of the model is in order here. A complete and detailed study of the techniques used to derive the model and the intricacies of parameter adjustment can be found in McRuer (5).

The simple crossover model of a pilot/vehicle combination for a wide variety of controlled elements has been shown to be

$$\left[Y_p Y_c(j\omega)\right]_{\omega \doteq \omega_c} = \frac{\omega_c}{j\omega} e^{-j\tau_e \omega} \tag{26}$$

when the operator is performing a compensatory tracking task. The elements are defined as

Y_p = the pilot describing function

Y_c = the controlled element or vehicle transfer function

ω_c = the system crossover frequency, i.e., the frequency where $|Y_p Y_c| = 1$

τ_e = the effective time delay, i.e., the high-frequency transport lag characteristics observed in the human operator while tracking. Includes neural conduction time delay, cerebral computation times, and limb dynamics time constant.

The model is a frequency domain description of the open-loop system characteristics in the region of crossover and in the presence of sinusoidal, random-appearing inputs. The model given by Eq. 26 describes only the linear behavior of the operator, i.e., that portion of the system output which is correlated with the input. An operator also generates an output that is uncorrelated with the input. This "noisy" portion is called the remnant, and is defined to include all pilot output power not correlated with the input.

The particular values of the two parameters, ω_c and τ_e, which would be exhibited in a manual control situation, depend on numerous factors, including the input characteristics, the form of the controlled element, the nature of the environment (e.g., fixed- or moving-base simulation), motivation, and the nature of the task. By following the adjustment rules given by McRuer (5), the parameters can be closely estimated.

A pilot model corresponding to the crossover model of Eq. 26 for the controlled elements used in this study is

$$Y_p = K_p \frac{(T_L j\omega + 1)}{(T_I j\omega + 1)} e^{-\tau_e j\omega} \tag{27}$$

where K_p = the pilot gain

T_L, T_I = the lead, lag equalization time constants generated internally by the pilot

Note that in order for the crossover model of Eq. 26 to correctly describe the total open-loop system, the pilot must exactly cancel any lead or lag in the controlled element near the crossover region. Available data indicates that he is able to do so, except when the controlled element is the "critical task" of Jex (22). There he is constrained to a behavior which causes the pilot to appear nearly as a gain with a transport lag, so that the total open-loop does not have the usual form of Eq. 26. Equations 26 and 27 will be used to fit the data of this study.

B. VALIDITY OF THE DESCRIBING FUNCTION DATA

1. Computational Approach

Describing functions of the pilot and of the total open-loop were computed using a digital routine (BOMM, Ref. 33) which determined the ratios of the Fourier coefficients of the appropriate time series. Some spectral densities and statistical measures were also computed. The describing functions of interest are given by

$$Y_p = \frac{\delta_{e_i}}{e_i} \tag{28}$$

and

$$Y_p Y_c = \frac{\theta_i}{e_i} \tag{29}$$

where δ_{e_i}, e_i, and θ_i are the Fourier coefficients at the i^{th} frequency for the elevator deflection (pilot output), system error (pilot input), and system output.

An example of the results of the routine is shown in Figs. 14 and 15 for $Y_c = K/s$. The plot of Φ_{ee} and $\Phi_{\delta\delta}$ (power spectra of e and δ) shows the difference between coherent power (at input frequencies) and uncorrelated power (or noise) by denoting the coherent power with the circular symbol. Thus it can be seen that the signal-to-noise ratio was a problem at the lower input frequency of Φ_{ee}. Because the signal at that frequency was obviously contaminated with noise, the corresponding describing function points were marked "unreliable" with a flag in Fig. 15. All of the describing functions of the experiment were treated in a similar manner and are included in Appendix C. Generally, the lower three to five frequency points were found to be unreliable.

The mid- and high-frequency describing function data appear to have been calculated from high quality (noise-free) experimental data. It is these data that should be compared between controlled element forms for internal consistency, and with the previous work of McRuer (5) for compatibility.

The describing function data is included in Appendix C along with a tabulation of the fitted parameters and rating data. Because of the economics involved, describing function data could be computed only for the single-loop runs of JDM. The rating data for pilot MDK, however, is included in Appendix C.

2. Compatibility of Effective Time Delay, Input, Crossover Frequency, and Phase Margin Effects

The points selected in the experiment for comparison with past work were, from Table XIII, $Y_c = K_B/s$ and K_B/s^2, where K_B is the "best" gain as determined in a brief preliminary trial. Several input combinations were used to allow a check of variation with input bandwidth. Figure 16

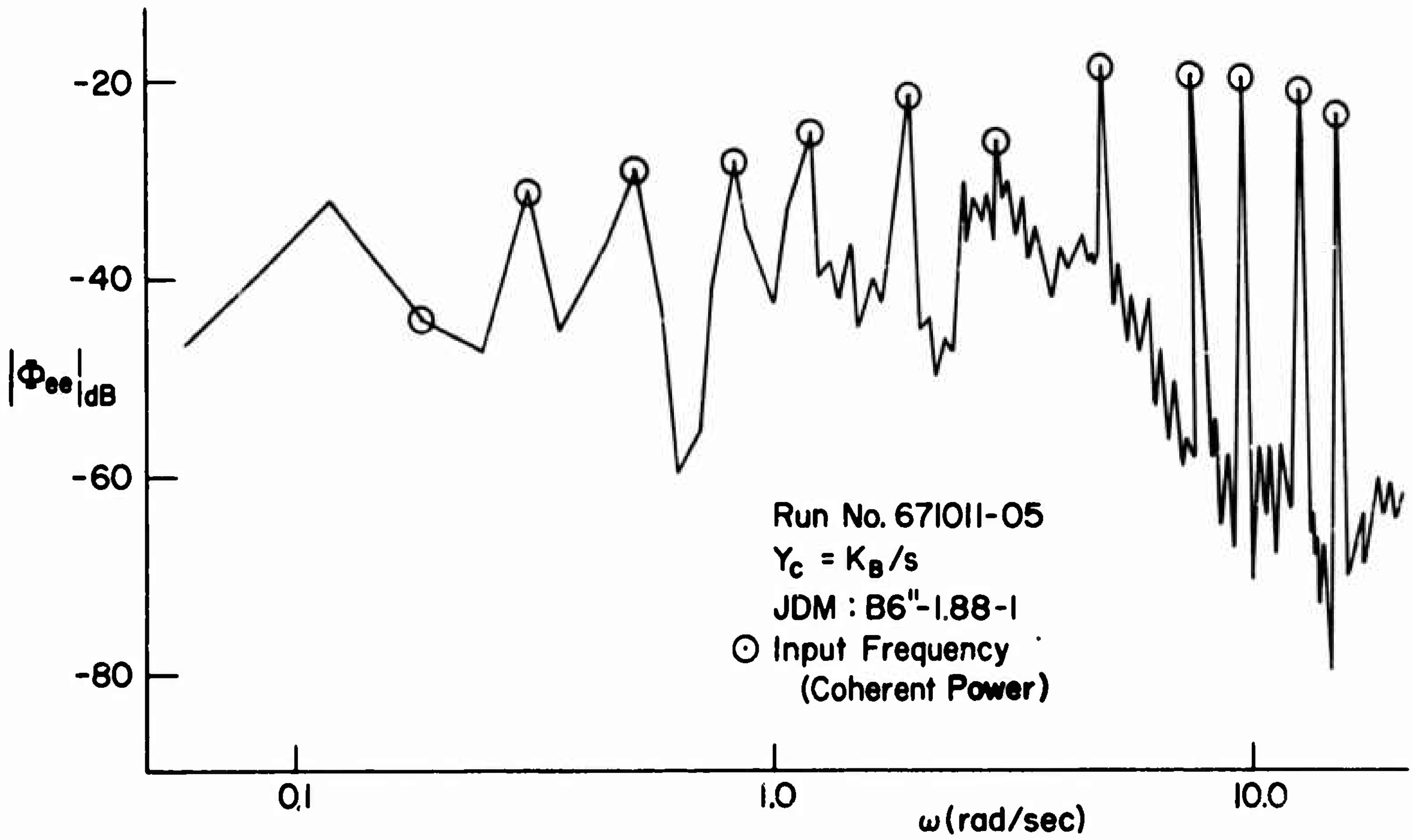


Figure 14a. Typical Error Power Spectrum Showing Problems with the Noise Level at the Low Frequencies for $Y_c = K/s$

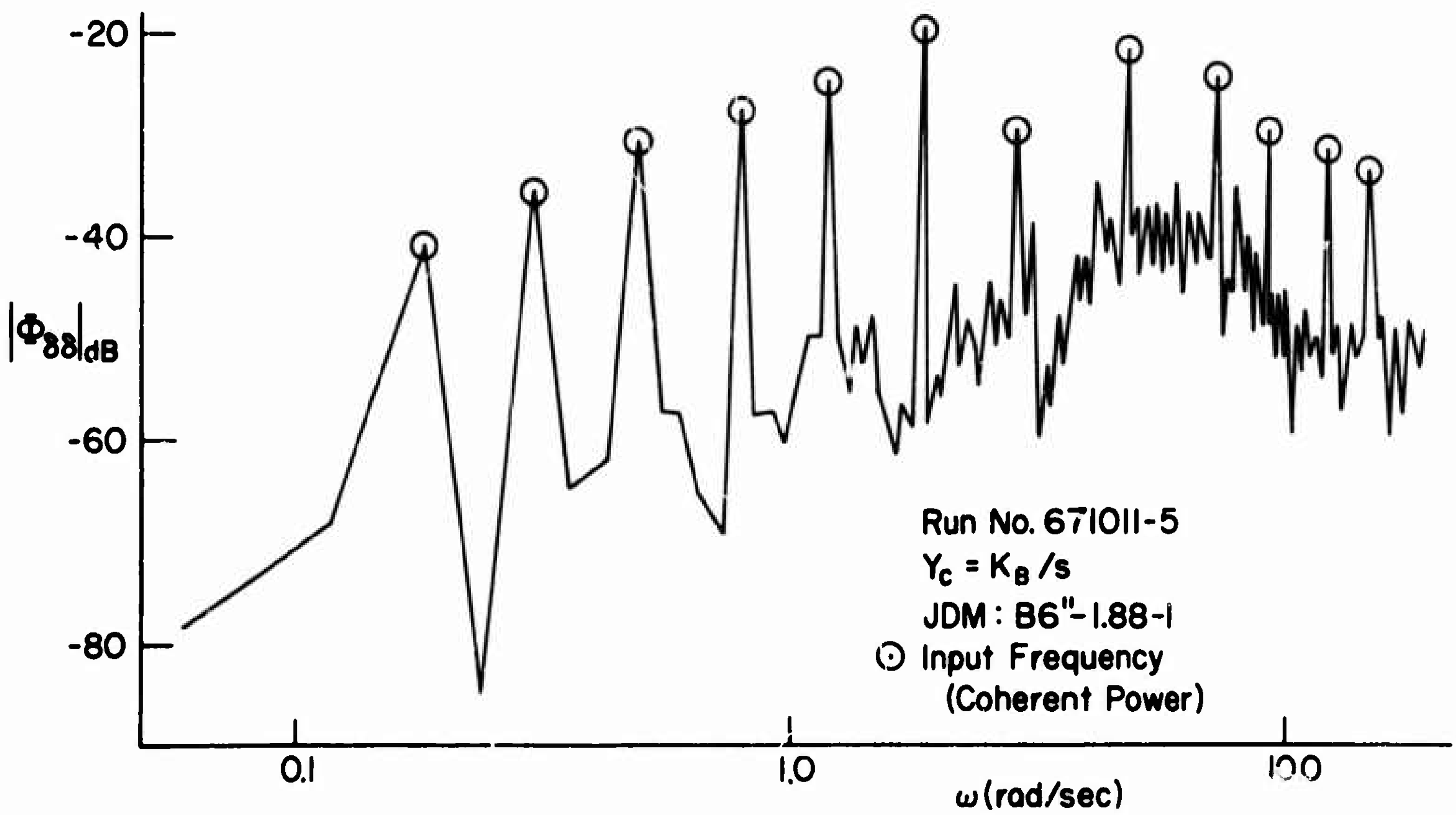


Figure 14b. Typical Pilot Output Power Spectrum Showing Good Signal-to-Noise Properties at All Input Frequencies for $Y_c = K/s$

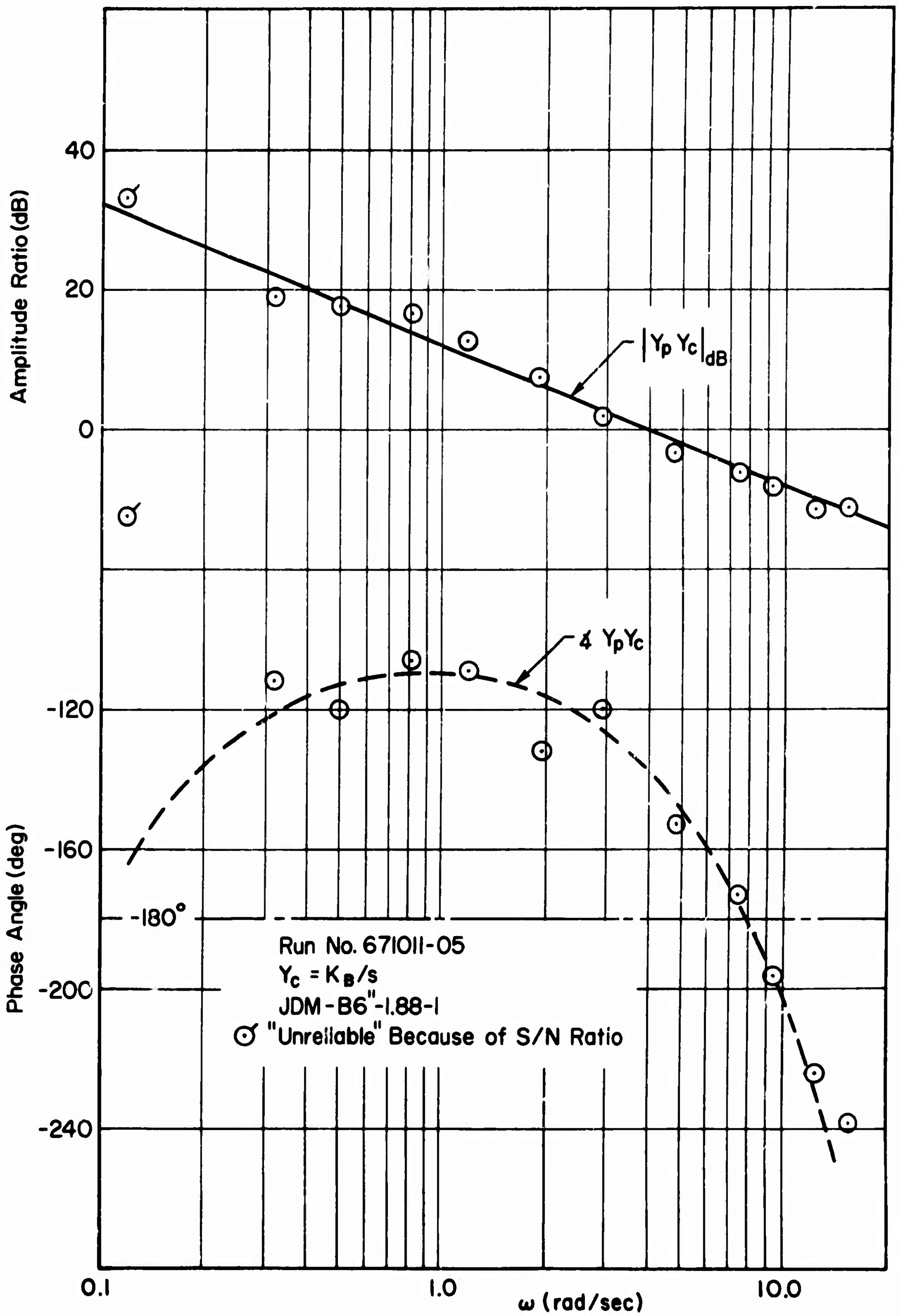


Figure 15. Total Open-Loop Describing Function Computed for $Y_c = K/s$

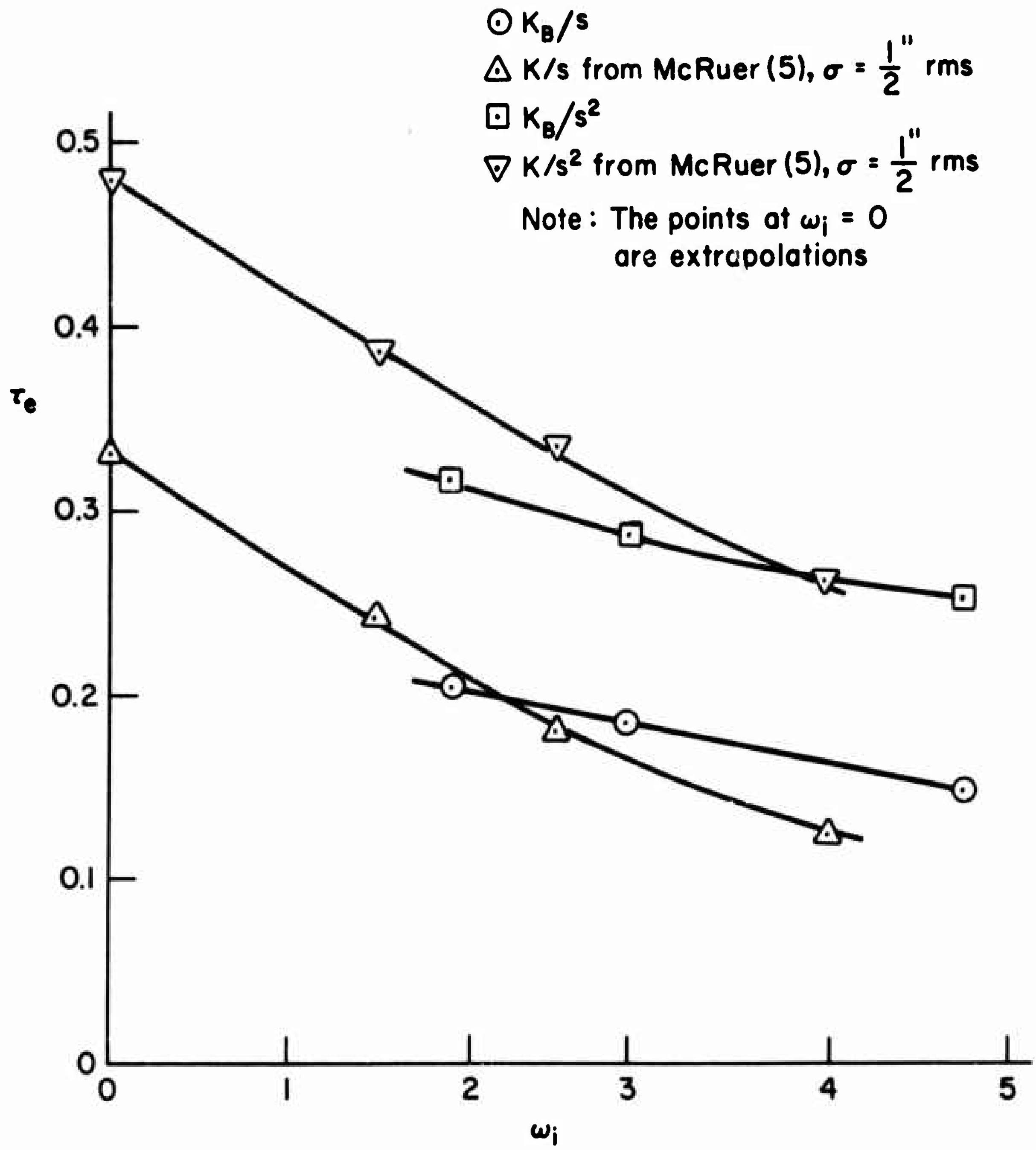


Figure 16. Variation of Effective Time Delay with Forcing Function Bandwidth (σ = 1 cm rms)

shows the effects of ω_i on τ_e at the intermediate input amplitude of 1 cm rms. Plotted also are the comparable curves from McRuer (5). It is seen that the trends are consistent, i.e., that although the slopes are different, the values of τ_e are very nearly the same. The difference in slopes could be accounted for by the differences in the control axis and stick (pitch/center-stick versus roll/side-stick), but the differences are considered small enough to show that the τ_e trend is compatible and consistent.

Further checks are provided by crossover frequency and phase margin trends with input bandwidth. Figure 17 shows a comparison of crossover frequency trends. The agreement with Ref. 5 is good. The regression phenomenon can be seen for the acceleration command element, i.e., the operator actually reduces the error magnitude by reducing his gain and bandwidth slightly when the input bandwidth is large. Figure 18 shows a comparison of phase margins. The agreement is excellent for K/s^2, but rather poor for the K/s elements. Since the high-frequency phase is so sensitive to the τ_e curve fit, the comparison does not indicate countertrend and is thus considered inconclusive.

An interesting alternative way to look at the τ_e data exists which should prove useful in the estimation of τ_e. McRuer (5) shows a dependence of τ_e on the form of the controlled element as well as the input bandwidth. The τ_e seemed to depend on the equalization generated internally by the pilot. A concise method of portraying the equalization can be obtained by defining a parameter which is sensitive to both lead and lag. One such parameter is the slope of the pilot's amplitude ratio at the crossover frequency, where the choice of crossover frequency reflects that the pilot is most sensitive to characteristics at crossover during tracking [see, for example, McDonnell (27) or Ashkenas (21)]. Data were assembled from this study and from McRuer (5) to test such a parameter. Figure 19 shows the effective time delay for several controlled elements plotted as a function of L, where

$$L = \frac{1}{20}\left(\text{Slope of } |Y_p|_{\text{dB/decade}}\right)_{\omega=\omega_c} \qquad (30)$$

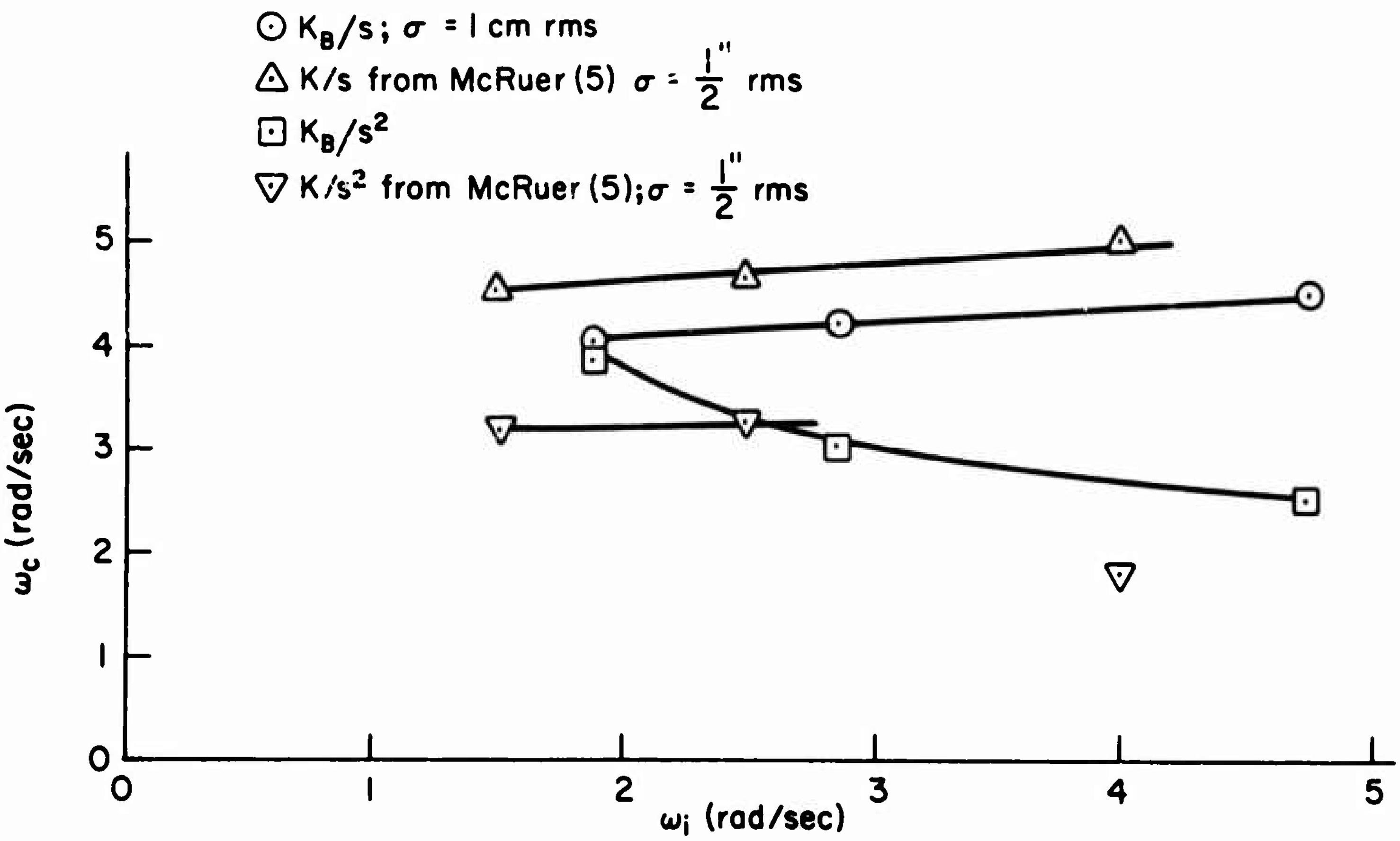


Figure 17. Variation of Crossover Frequency with Input Bandwidth

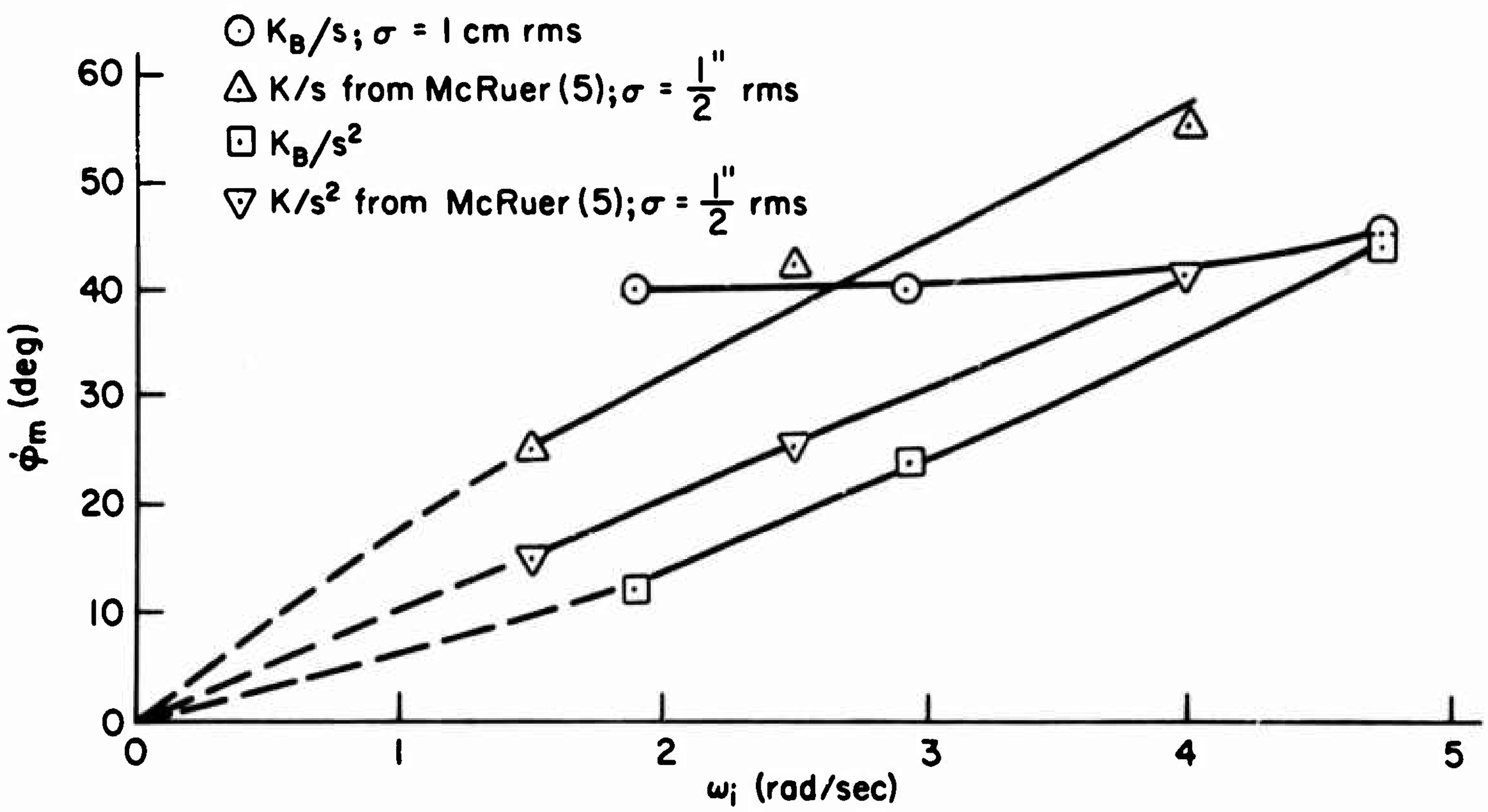


Figure 18. Variation of Phase Margin with Input Bandwidth

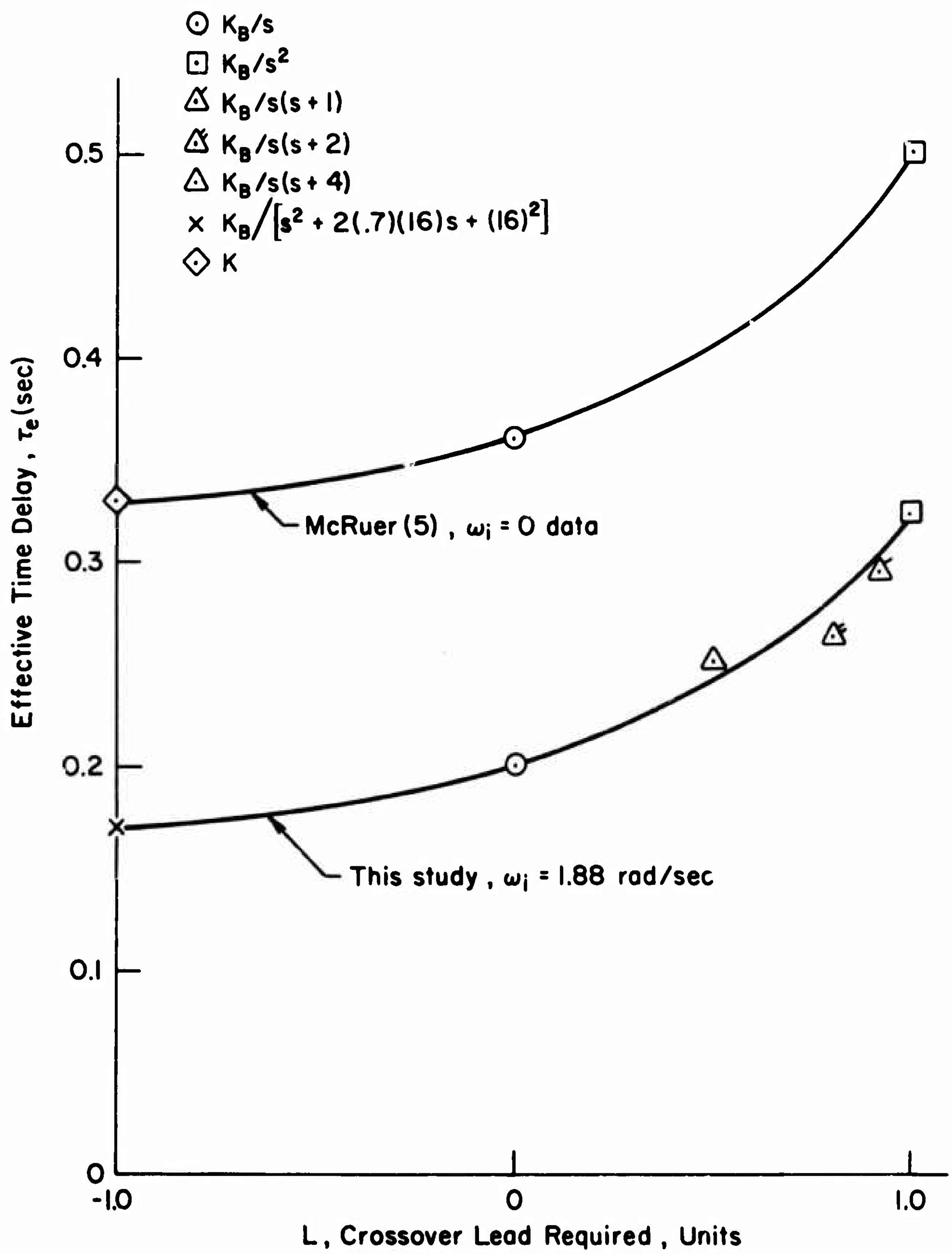


Figure 19. Variation of Effective Time Delay with Slope of Pilot Equalization

Thus, for a $Y_c = K/s$, no lead would be required and $L = 0$. For $Y_c = K/s^2$, lead is generated at a very low frequency, so $L = 1$. It is apparent that a remarkable pair of curves results which is a function of ω_i, the input bandwidth. Such a family has considerable potential as an aid to estimate τ_e. The limited amount of data shown in Fig. 19 also further demonstrates compatibility with past work.

3. The Relationship Between the "Best" Gains

It was hypothesized by McDonnell (27) that, for a given tracking task, the selection of the "best" gain for the controlled element is based on the amplitude ratio of the element at crossover, i.e., where $|Y_p Y_c| = 1$. Thus, if the gain at crossover for one form of controlled element is known, the gains for other forms should be estimable by setting crossover amplitudes equal. Since the best gains were determined by pilot JDM for several forms, the hypothesis can be checked with the data of this study. Table XVII lists the data necessary to compute crossover gains together with the computed values. If the gain of the subcritical task is excluded

TABLE XVII

AMPLITUDE OF THE CONTROLLED ELEMENT AT CROSSOVER

Y_c/K_B	ω_c(rad/sec)	K_B	$\lvert Y_c(\omega_c)\rvert_{dB}$
$1/s$	4.0	0.586	−16.7
$1/s(s+4)$	4.0	2.15	−20.5
$1/s(s+2)$	4.0	2.15	−18.4
$1/s(s+1)$	3.4	2.15	−15.0
$1/s^2$	4.0	1.17	−22.9
$1/(s-2)$	4.7	3.45	−3.4
$1/[s^2+2\times0.7\times7.8s+7.8^2]$	4.5	8.38	−13.2
$1/[s^2+2\times0.7\times16s+16^2]$	3.1	35.2	−16.9

(it obviously is not comparable with the others), the mean of the gains, in dB, is −18 dB, with extremes of ±4.8 dB. This is regarded as reasonable support for the hypothesis since the gains which could be selected were in discrete steps allowing an uncertainty of approximately ±50 percent in the final setting. At the very worst, close estimates to the best gain can be made if the crossover model is valid. Connections between the hypothesis and the subcritical task gain are not known at this time.

C. CORRELATIONS OF PILOT RATING WITH THE EXPERIMENTALLY DETERMINED PARAMETERS

Approximately 50 **compensatory pitch tracking** runs were made by each of the two pilots. Since eight different rating scales were used by the pilot for each run (see Table XVI), and approximately a dozen parameters were measured during a run, a selective correlation will have to be made for reasons of economy. Because of the wide familiarity with the Cooper rating, it will be used to make the initial correlations with system and pilot parameters. Correlations between ratings can then be made to test the selectivity and sensitivity of the individual trait ratings. Any special trends which look promising can then be brought out explicitly by returning to a correlation of the individual rating scale with the system parameter of interest. The number of cross plots can thus be kept to a minimum without running the risk of missing key trends.

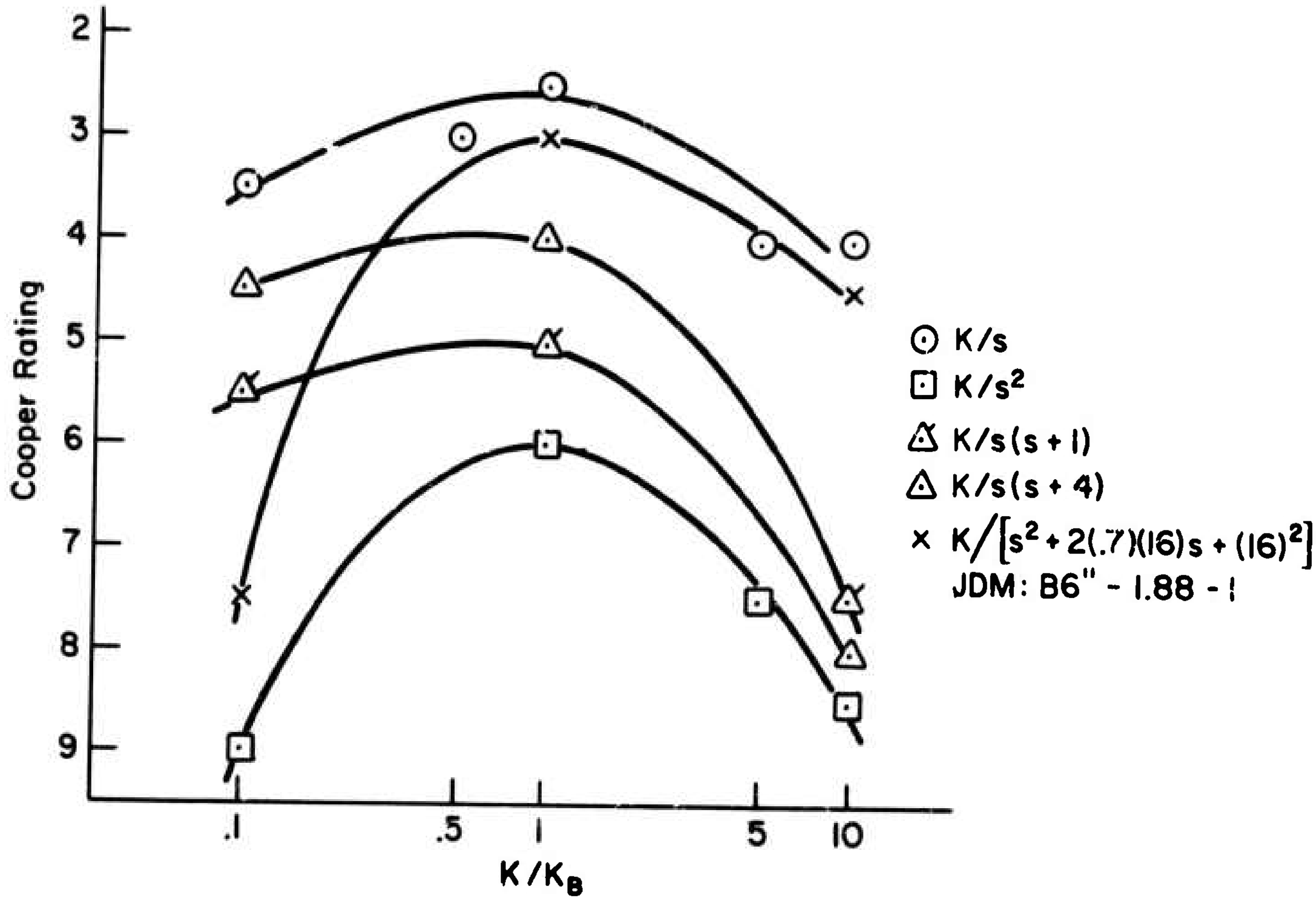
Some of the data presented will be redundant because of the functional dependence of several parameters. Thus, for example, plots of pilot gain and controlled element gain versus ratings would be identical because the adaptive nature of the operator results in $K_pK_c = \omega_c$ = constant. Nevertheless, since we are looking for consistency and the widest applicability possible, all pertinent parameters will be considered.

1. Correlation of Pilot Rating with Pilot Parameters

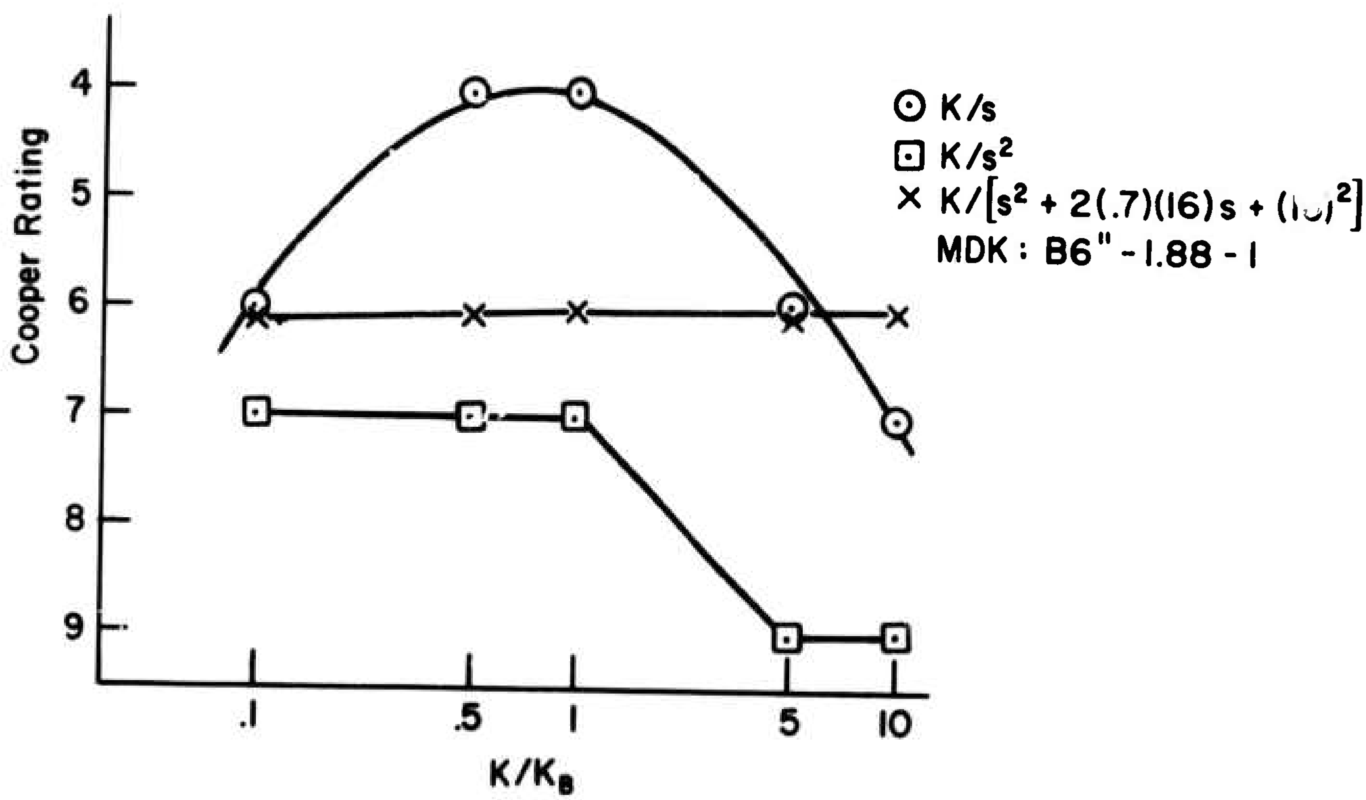
a. <u>Variation of Pilot Rating with Gain</u>. The operator is capable of adapting over a very large dynamic gain range with little change in performance, so the pilot's opinion of various gains is of extreme importance. Figure 20 shows the results of a dynamic range of 100. A preliminary set of trials was carried out to determine K_B. The gain was then varied between $0.1K_B$ and $10K_B$. Since $K_pK_c = \omega_c$ = constant for a given Y_c form, either K_c or K_p can be plotted to show the desired trends. The selected parameter for Fig. 20 was the ratio of the controlled element gain to the previously determined "best" controlled element gain. The resulting trends in Fig. 20a show the expected dome shapes. It is interesting to note that there appears to be a "family compatibility" with all but the second-order complex pair element results. The opinion trend for all elements seems to deteriorate more quickly for the high controlled element gains. The comment was made during the series that the results of an inadvertent stick motion with the high element gain was considerably more disagreeable for all controlled elements than the large stick displacements (and forces) necessary with the low element gains. On the other hand, when a low element gain was used with the complex pair, extremely large stick forces had to be held, whereas with the other forms the large input excursions could be integrated out. The rapid deterioration of opinion for the complex pair element is therefore quite reasonable.

Figure 20b shows the Cooper ratings obtained from the other pilot, MDK for gain variations. As mentioned earlier, describing functions could not be computed for MDK because of the limited funds available. However, the rating data shown give us a hint as to the kind of problems introduced by pilot "set." MDK was obviously a much less sensitive rater, i.e., he was reluctant to make fine distinctions between configurations. His comments indicated that he preferred to base his ratings on the category descriptors as much as possible because the finer distinctions were not clear. No further implications can at present be drawn from these data, but they are included so that a data base will be started for future studies of pilot set. Additional MDK data are included on "trait" ratings in a subsequent subsection.

b. **Variation of Ratings with Effective Time Delay.** Pilot parameters, including the effective time delay, were read from curve fits of the describing function data. A tabulation of the parameters, as well as the curve fits themselves, are shown in Appendix C. Figure 21 shows the



(a) JDM



(b) MDK

Figure 20. The Variation of Pilot Rating with Controlled Element Gain for Two Pilots

effects on opinion of the effective time delay. The gains for all elements were optimum. The variation looks quite linear with the exception of the subcritical task, which contains a nonminimum phase pole. Recall that it is the subcritical task which cannot be fitted with the simple crossover model with any great success, and which has a constraining effect on the pilot.

The effective time delay is affected by two factors. The effect of the equalization generated by the pilot on the τ_e is shown by the solid line in Fig. 21. Input bandwidth effects are shown by the dashed line. The carpet plot of Fig. 21 sums up the relation between both equalization required and input bandwidth quite neatly, so that if used in conjunction with Fig. 19, estimates of ratings should be improved.

c. **Variation with Equalization**. Prior to the experiments of this study, very little data existed where lead was measured at the same time the ratings were taken. Thus the majority of the connections between lead and ratings were inferred [see, for example, Ashkenas (21)]. A compounding problem was the uncertainty about the lead placement. It has been assumed in most recent work that the pilot exactly canceled controlled element lag with his lead generation over an approximate range of $0.1 < T_L < 5$ sec. Thus lead equalization relationships with pilot ratings have been abundant, but also questionable.

The data points of Fig. 22 overcome the two shortcomings noted above. Best gains were used on all configurations, and the bandwidth and amplitude of the input was held fixed. Scrutiny of the describing function data in Appendix C will reveal that $|Y_pY_c|$ does indeed look like K/s over the frequencies where the lag occurs, indicating that the pilot does cancel the lag with his lead. It was necessary to infer only one lead value — that for $Y_c = K/s^2$. It has been shown in McRuer (5) that in that case $T_L \doteq 5$ sec, which is below the lowest frequency that can be resolved with one or two runs.

A comparison of the rating data with previous data [for example, Ashkenas (21)] shows that the difference in ratings between K/s^2 and

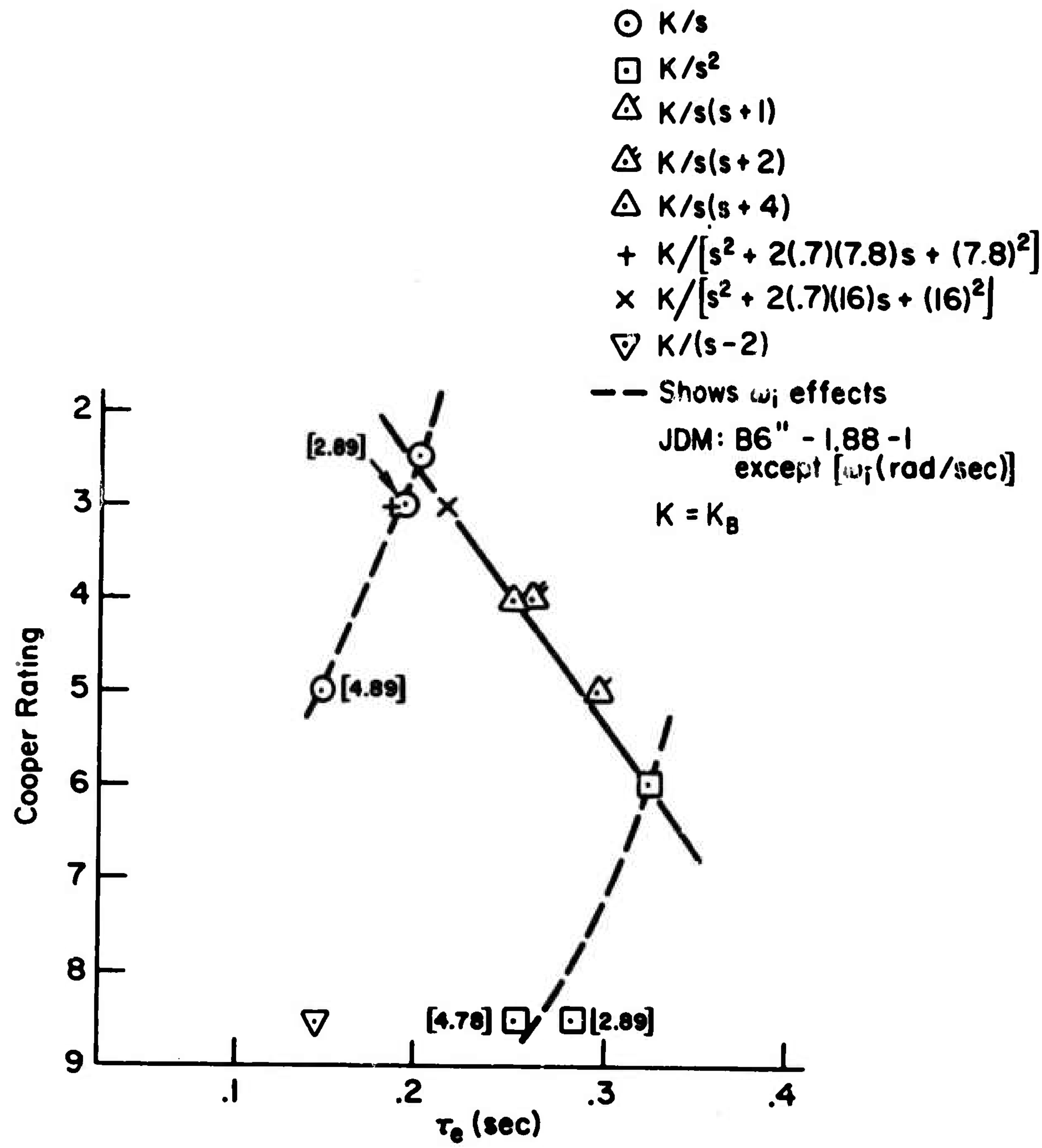


Figure 21. Variation of Pilot Rating with Effective Time Delay

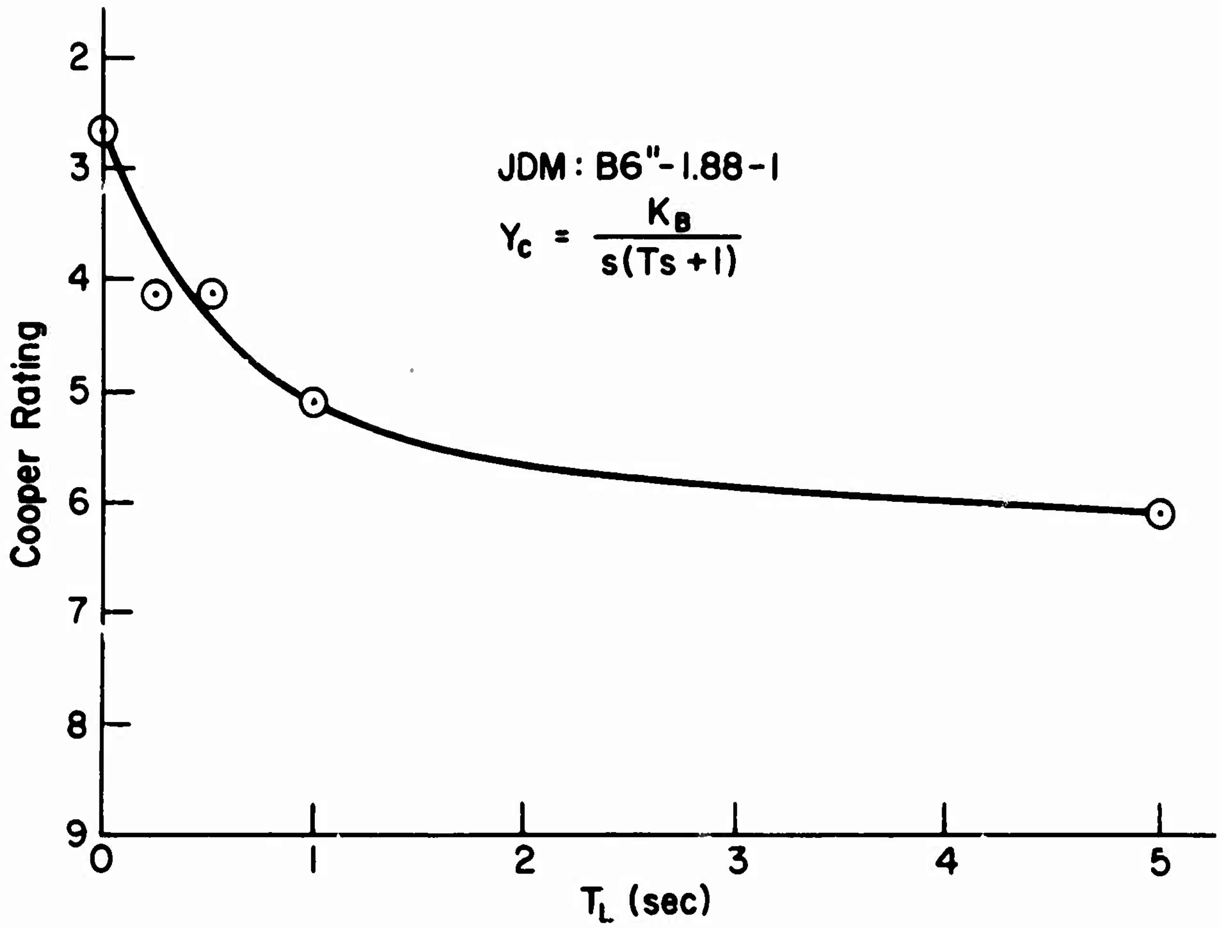


Figure 22. Variation of Pilot Rating with Pilot Lead

$K/s(s+1)$ is not as large in the current series as would be expected. A difference of only one Cooper unit was obtained here as compared with 3-4 units elsewhere. This compatibility problem cannot be resolved because of the already mentioned uncertainties in the older data, together with a lack of documentation regarding task, control stick, motivation, simulator quality, etc. For example, opinion is thought to be very sensitive to crossover frequency when the lead is near crossover. Thus, in the current series, a wider difference in ratings would probably have resulted if the pilot had lowered his gain slightly.

2. Correlation of Ratings with Closed-Loop Parameters

There are myriad closed-loop parameters which could be computed, but perhaps three are of significance in identifying trends. A measure of the "tightness" of the loop closure is provided by the crossover frequency, and we have previously maintained that it remains essentially invariant with gain changes (see p. 73). It would therefore be instructive to check it. Stability margins and performance are also of interest. Since phase margin is used almost universally, it is appropriate to use it here. Finally, performance could conceivably influence pilot ratings to a high degree, hence an error measure will be computed and checked.

a. **Crossover Frequency Trends.** It was shown in Fig. 17 that the crossover frequency, ω_c, is essentially invariant with input bandwidth, ω_i. Checks of ω_c as a function of gain are also available for K/s and K/s^2. Shown in Fig. 23 are the crossover frequencies for several gains (the 0.1 K_B/s^2 describing function calculations had an extremely poor signal-to-noise ratio, hence ω_c was not available for it). The change in ω_c due to gain is seen to be about 1 rad/sec over a dynamic range of 100. With such a small variation, it is a foregone conclusion that a correlation between ratings and ω_c would be poor.

b. **Correlation of Phase Margin and Ratings.** The phase margins for the best gain configurations are plotted in Fig. 24. With the exception of the subcritical task, the ratings vary fairly linearly with phase margin. It could be argued that the pilot is downgrading the configurations because of his increasing discomfort with the lowering stability margins. It could also be argued that the "cause" is the requirement to equalize. Since pilot comments were of no help, it is pointless to speculate about cause and effect. However, the phase margin can be written as

$$\varphi_m = \frac{\pi}{2} - \tau_e\omega_c \qquad (31)$$

We have seen that crossover frequency is approximately constant as a function of gain, and that a small incremental difference exists between forms, so we would expect that φ_m will vary inversely as τ_e. A comparison of Figs. 21 and 24 shows that to be the case.

c. **Performance and Ratings.** The pilots were instructed to rate the configurations in the context of the task, where the task specification included a performance error specification. The resulting objective measures of performance should thus be interesting to compare with the

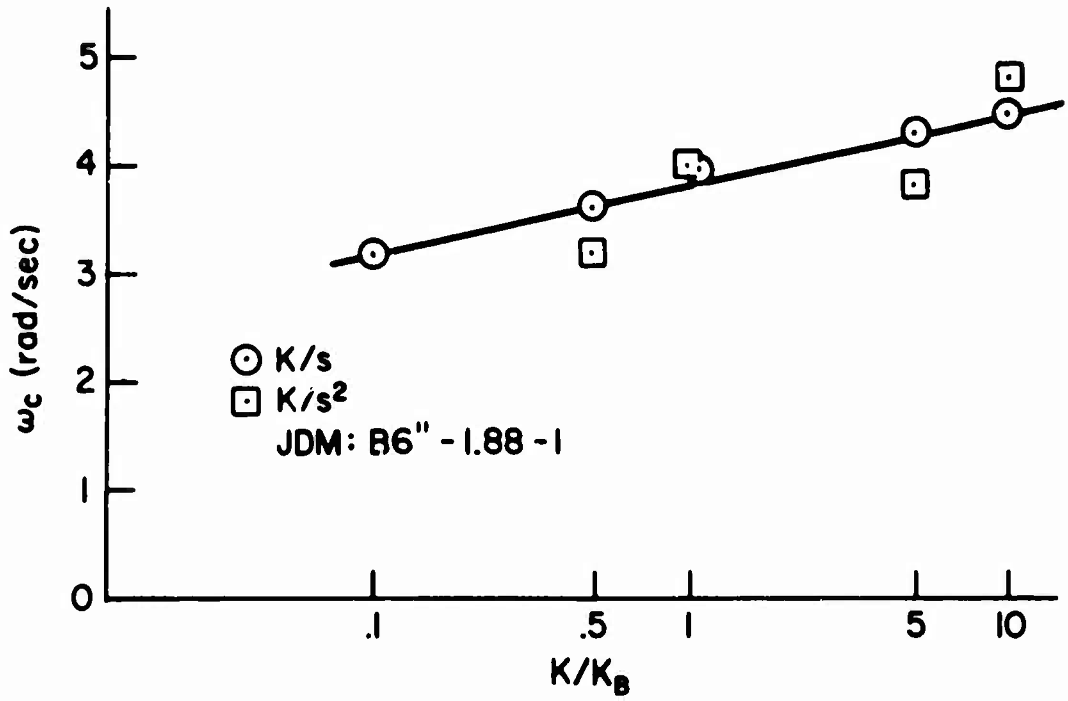


Figure 23. Variation of Crossover Frequency with Gain

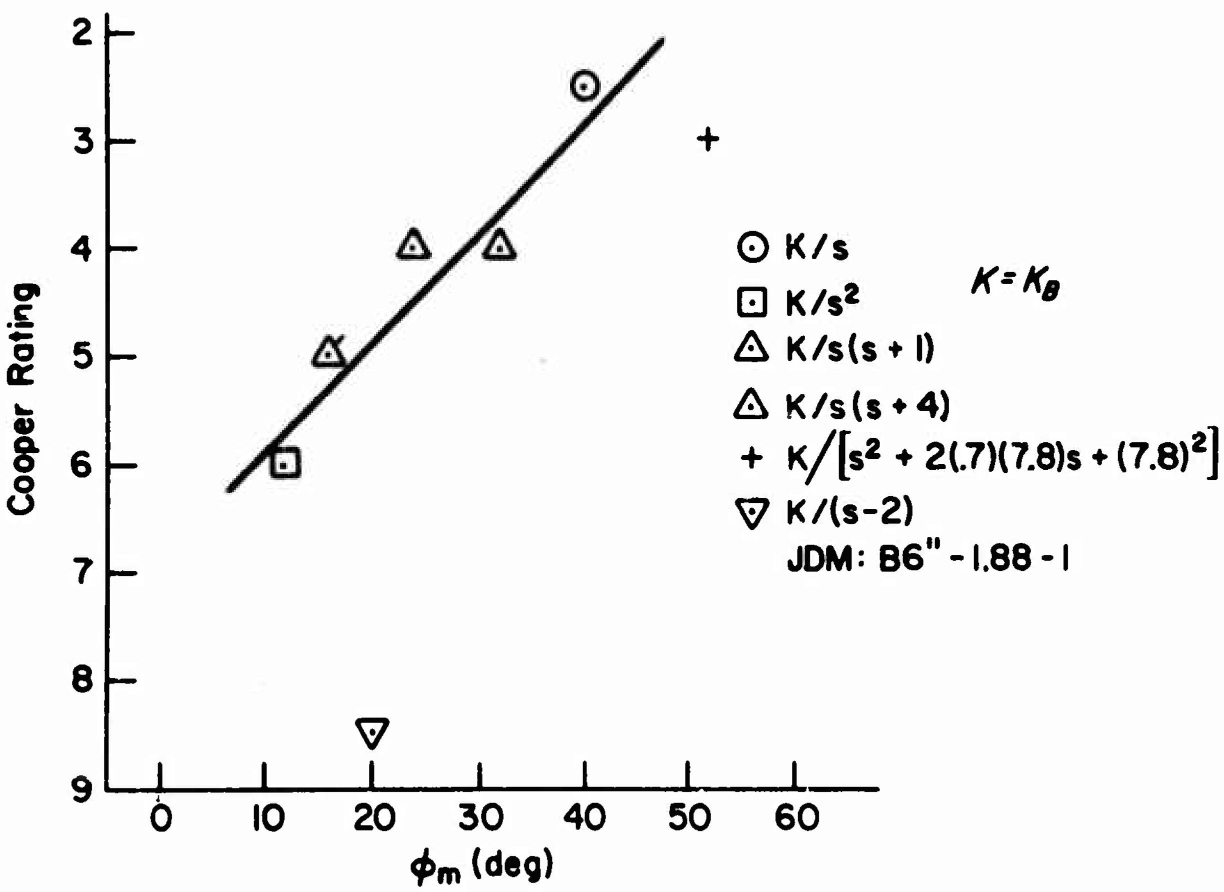


Figure 24. Variation of Cooper Ratirg with Phase Margin

ratings. Performance was computed by measuring the average absolute value of the system error, i.e.,

$$\overline{|e|} = 0.01 \int_0^{100} |\theta_e| dt$$

Figure 25 presents further evidence that the crossover characteristics stay approximately constant as a function of gain. Performance, then, gives no indication of the rating changes due to gain for a given form.

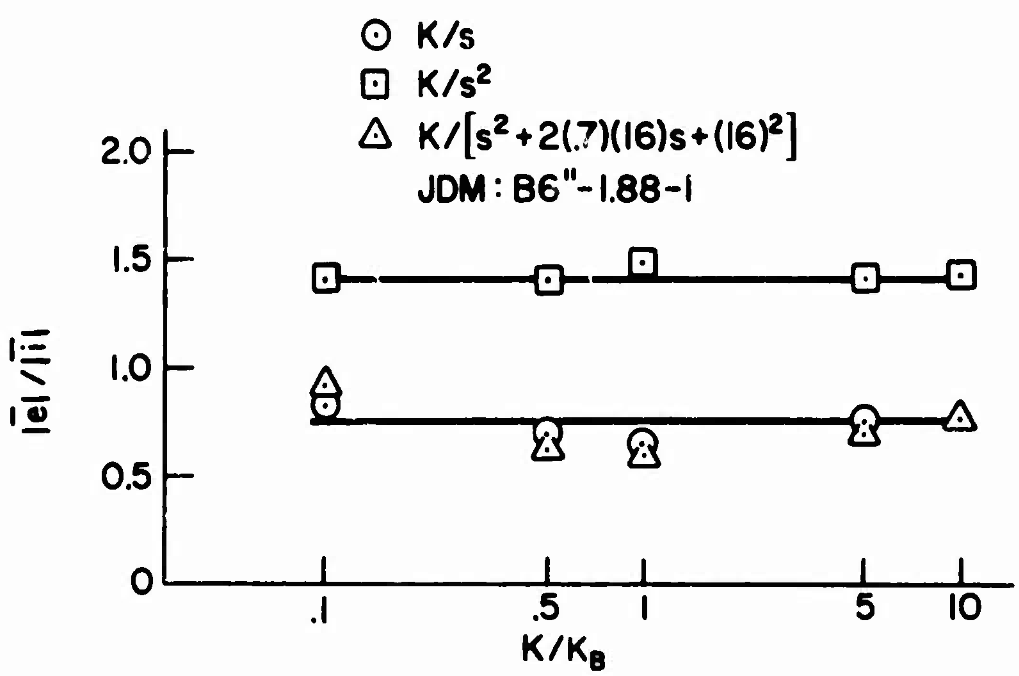


Figure 25. Performance Variation with Gain

On the other hand, Fig. 26 shows that there is a direct correlation of performance and ratings between the "best" gain configurations of several controlled element forms. Shown in the figure are four data points for input bandwidths other than 1.88 rad/sec. The correlation for the low ratings is seen to be quite good.

If the pilot is really rating partly on performance, a look at the actual magnitude of the error could prove interesting. Figure 27 shows the absolute value of the system error averaged over the 100 sec run length for K/s and K/s^2, and with the three input levels. The correlation

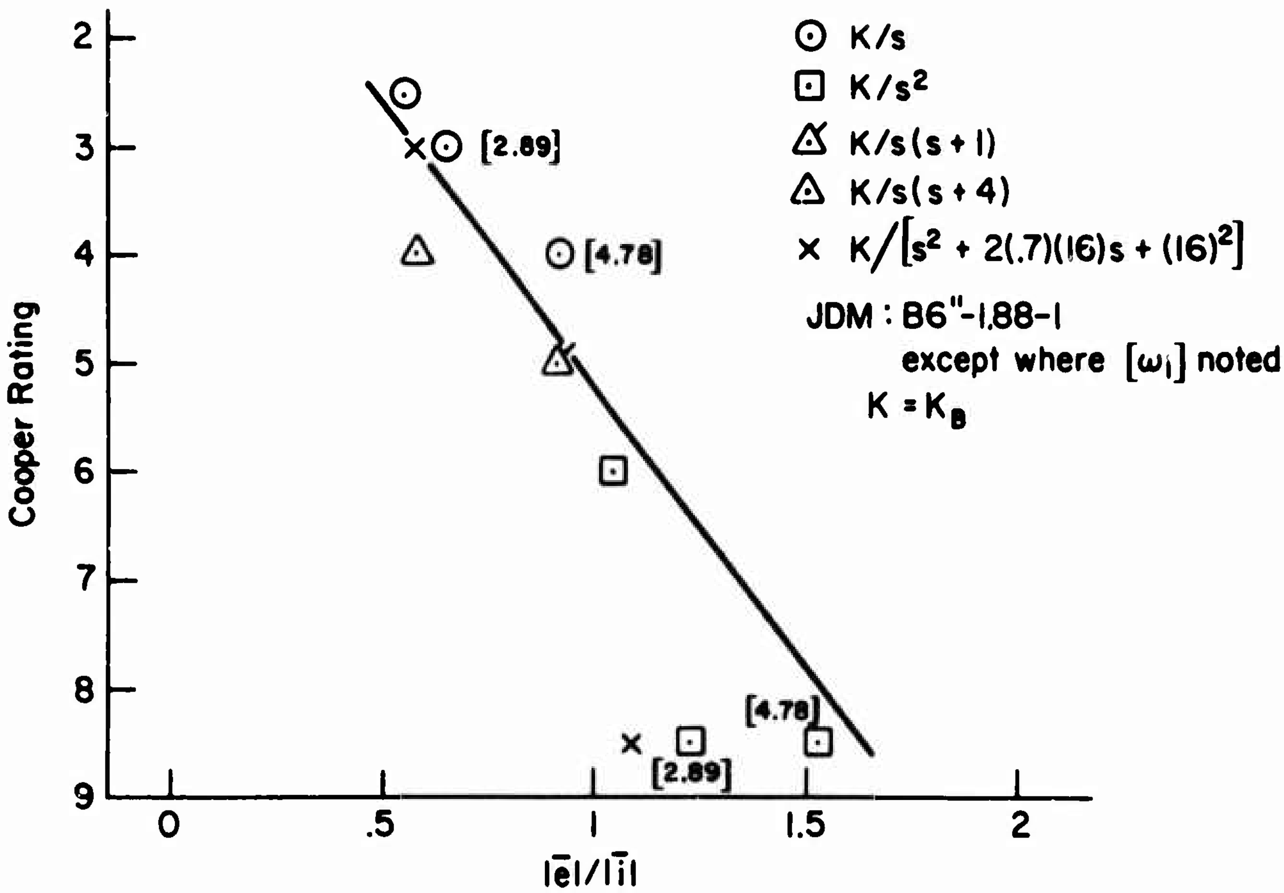


Figure 26. Correlation of Rating with Performance

is excellent. The regression line is identical to the line in Fig. 26, but for the sake of clarity the two figures have been kept separate. Shown also in Fig. 27 is the performance criterion value specified in the task definition. For this particular pilot, the intersection seems to be at about the three level on the Cooper scale.

d. **Connections Between Remnant and Ratings.** The pilot's stick output power can be considered to be the sum of the power which is correlated with the system input (the linear portion) and the uncorrelated power, or noise, which is by definition the remnant. The relative remnant, ρ_{ac}^2, is defined as the ratio of the correlated power to the total power, or

$$\rho_{ac}^2 = \frac{\overline{c_i^2}}{\overline{c^2}} = \frac{\overline{c_i^2}}{\overline{n^2} + \overline{c_i^2}} \qquad (32)$$

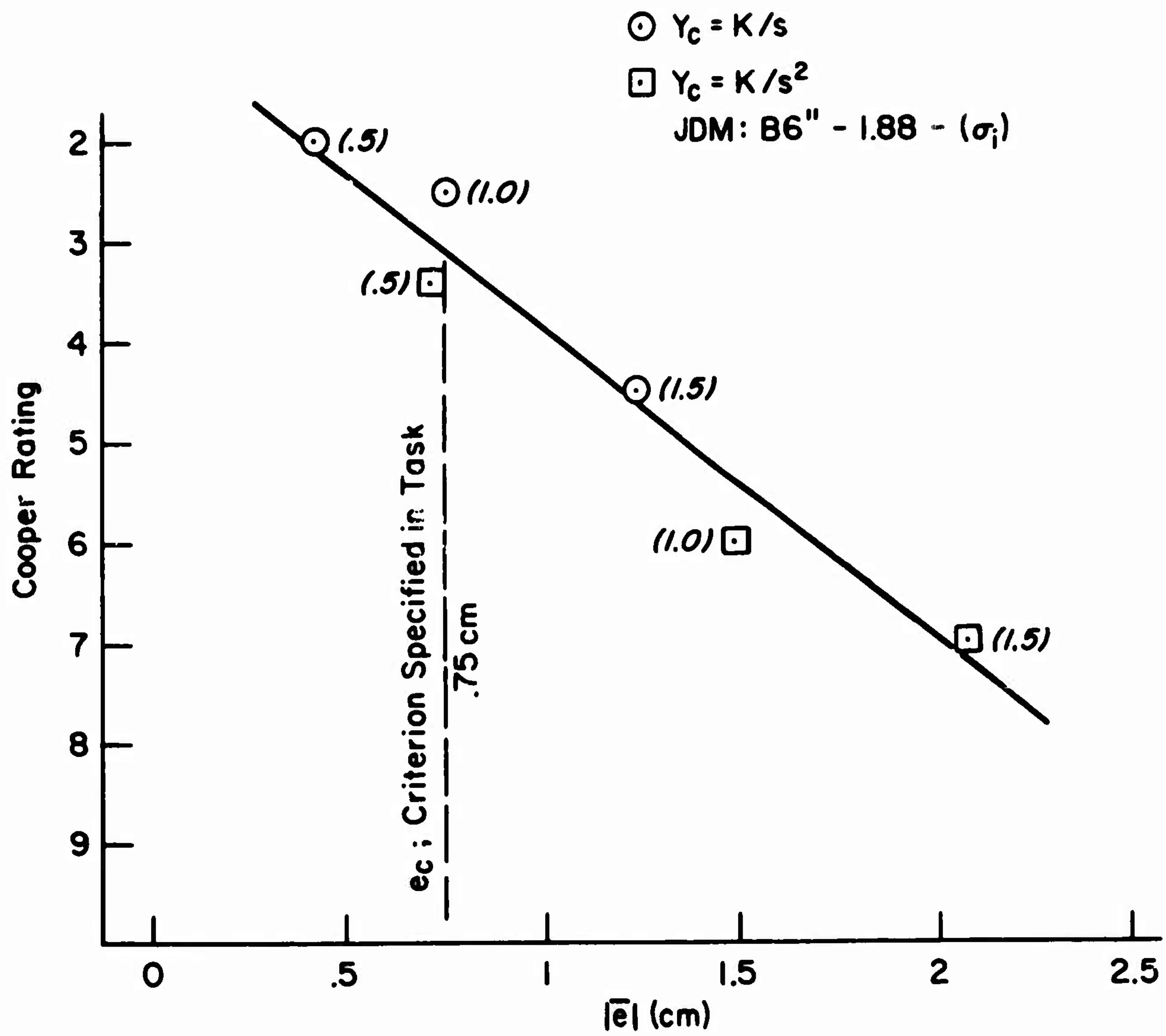


Figure 27. Influence of Error Magnitude on Ratings

Thus, when the operator is introducing only a small amount of noise, either through nonlinearities, time variations, or noise injection, the ρ_{ac}^2 will be nearly unity. When the operator's output is all noise, the ρ_{ac}^2 will be zero. Since the amount of remnant in the system could have a significant effect on pilot ratings, the relative remnant was computed simultaneously with the describing functions.

The variation of the relative remnant was investigated as a function of four key parameters: the controlled element gain; the effective time delay which, it will be recalled, reflects the equalization required of the pilot; the amplitude of the system input; and the bandwidth of the system input. Figures 28a and b show the effects of controlled element gain on the remnant, and the correlation of rating with the remnant. The trend of ρ_{ac}^2 with gain demonstrates that the pilot performs more linearly with larger stick excursions when the element is a K/s, but that his performance with a K/s^2 is approximately one-half noise and is little affected by gain. The corresponding rating results, Fig. 28b, show little correlation with remnant, indicating that the remnant variation with gain is probably not a primary causal factor of the rating variations.

The relation of remnant and τ_e is the most interesting of the quartet. The configurations all have best gains and the same input, so only the form differences are influencing the remnant. The straight line shown in Fig. 29 fits the data reasonably well, with the exception of the subcritical task point. It will be remembered that this is the case which is not adequately described by the crossover model. Thus it could be argued that a measure of task difficulty, at least for a comparison of different forms, is given by the relative remnant. It is felt, however, that τ_e is considerably more direct and can be estimated, so is the more desirable of the two measures to apply to the rating problem.

The effects of the input are shown in Figs. 30 and 31. It is apparent that no direct or significant correlations exist, which leads to the conclusion that it is effects of the input on other parameters (namely, the $\Delta\tau_e$ and performance, as we have seen) that causes the deterioration in rating.

The remnant data presented in Figs. 28 through 31 are consistent with McRuer's (5) data, with the possible exception of the variation with gain for K/s^2. McRuer found a definite decrease in ρ_{ac}^2 with increasing gain, while this study notes a slight increase in ρ_{ac}^2. The data has been carefully checked, so the discrepancy must remain unexplained.

3. Correlation of Ratings with the Environment

It has been emphasized several times to this point that the configuration must be rated in the context of the task in order for the ratings to be valid indicators of the vehicle suitability for the task. We would thus expect ratings to be dependent on the environment, or system input

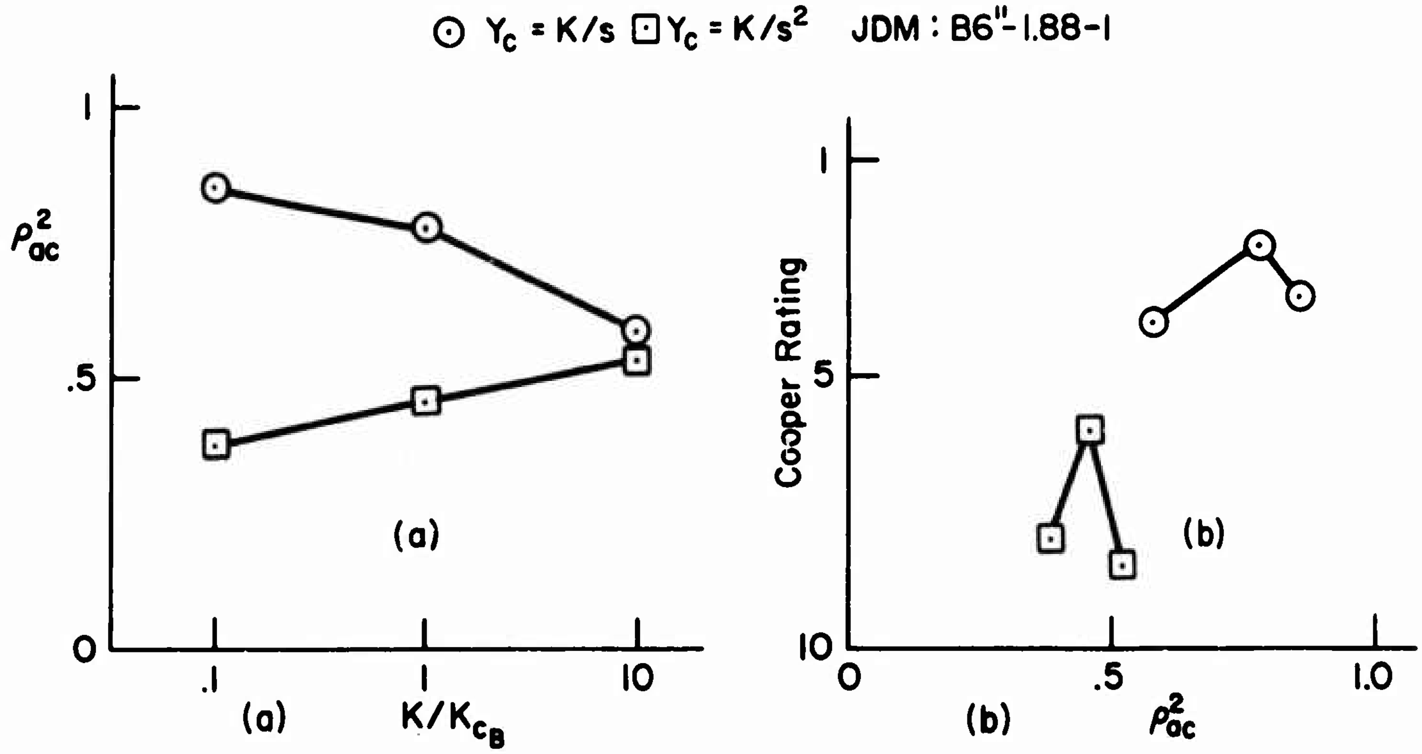


Figure 28. Variation of Ratings with Remnant as a Function of Gain

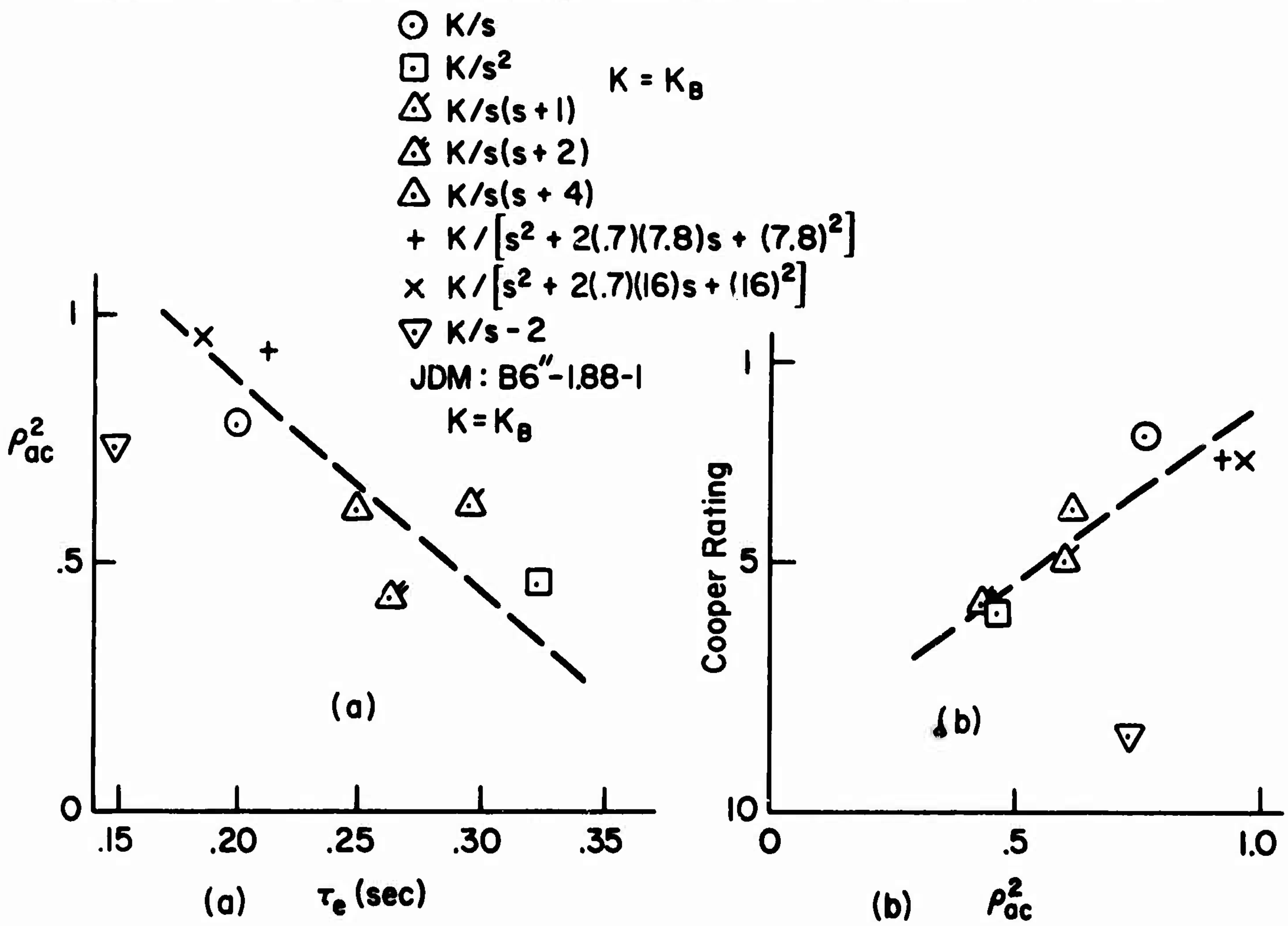


Figure 29. Variation of Ratings with Remnant as a Function of Controlled Element Form

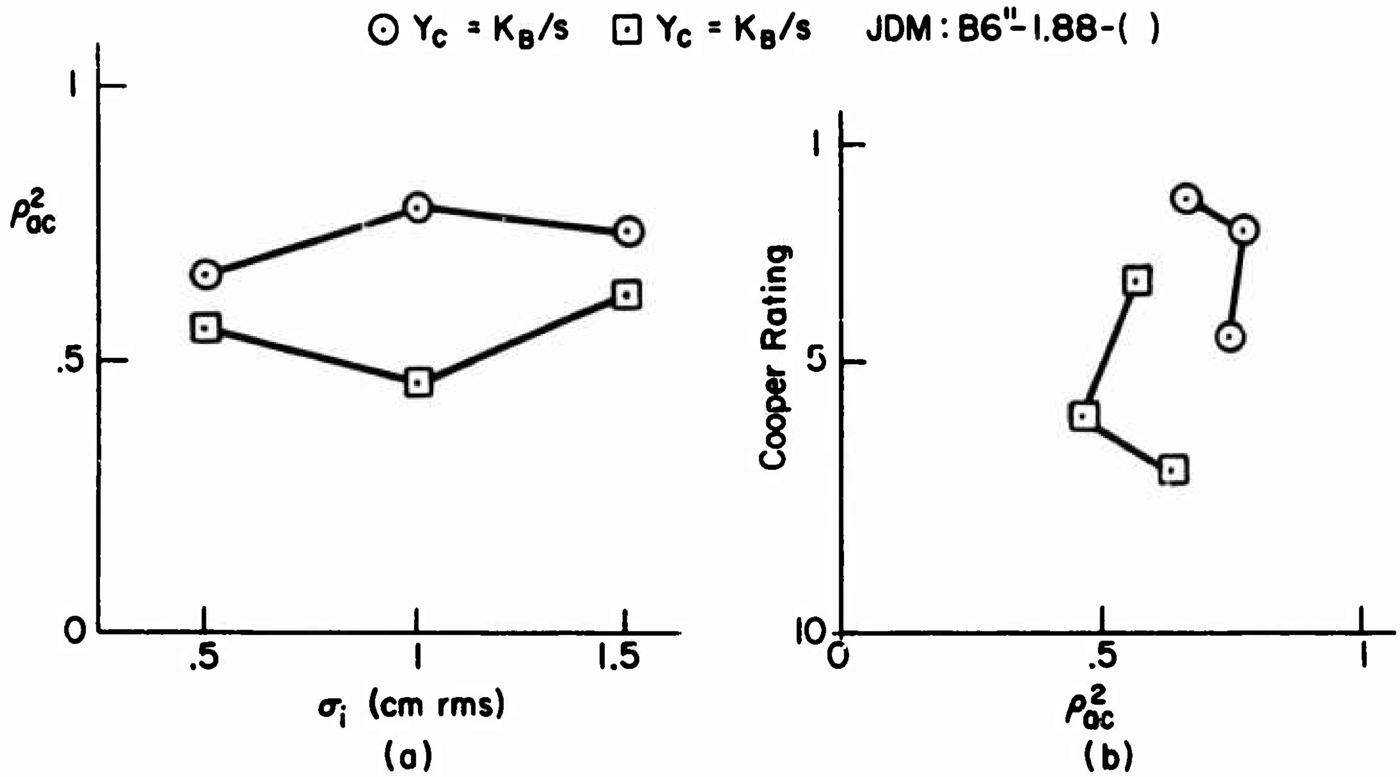


Figure 30. Variation of Rating with Remnant as a Function of Input Amplitude

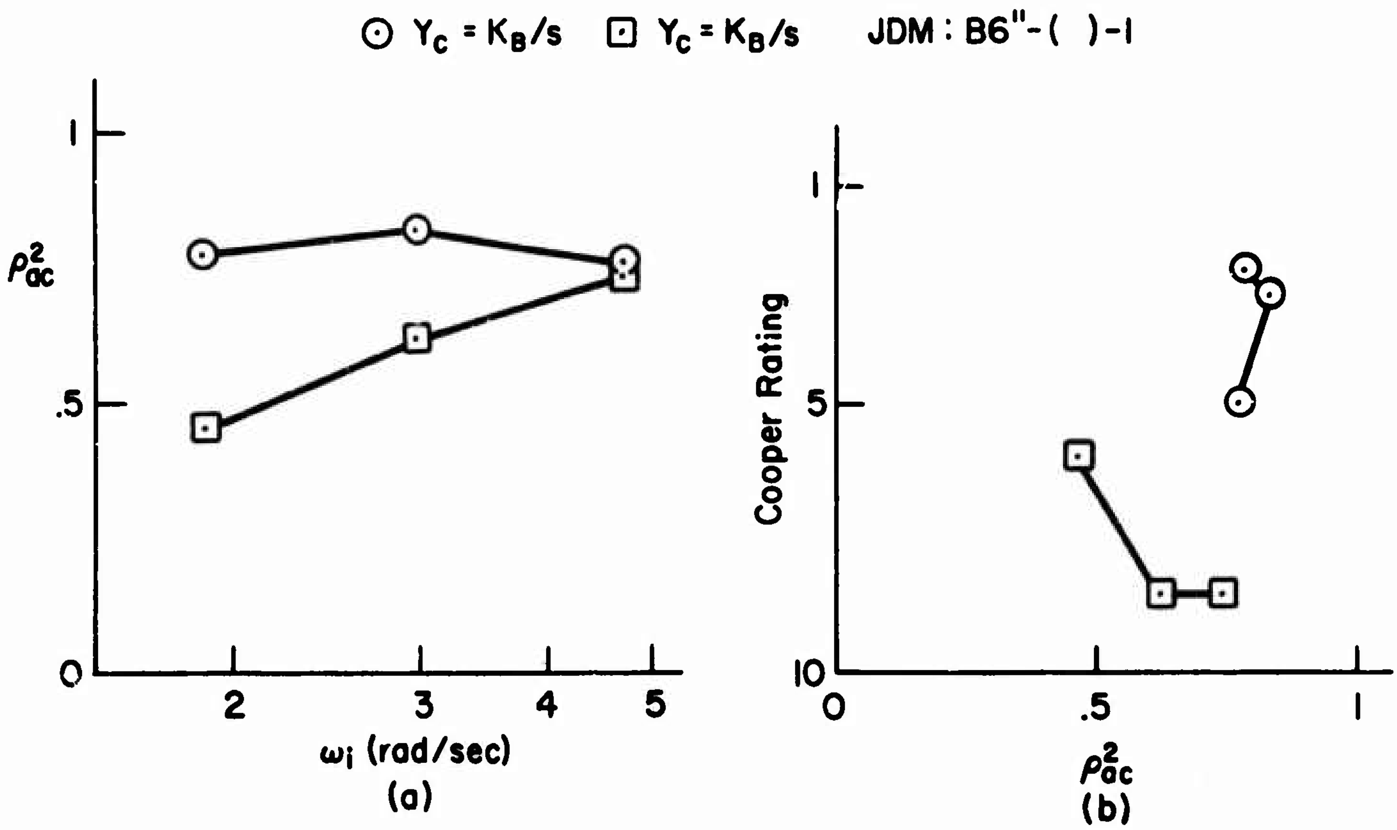


Figure 31. Variation of Ratings with Remnant as a Function of Input Bandwidth

in our case, as well as the configuration and task specification. Results supporting this contention have already been noted, where we have seen changes in rating as a function of τ_e, for example, which can be, in turn, almost totally dependent on the input bandwidth (see the dashed lines in Fig. 23). Here we shall plot the input effects directly, which is just an alternate way of looking at the data.

The data shows, in Fig. 32, that for small amplitude inputs the bandwidth must be increased to fairly large values before the pilot is appreciably affected. As the amplitude is increased, however, the pilot becomes very sensitive to bandwidth. This phenomenon could be a manifestation of the indifference threshold discussed in McRuer (5). When a good deal of lead is being generated, as with the K/s^2, an increase of ω_i from 1.88 to 2.89 rad/sec caused an increment in ratings of 2.5 to 3 units.

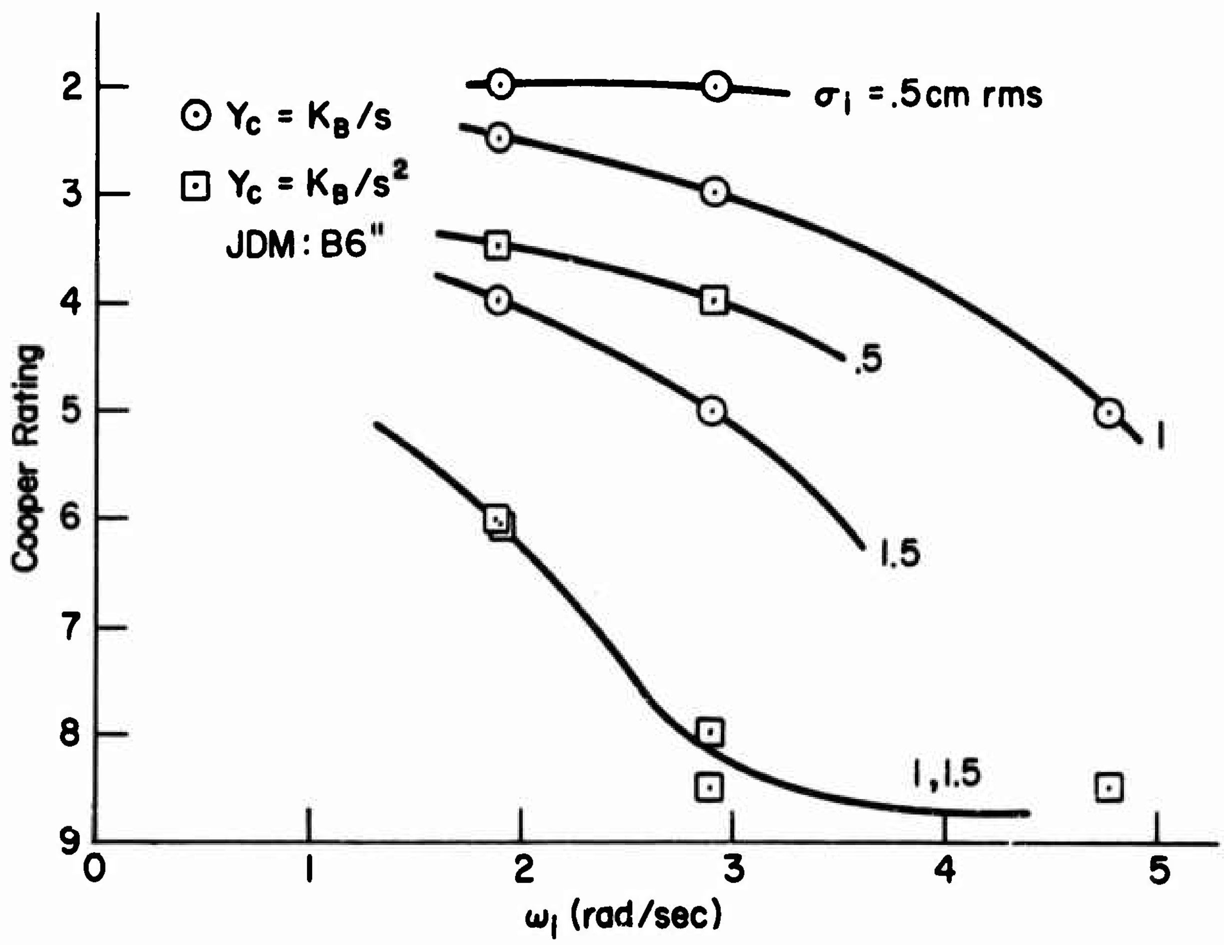


Figure 32. The Effect of Input Characteristics on Ratings

4. Correlation of Ratings with Secondary Task Score

As detailed in Section IV, a secondary loading task, in the form of an unstable roll tracking task, was utilized as a measure of pilot attention required to maintain primary task performance, or the "excess capacity" the pilot has for performing other tasks while maintaining primary performance. The scores obtained from the cross-coupled secondary task represent its degree of difficulty; consequently, they also represent the "degree of ease" of the primary task.

Secondary scores were obtained for all configurations and inputs, and have been correlated with ratings in various ways. Figure 33 shows how

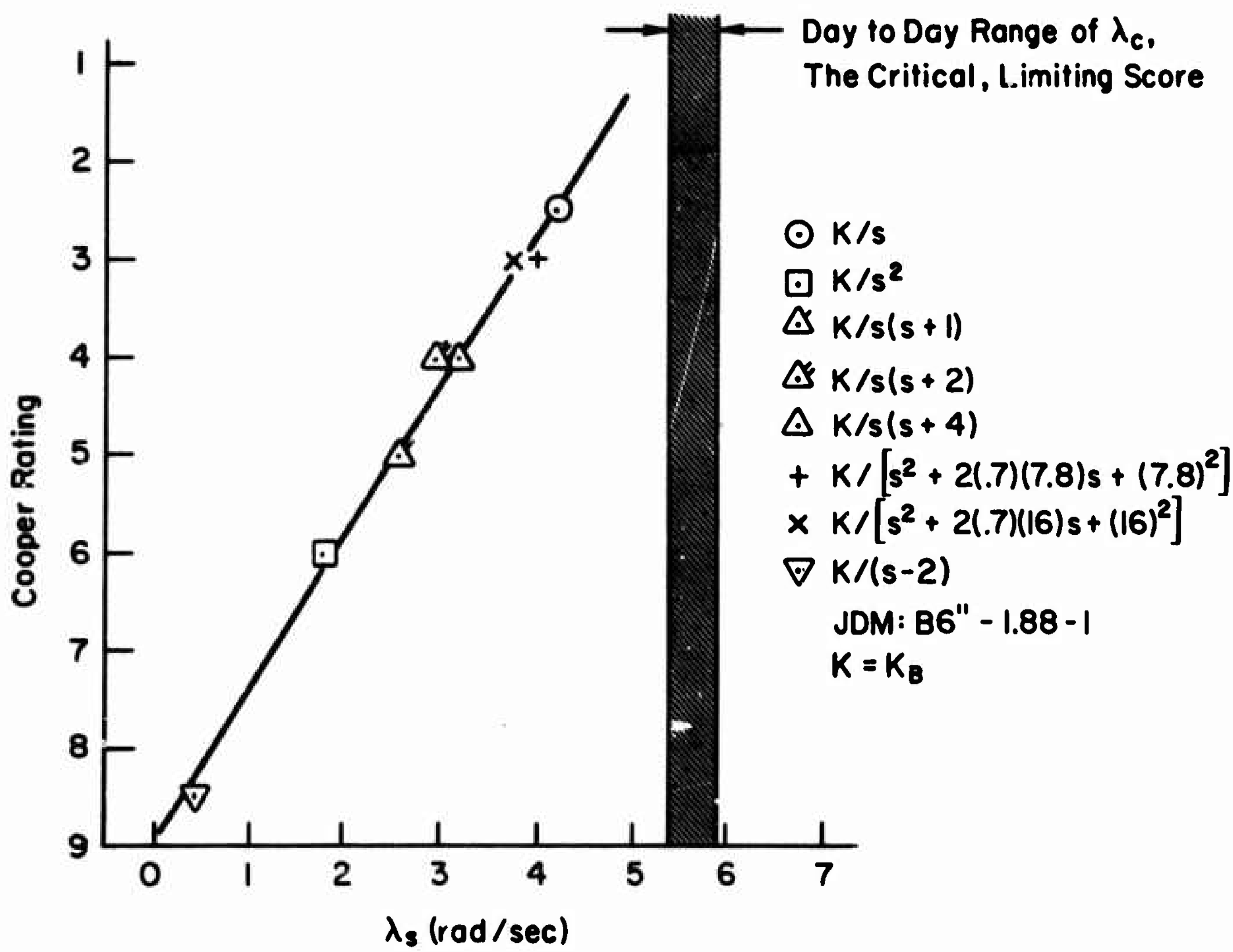


Figure 33. Secondary Task Score Variation with Ratings for Best-Gain Configurations

the scores for the best gain configurations of each form compare with the Cooper ratings. The agreement is extremely good. Even the subcritical task, which has been a notable culprit in other correlations, seems to fit in linearly with the other data. Recall that a $\lambda_s = 0$ corresponds to 100 percent of the pilot's attention being focused on the primary task, while a $\lambda_s \doteq 5.5$ means that no attention is required to maintain primary performance.

The effects of gain variation are shown in Fig. 34. Here again, the correlation is remarkable. The data point for $Y_c = 0.5\ K_B/s^2$ is considered either to have been rated incorrectly or set up incorrectly on the computer, since the rating assigned was considerably better than the "best" rating, i.e., the rating for K_B/s^2.

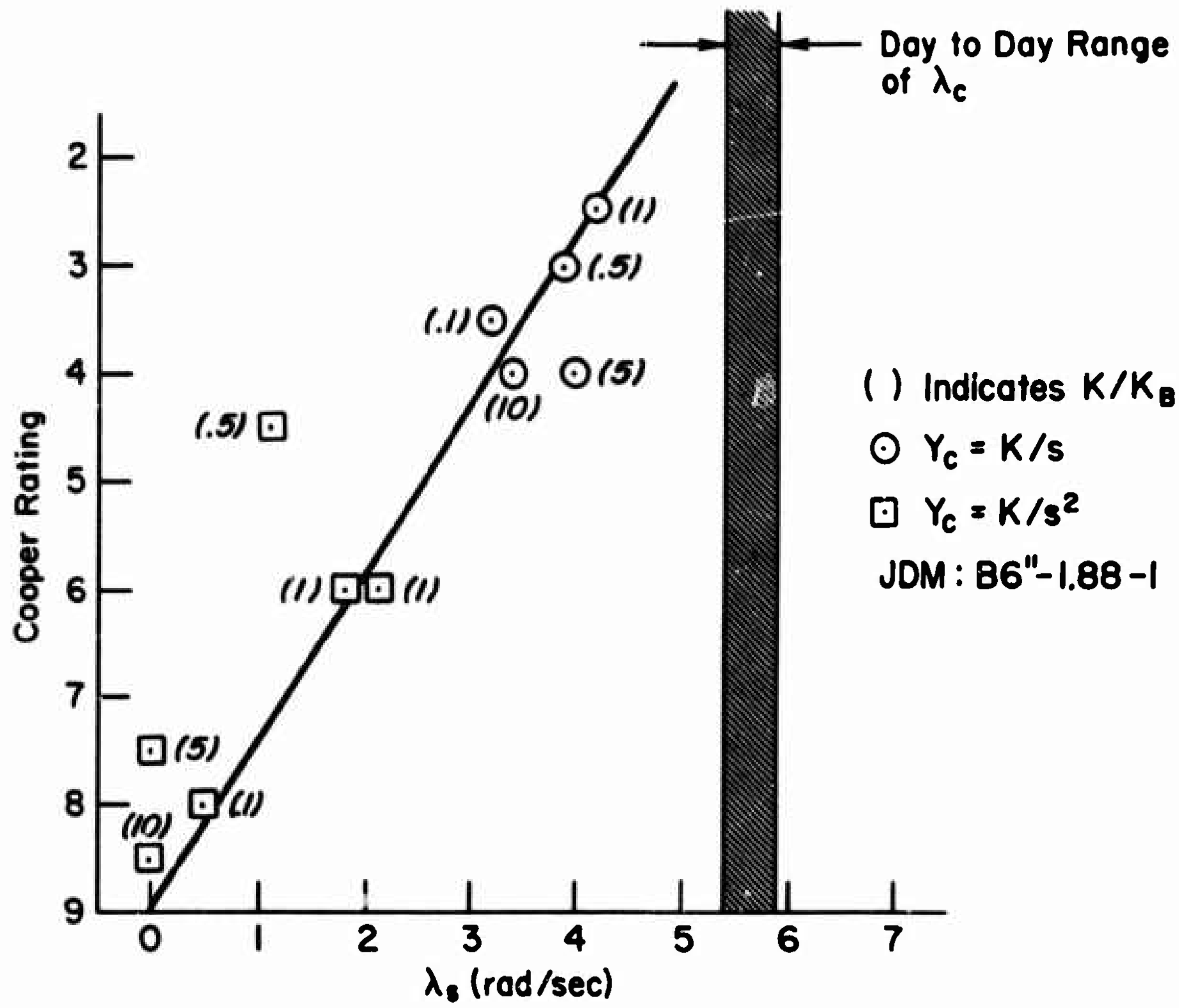


Figure 34. Variation of Secondary Task Score with Controlled Element Gain

The effects of changing the input parameters are seen in Fig. 35. The scatter has increased somewhat, but agreement is still good. The entire experiment has been plotted for subject JDM in Fig. 36. Of the 45 configurations, 73 percent are within one Cooper rating of the regression line.

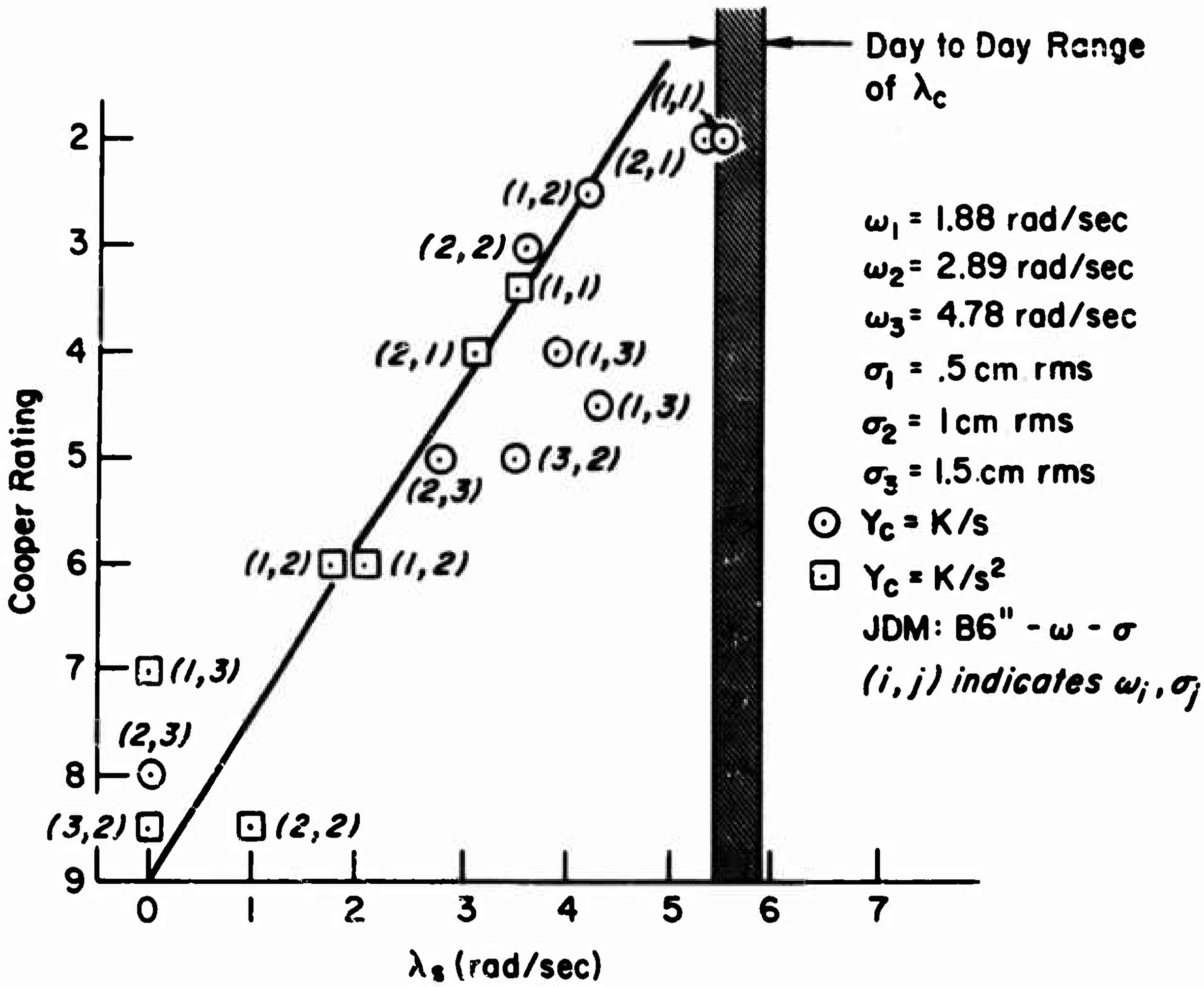


Figure 35. Variation of Secondary Task Score with Input Amplitude and Bandwidth

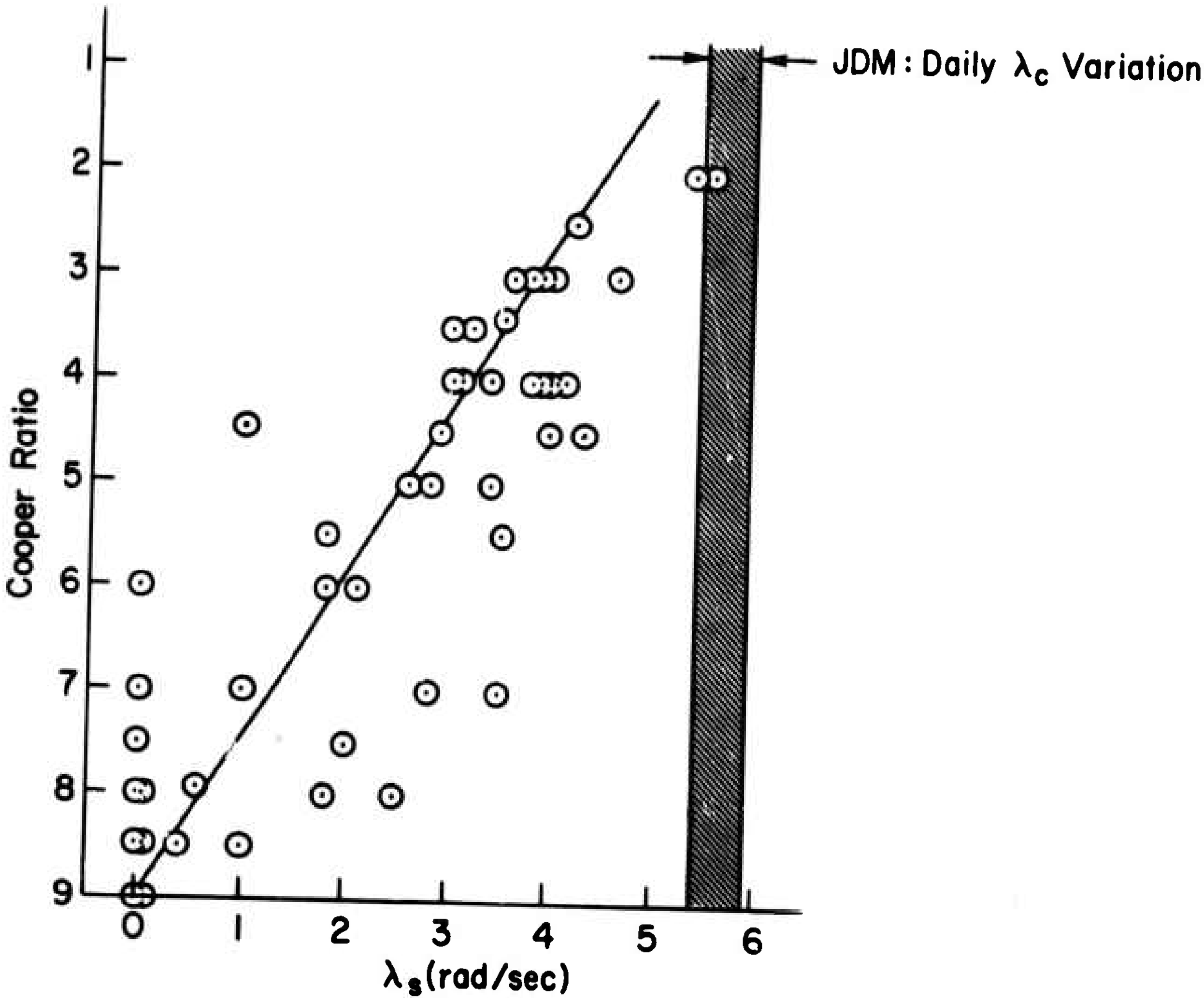


Figure 36. Secondary Task Scores for All Configurations and Inputs

5. Correlation of Ratings with Neuromuscular Tension

Past experience [McDonnell (29), McRuer (5)] with difficult tasks has indicated that in many cases the pilot becomes extremely tense, that is, exhibits a high degree of neuromuscular tension. It was hypothesized that this tension, or effort, would be a chief determiner of pilot rating, since "effort" or "work" invariably comes up in any discussion of handling qualities ratings. Thus the pilots were instrumented with electromyograph (EMG) probes to attempt to measure such a parameter. The most sensitive area on the arm was determined to be the triceps, where electrodes were attached and monitored during the runs.

The results showed that neuromuscular activity increased only as controlled element gain decreased, as would be expected. Average tension level, as a function of element form (and consequently as a function of τ_e), appears indeterminate, as is shown in Fig. 37. Especially surprising was the relatively low value for the subcritical task, which was expected to be the largest in view of subjective comments made during other experiments (McDonnell, Ref. 29). It is apparent that the average tension level is perhaps a less reliable indicator of limb activity than measures of external performance, such as average stick motion, while its significance as a measure of pilot rating in terms of internal effort is doubtful. It

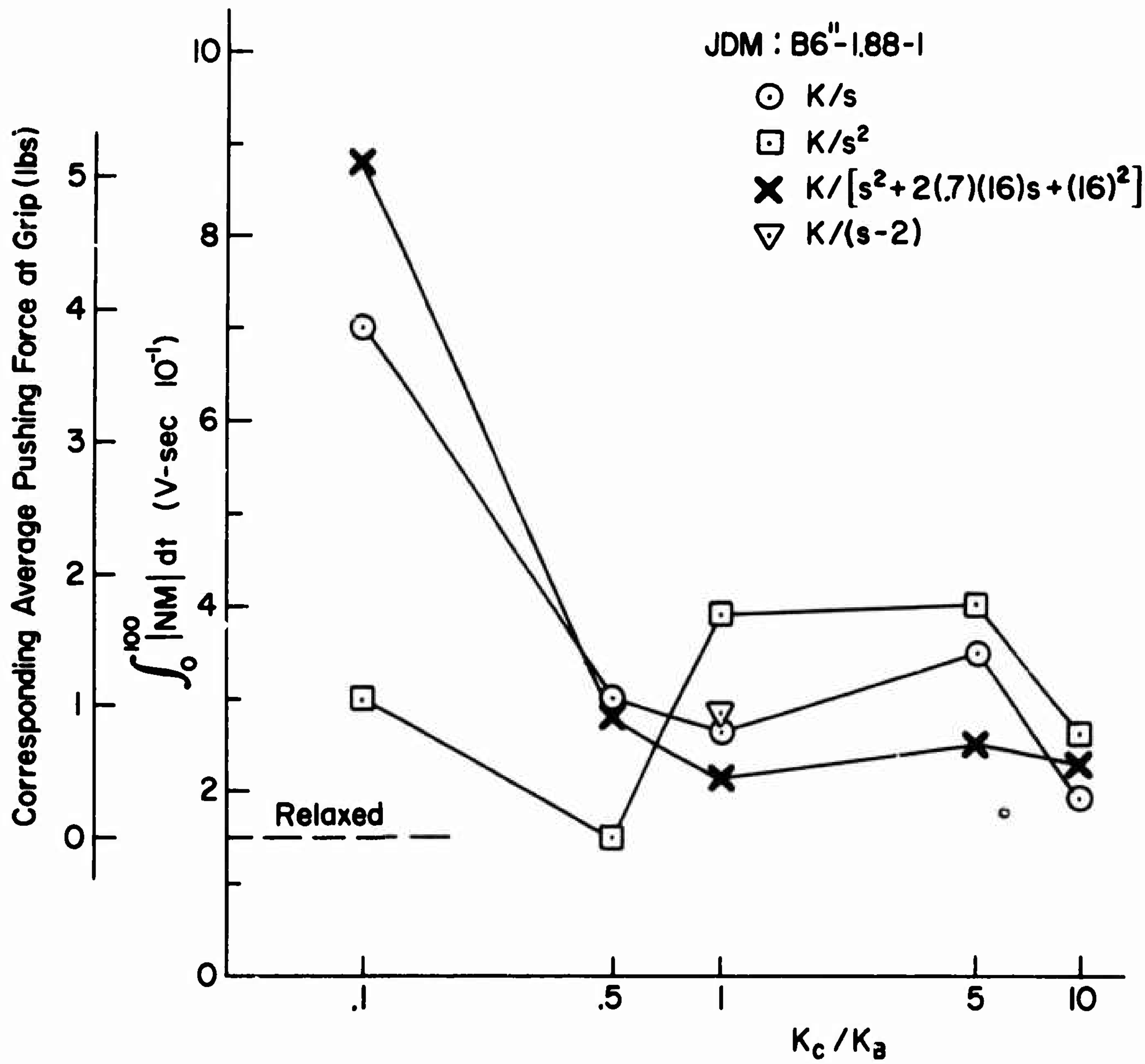


Figure 37. Average Tricep Tension While Tracking

is concluded from the data that the average internal tension is not a primary causal factor in pilot ratings.

6. Comparison of Cooper and Cornell Ratings

A limited amount of rating data, heretofore unpublished, was taken in 1963 as part of a large program (McRuer, Ref. 5). The pilot used and compared the Cooper scale and the Cornell scale for two configurations, $Y_c = K/s$ and $K/s(s-\omega_n)$. It would be of interest to compare those data with the results of the current series. The task carried out was compensatory tracking, where a laterally moving dot was controlled with a roll side stick. The pilot interpreted an inch of lateral dot displacement as 30 deg of bank angle. The criterion, or performance required for the task, is not clear quantitatively, but the pilot considered the task to be approximately straight and level cruising flight. It is interesting to note that the pilot felt that he had to maneuver the configuration in an open-loop fashion without an input in addition to the compensatory tracking before he would give a rating. Thus, in terms of the structure evolved in Section II of this report, he was rating on an undefined combination of tasks.

The plotted Cooper rating data of Figs. 38 and 39 is taken from Table D-I in Appendix D. The Cornell ratings shown in the figures are not tabulated. Comments made by the pilot indicated that he felt the two scales were identical at the good end and were approximately a point different at the bad end, with the Cornell rating being the larger of the two. Figure 39 reflects the point difference between the scales in the 6 to 10 region. No comments were made about the midranges, but Fig. 38 shows that the difference between the scales there increases somewhat linearly with the ratings.

An interesting observation on variability: Fig. 38 shows a marked increase in scatter for the poorer ratings, thus supporting our earlier findings regarding the sensitivity of the rating scales.

A comparison between the earlier data and the ratings obtained in this study was made by normalizing the gain of the earlier controlled

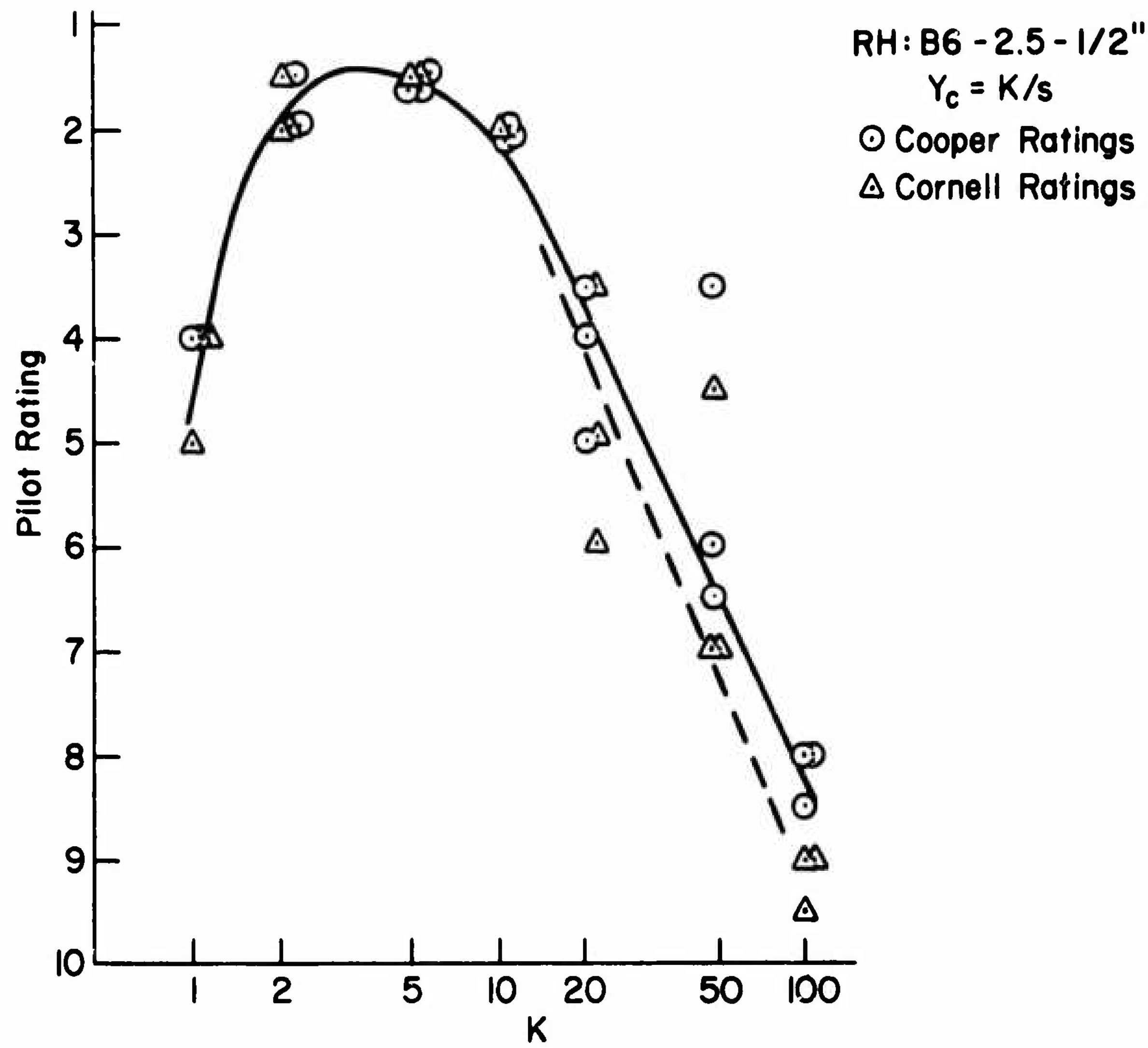


Figure 38. A Comparison of Cooper and Cornell Ratings for $Y_c = K/s$

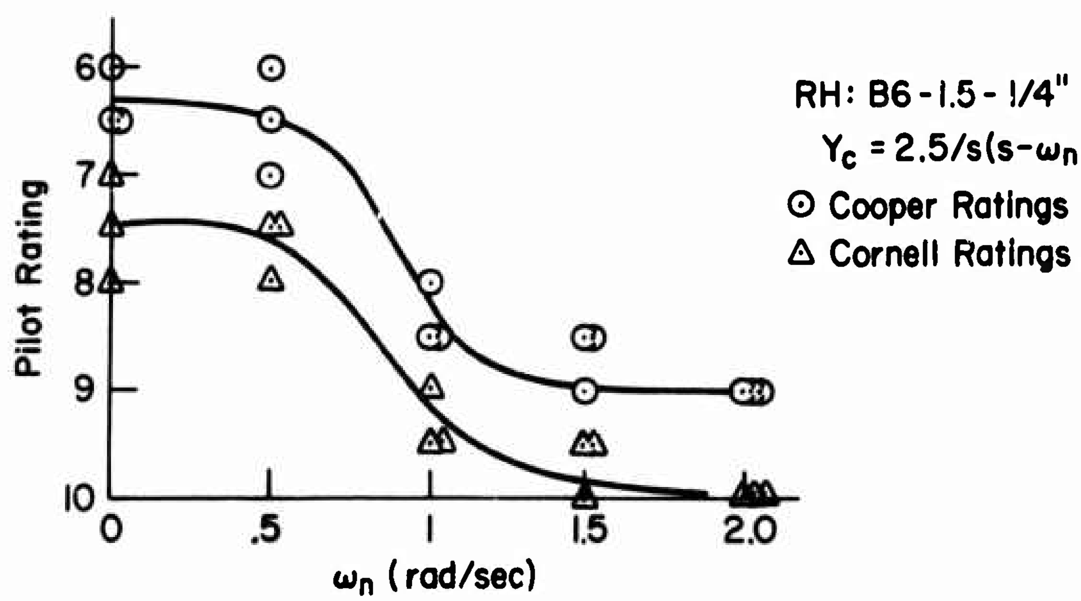


Figure 39. A Comparison of Cooper and Cornell Ratings for an Unstable Second-Order Controlled Element

element. The differences, shown in Fig. 40, are quite dramatic. A plausible explanation is the difference in tasks. As noted earlier, RH was rating on the basis of a qualitative cruise-like condition, and based his ratings in part on open-loop, no-input characteristics. Although the differences are not conclusively due to task definition, the importance of making a complete and concise specification of the task can be appreciated.

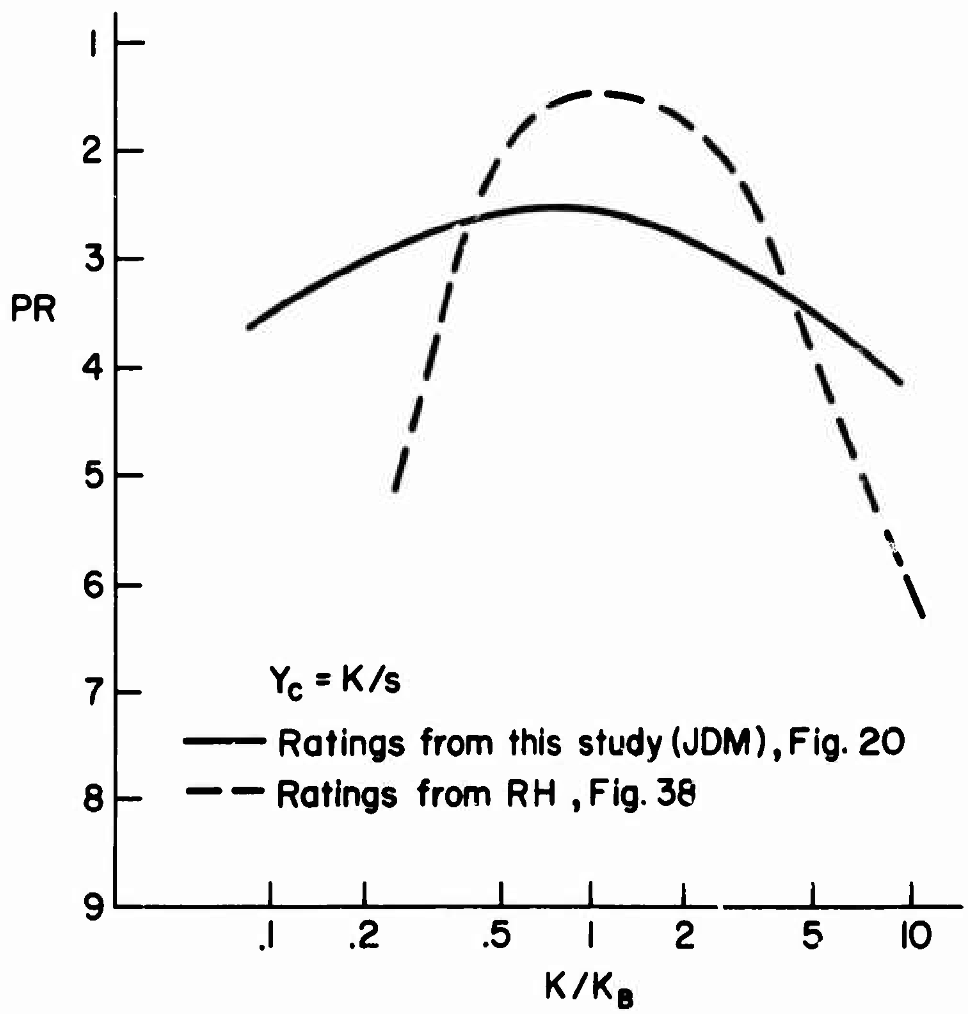


Figure 40. A Comparison of Cooper Ratings for Two Tasks and Two Pilots

D. CONNECTIONS BETWEEN EXPERIMENTALLY MEASURED RATINGS

In addition to the many parameters obtained from the describing functions, several ratings were taken for each configuration. The scales selected are given in Section IV.C, and included the Cooper scale, the revised Cooper scale, a "Handling Qualities" scale, and four "trait" rating scales. The "Handling Qualities" scale (HQ) was intended to overcome some of the difficulties of the Cooper scale by providing a continuous sequence of compatible descriptors across the entire scale. The trait ratings were solicited with the hope that they would provide specific information to the experimenter on the nature of the deficiencies. The connections between these ratings will be examined subsequently.

Because of the large amount of interest shown in the "Cooper boundaries," i.e., the divisions between satisfactory and unsatisfactory (3.5) and between acceptable and unacceptable (6.5), the experiment was designed to test the existence and stability of them by the following procedure:

- The Cooper rating was solicited for the configuration.
- Another card was presented with the questions:

  1. Is the vehicle controllable during the task?
  2. Is the vehicle acceptable for the task? (May have deficiencies which warrant improvement, but is adequate for the task.)
  3. Is the vehicle satisfactory for the task? (i.e., adequate for the task without improvement.)

Upon scrutiny of the data it was apparent that the experiment would not yield the correct results because the short-term retention of the pilot enabled him to rate consistently between both ratings. Thus, in the entire experiment with both subjects, no variation was found in the "boundary" versus Cooper ratings. The boundary ratings will therefore not be considered further.

It was concluded that in order for such an experiment to yield valid results, pilots would have to be used who had no previous knowledge of the Cooper scale, and each configuration would have to be presented twice, once for each rating. The Cooper scale would need to be modified so as not to include the boundary adjectives, but only the descriptors and numerical values. A comparison could then be made between the boundary ratings and the descriptors. Unfortunately, the experiment would be quite lengthy.

1. Comparison of Cooper, Handling Qualities, and Cooper-Harper Ratings

A comparison of the Cooper ratings with the Cooper-Harper ratings for all configurations showed that with one pilot (JDM), out of 57 ratings, 3 were 1 unit different, 16 were 0.5 unit different, and the rest were identical. With the other pilot (MDK), out of 84 ratings, 2 were 3 units different, 10 were 2 units different, 70 were 1 unit different, leaving only two with no difference at all. In virtually all the configurations where differences between the two ratings did occur, the Cooper-Harper rating was the larger (worst) of the two, indicating a possible slight bias toward the bad side. It is obvious that the bias is a function of the pilot, since pilot MDK had an essentially fixed difference of 1 unit. The cause of the bias is unknown, especially in light of the fact that the satisfactory-unsatisfactory/acceptable-unacceptable boundaries are identical in both scales. In the subsequent discussion, no distinction will be made between the Cooper and Cooper-Harper ratings, thus reducing the number of plots required.

In Section III, the semantic relationship between the various and sundry phrases, including Cooper's, was determined and is shown in Figs. 41 and 42 as the "Line of Semantic Agreement," i.e., the calibration between Cooper ratings and the ψ scale that was found from the semantic experiment described in Section III and given by Eq. 8. The actual ratings obtained in the simulation are plotted and can be compared to the calibration line. The numbers on the data points indicate how many identical ratings were obtained. Bear in mind that the calibration line is a theoretical relationship based on data obtained from a semantic experiment, whereas the data points are actual rating data. As such, the "true" ratings are unknown and are best estimated from the data. The differences between the data and the calibration line are definitely one-sided. A possible explanation for this is determined by returning to the original questionnaires (see Appendix A). There it can be seen that both pilots used in the experiments were more pessimistic than average, which could explain the bias noted in the plots.

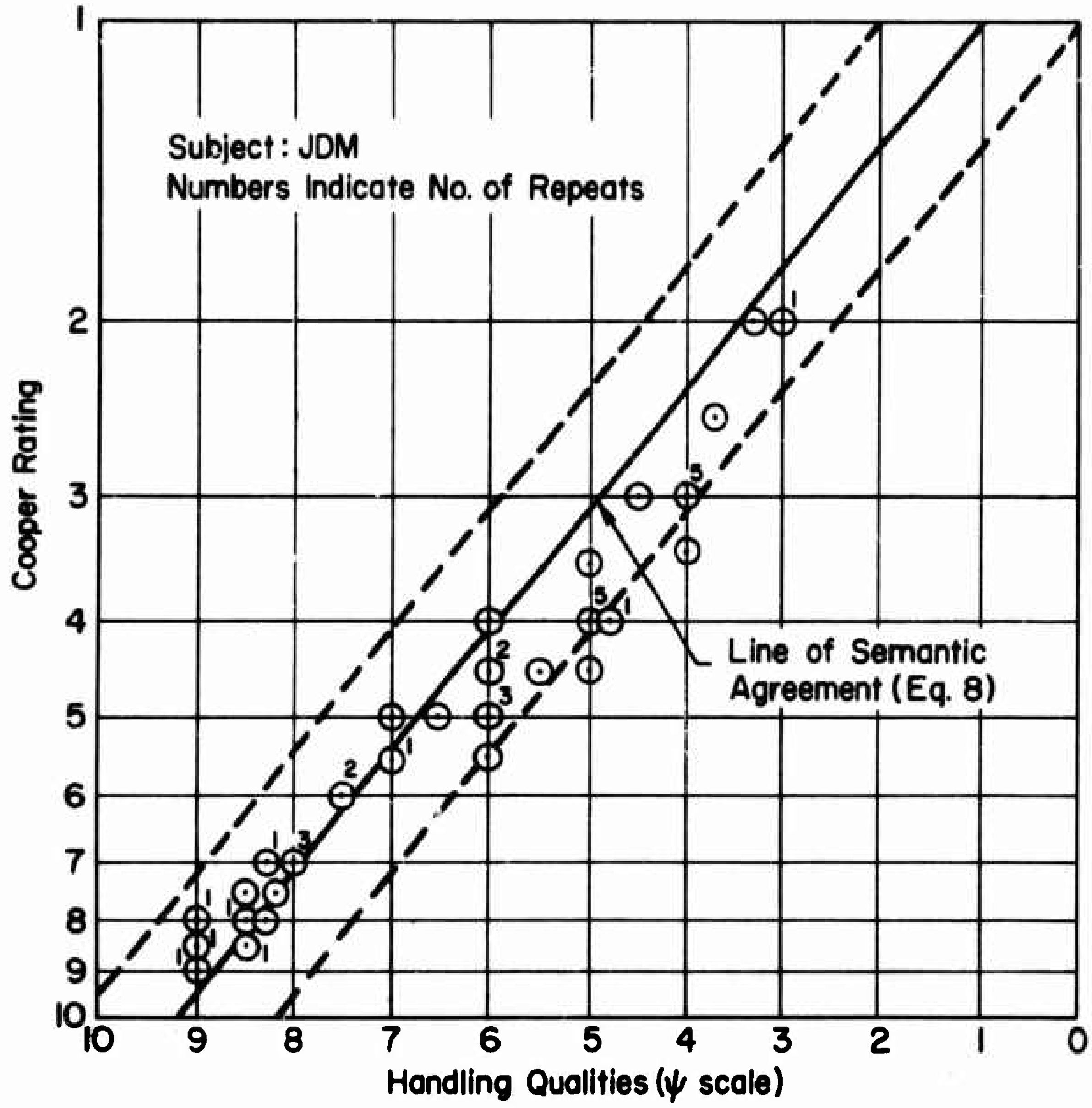


Figure 41. Comparison of Cooper Ratings and Handling Qualities Ratings for All Configurations; Subject JDM

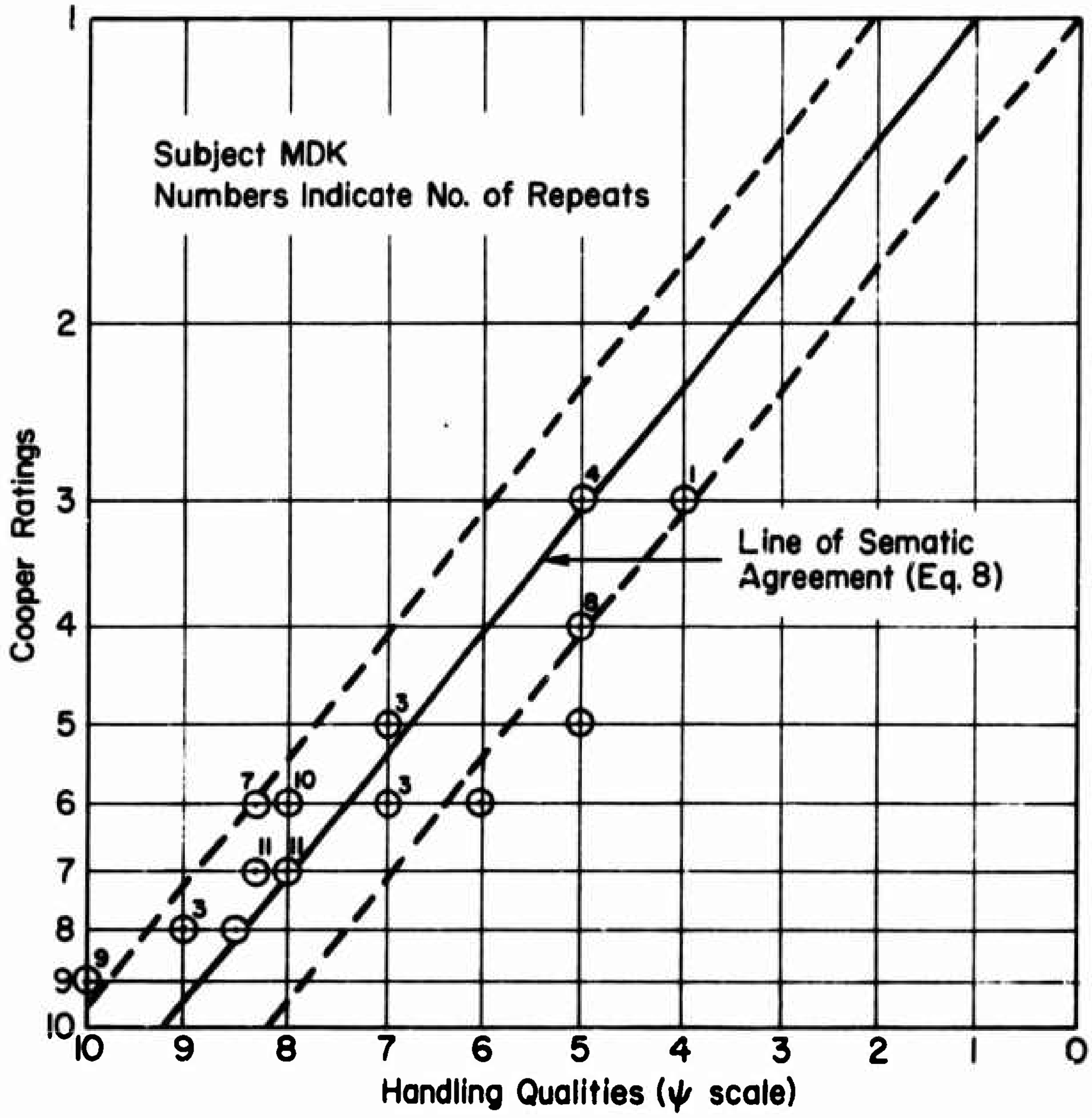


Figure 42. Comparison of Cooper Ratings and Handling Qualities Ratings for All Configurations; Subject MDK

If a pilot introduces a systematic variability in all ratings, the effect on the data of Figs. 41 and 42 would be to slide the data points down the calibration line (or parallel to it if a bias is present). If the pilot has a purely random variance (as he seems to have in the semantic experiment as determined by comparing the scores to overall means) the observed variance could be as large as the variance noted in the semantic experiment, i.e., the square of the discriminal dispersion.

The discriminal dispersion has been shown by the dashed lines in Figs. 41 and 42. Virtually all of the data are contained by these lines, which indicates: 1) the bias present in each pilot's ratings is within 1σ of the average pilot, and 2) **it appears that most of the variability is due to semantics and not to the evaluation process.** Remember that we are not considering the variability of ratings for repeated configurations or bias differences between pilots, but only the relative semantic ambiguity between the Cooper descriptors and the Handling Qualities descriptors.

It is concluded on the basis of these data that our earlier findings that the Cooper scale becomes more sensitive at the bad end are correct, and that in an actual rating situation the resolution capability of the pilot is being taxed beyond its power when significance is placed on differences of 1 Cooper unit with only a few observations at the bad end of the scale.

The fact that there is semantic consistency in the ratings of two pilots does not mean that they will closely agree upon the merits of a particular vehicle. It is an indicator of the level of confidence that can be placed on resultant ratings, considering also the pilot's "set" (how his preference is affected by training, experience, etc.) and sensitivity to vehicle parameter changes (how his deterioration in ratings is affected by motivation, ability, and self-assessment of performance).

The priority of attributes to be possessed by a pilot is fairly clear. It is of absolute importance that the pilot have a good ability to use words. Unfortunately, the administering of a test which would give data similar to that of Figs. 41 and 42 is not at all an easy matter. One alternative is to use the conventions of the past, i.e., choose pilots who have a strong educational and technical background. The participants of the semantic survey were all carefully chosen. Out of 67 raters, 4 had to be disregarded because of glaring inconsistencies. Since it was impossible to check the causes, it was assumed that lack of motivation or a misunderstanding of the instructions were most likely the causes, not an inability with words.

Another alternative would be to construct a very limited version of the semantic survey (maybe ten key phrases) to administer to possible rating candidates. Criteria could be established for acceptance or rejection of the rater based strictly on verbal ability. The candidate would also be required to have the education, background, and experience appropriate to the rating task.

Considering the results of the survey, it is doubtful that such a screening is necessary if raters do have the appropriate background and are thoroughly motivated.

2. Comparison of Global Ratings with Trait Ratings

In addition to the global ratings (as overall ratings are often called, i.e., Cooper, Handling Qualities, Cooper-Harper), ratings of Response Characteristics, Control, Demands on Pilot, and the Effects of Deficiencies were obtained. The phrases used were those previously scaled in Section III.C and shown in Table XVI. The intent of such trait ratings was that they would very likely be closely related to physical characteristics of the vehicle or system and thus aid the engineer in determining the appropriate improvement, or at least in identifying the problem.

Table XVIII shows some anticipated interactions between the traits and several important pilot, vehicle, and system parameters. As an example, if the controlled element form and input are held fixed in a closed-loop tracking task, but the vehicle gain is changed, we know that pilot rating will change (Fig. 20), but that performance in terms of what the pilot sees will remain constant. Thus, as a function of gain, it was anticipated that the rating of "Response Characteristics" would remain approximately constant, while the ratings of "Ease of Control" and "Demands on the Pilot" would vary widely.

TABLE XVIII

ANTICIPATED PILOT/VEHICLE SYSTEM CORRELATES FOR TRAITS

TRAIT	SOME ANTICIPATED PILOT, VEHICLE, AND SYSTEM PARAMETER CORRELATES FOR:	
	OPEN-LOOP MANEUVERS	CLOSED-LOOP TRACKING
1. Response Characteristics (RC)	Vehicle numerator and denominator time constants, T_L, T_I, command input	ω_i, σ_i, system bandwidth, remnant level (φ_{nn})
2. Ease and Precision of Control (C)	Vehicle damping and natural frequency, stick characteristics, K_c (or K_p)	φ_m, K_m, φ_{nn}, e(t), stick characteristics, K_c (or K_p), T_L, T_I
3. Demands on Pilot (DP)	Complexity of open-loop response to command input, stick characteristics	τ_e, T_L, T_I, K_p, K_c
4. Effects of Deficiencies on Performance (ED)	Overshoot, rise time, settling time	$\overline{e^2}/e_c$

Figures 43, 44, and 45 show a summary of the results of the trait ratings for both pilots. Observations of a general nature are that:

1. There is a somewhat uniform trend between the Handling Qualities (HQ) rating and the corresponding trait ratings, i.e., all traits seem to suffer when the global rating deteriorates.

2. When there is disagreement between pilots on the overall adequacy of the configuration for the task, the contributing factors are reflected by all of the traits. A pilot "set," then, seems to be exhibited by all of the traits. This could mean that (a) the traits measure independent features of the vehicle which all vary a similar amount, or that (b) the traits are all describing the same phenomenon.

In some specific instances, lack of consistency can be observed. Figure 44a shows that one pilot rated a low-gain configuration much <u>less demanding</u> than the high-gain case, even though it took as much as 100 times

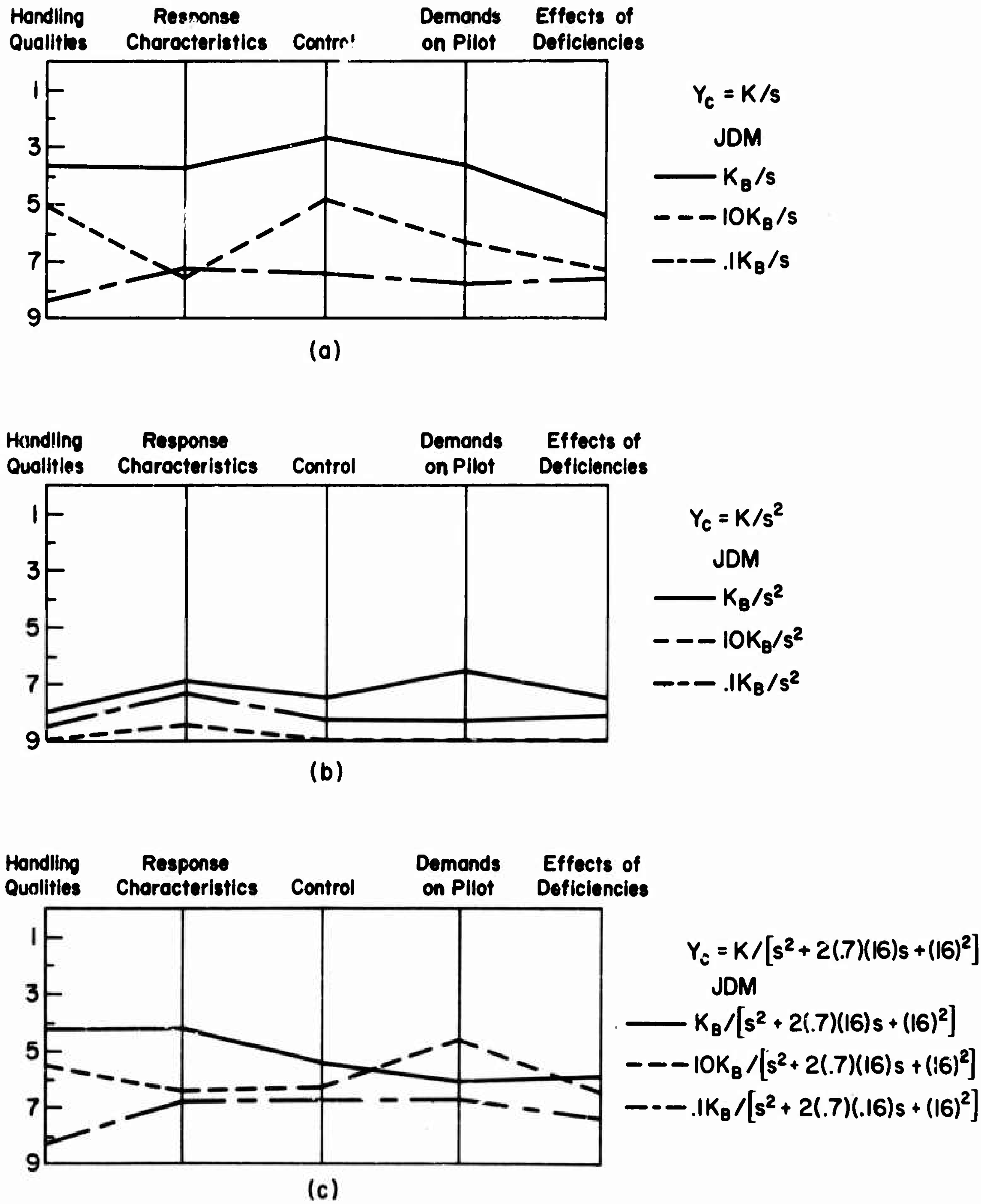


Figure 43. Variation of Trait Ratings with Gain for JDM

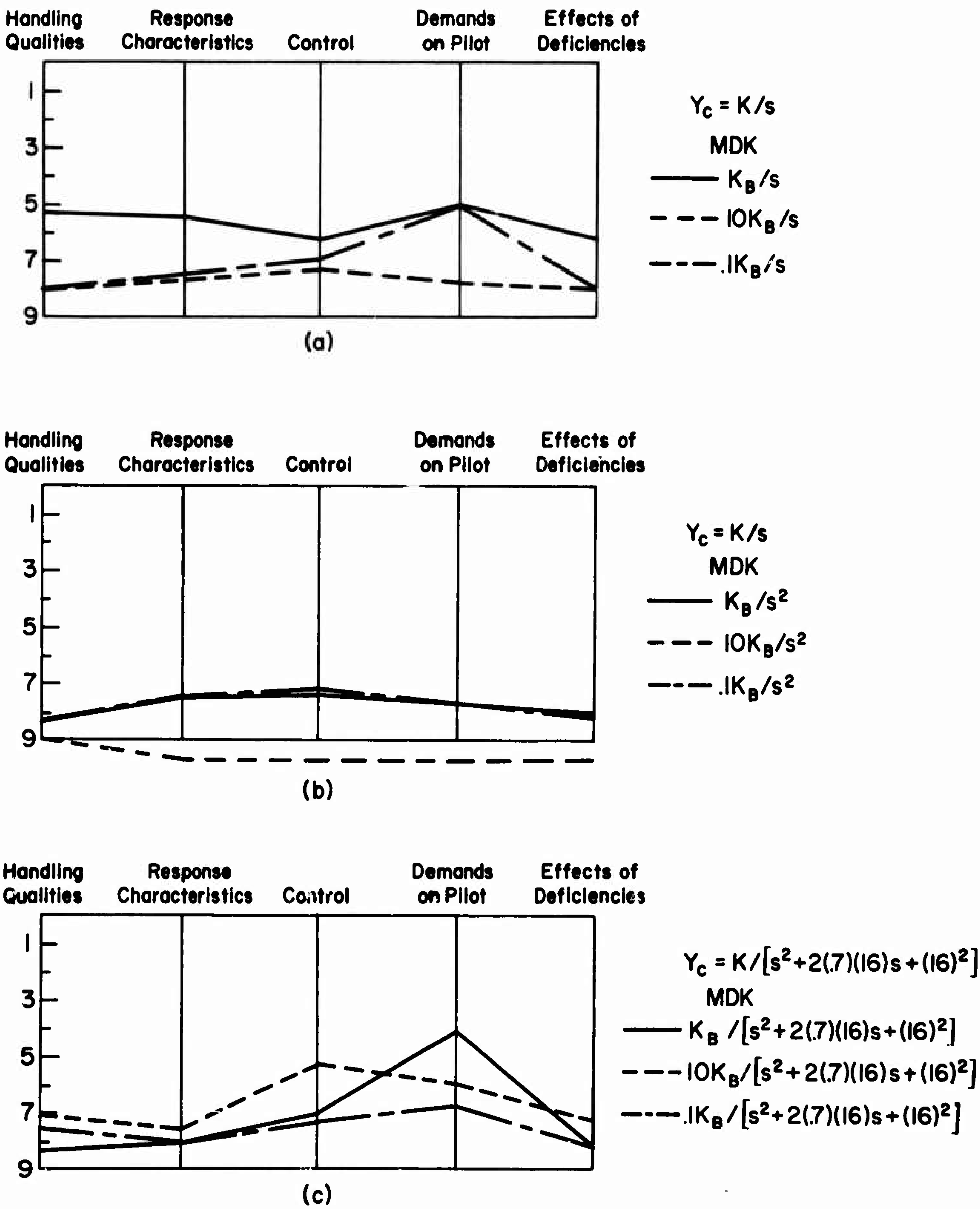


Figure 44. Variation of Trait Ratings with Gain for MDK

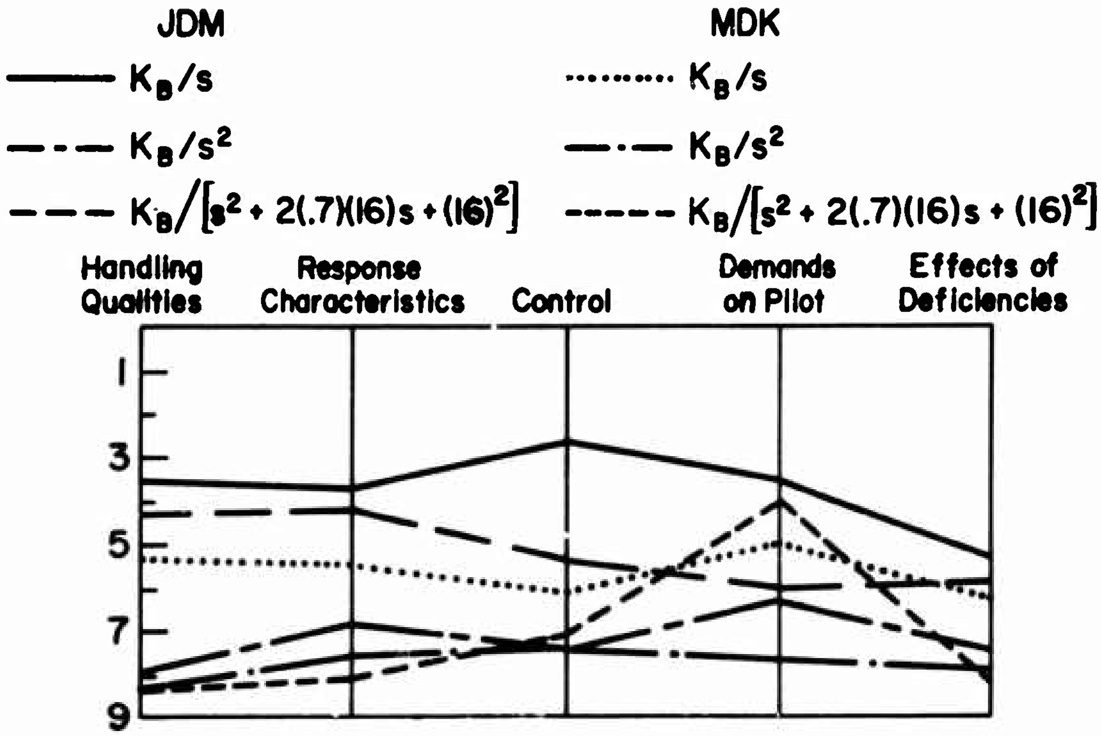


Figure 45. Controlled Element Form Effects on Trait Ratings

the stick travel and force to obtain equivalent performance. Figure 45 shows that for the complex-pair controlled element, the demands on the pilot were rated in opposite directions by the two pilots.

Taking into consideration the observed trends and inconsistencies, the usefulness of the trait ratings as supplementary indicators appears to be very limited. The connections between the traits and specific parameters were originally intended to be investigated via computerized correlation and factor analysis techniques. However, on the basis of the results of Figs. 43, 44, and 45, it is concluded that a considerably larger population of pilots would need to be sampled before any useful results could be obtained. The scaled trait descriptors could be used, however, to construct a specialized global scale, should an experimenter need one.

A possible alternative to the scaled trait ratings would be Osgood's (30) semantic differential type of rating scale, where the extremes of several subjective qualities are presented to the pilot and he is forced to select some degree of goodness of each by placing a mark on the line joining the two extremes. The disadvantage of such a technique is that no meaningful numerical values can be assigned the resultant ratings. Perhaps a fruitful area of research would be the use of psychometric methods to scale the data obtained in semantic differential or forced choice form in a display evaluation (34), for example.

At this time it must be reluctantly concluded that the scaled trait ratings are of no apparent value in pointing out areas of deficiencies to the engineer.

E. GENERAL APPROACH TO RATING ESTIMATES

Because of the lack of data pertaining to pilot "set," or individual differences between pilots, it is premature to attempt to construct a pilot rating model. However, it is felt that the data are sufficient to enable estimates of <u>increments</u> of ratings due to vehicle and environmental changes. The general approach is outlined below.

Because of the complex nature of pilot adaptation, caution is absolutely necessary when attempting to anticipate a rating for a given configuration. The two primary questions that must be answered are: 1) what performance can the pilot attain relative to that specified, and 2) how near to his adaptation limits is the pilot while maintaining the performance. The first question is answered by conducting an analysis of the pilot/vehicle system. In the case of compensatory tracking, the adjustment rules of McRuer (5) generally provide a good estimate of overall performance that can be expected. If performance is worse than that specified in the definition of the task, decrements in rating similar to that shown in Fig. 27 would be expected.

The second question can be answered by estimating the individual pilot parameters. If the crossover model of the operator is valid, pilot ratings would be expected to be proportional to the effective time delay, Fig. 21, which in turn reflects both equalization and input effects. If the crossover model is not suitable, as in the subcritical task, a more detailed analysis would be in order to determine if the operator is near his limits. The effects of a regression (i.e., increase) of τ_e with a large ω_i were not investigated in the present experiments.

The pilot also has definite preferences for control stick characteristics. If his preferred gain is known, the decrement due to nonoptimum gains can be predicted from Fig. 20.

The question is always asked, "What does it mean when the sum total of all of these effects indicates a rating far worse than the worst on the scale — say, a Cooper rating of 20?" The answer is simply that the scale is not absolute, but only relative. Ratings must therefore be truncated at 9, which is somewhat analogous to admitting that most home thermometers would not yield a correct measure of 0° Kelvin!

Hopefully, rating variations have been shown with enough pilot and system parameters to enable the engineer to estimate relationships with confidence and with a minimum of analysis. A significant amount of work remains to be accomplished, however. The next section will detail recommendations to further improve the state-of-the-art, and will summarize the many conclusions reached throughout the study.

SECTION VI

SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS

The study program described herein has led to a large number of very interesting findings, which can be drawn together in this section to form a fairly complete picture of the current state of rating technology. The findings lend themselves to a natural division into two categories. The first part of this study was aimed at the problems of rating scales themselves, and led to a somewhat separate and independent set of conclusions. It will be discussed first. Then the effects of the physical system on ratings can be discussed.

A. SUMMARY OF RATING SCALE FINDINGS

Rating scales are subjective in nature and therefore are scales of comparison. As such, they should have no absolute values associated with them. The use of rating scales will result in such phenomena as pilot biases due to personal preferences based on training, experience, and general background; differences due to interpretation of the objectives of the rating situation; and biases and variability due to deficiencies in the scale itself. The first source of bias can be minimized by careful planning and definition of the criteria used in the experiment; the second and third are amenable to analysis and improvement.

A considerable amount of effort was devoted to the interpretation problem in Section II, where "ground rules" regarding definitions of missions, tasks, etc., were established. Thus, the bias due such factors can be minimized, and the interchangeability and consistency of experimental data should be much improved.

The problems with the scale itself were noted in Section II, and attacked in earnest in Section III. An application of psychometric methods yielded a set of scaled descriptors showing that

1. There is an underlying psychological dimension, or continuum (called the ψ scale herein), which has a constant subjective sensitivity along its length. A measure of the sensitivity is called

the "discriminal dispersion," and is essentially the standard deviation of the resolving power of raters to distinguish semantic differences in language. The constant sensitivity yields an interval scale, where the intervals are units related to noticeable semantic differences. The interval nature of the dimension allows ratings to be averaged, which has heretofore been mathematically inappropriate.

2. The Cooper scale (1) and Cooper-Harper scale (3) are very nearly functionally related to the ψ dimension. The error introduced by averaging Cooper ratings, rather than their ψ equivalent, is small provided enough trials have been made to ensure confidence in the ratings (see next paragraph).

3. The Cooper and Cooper-Harper scales are shown to be overly sensitive at the bad ends, so that attaching significance to a difference of one Cooper unit between ratings at the bad end would require a relatively large number of trials.

4. The results of the current experiments show an internal consistency between the Cooper phrases, ψ values, and Cooper ratings to such an extent that it is concluded that a scale based on the ψ-scale values would solve many of the problems which currently exist. Such a scale might appear as shown in Fig. 46. There, "degrees of goodness" of handling qualities are distributed along a 7-point scale, which has a uniform sensitivity along its length. The scale shown would be called a "global" scale, since it integrates all deficiencies into the one descriptor "handling qualities." Specialized scales could be similarly constructed by using the catalog of scaled phraseology included in this report.

The choice of a 7-point scale is somewhat arbitrary, although it is felt that it would be optimum in that it would be sensitive enough to detect significant differences in opinion but at the same time would not tempt the pilot into reporting differences which could not be statistically confirmed.

In any event, the ψ-scale values given in this report can be linearly transformed to any interval base from the 9-point scale on which they were based.

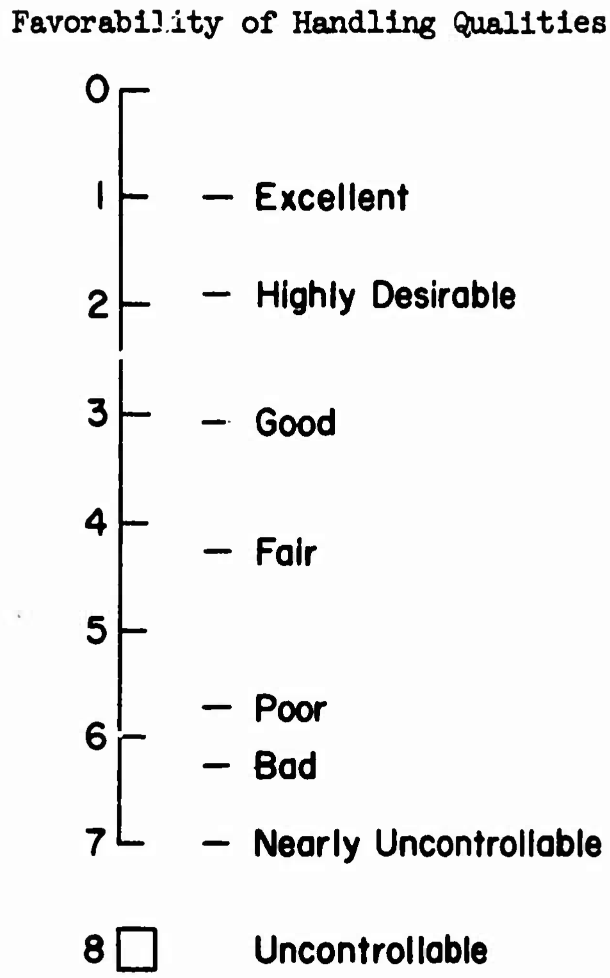


Figure 46. A Global Rating Scale for Handling Qualities Evaluation

Two rather negative and disappointing conclusions regarding the investigated scales are:

1. The verification of the existence of the Cooper boundaries (i.e., the satisfactory-unsatisfactory boundary at 3.5, and the acceptable-unacceptable boundary at 6.5), and the stability of them relative to the scale descriptors could not be determined. This is considered the final link necessary to prove the validity of the excellent decision tree type of process introduced in the Cooper-Harper scale (3). An experiment which would demonstrate boundary existence is suggested in Section V.D.

2. The trait ratings, which had initially been proposed to construct auxiliary scales for the purpose of rooting out specific physical vehicle deficiencies for the engineer were disappointing. The variability and lack of consistency between the two pilots indicates that the traits chosen for investigation are not selective.

> A large population of pilots, together with the computer aids of regression and factor analyses potentially could provide the desired relationships, but the likely attendant confidence levels would make the usefulness of such ratings doubtful.

The investigation of the possibility of obtaining numerical data when using the semantic differential technique has been suggested in Section V.D as a possible alternative to the trait ratings. Some additional research into scaling techniques would be required.

B. SUMMARY OF RATING CORRELATIONS WITH PILOT, VEHICLE, AND SYSTEM PARAMETERS

The considerable data available indicate that, where closed-loop compensatory tracking is the task, the pilot's increments in rating are based on the relative difficulty with which he obtains and maintains the specified performance. An estimate of performance is obtained directly. An indication of the difficulty involved, however, is not so obvious. Perhaps the most direct measures, judging from the data, are the gain required of the pilot, which directly determines muscular activity and sensitivity, and the equalization required of the pilot for stability.

The interactions between these parameters and the other system parameters are quite complex; nevertheless, a growing body of literature is available to aid the engineer in estimating them. Rating correlations with other parameters are also shown to be of potential use to the engineer in rating estimation, but are less direct.

The notion that task performance and difficulty are the causal factors of pilot ratings was further supported by an experiment measuring an "attention level" related parameter. A secondary task was used to "load" the pilot so that primary performance began to deteriorate. The correlations given in Section V.C show that good agreement exists between the level of difficulty attainable with the secondary task and the rating for primary task alone. This application of a secondary task to find the "attention level" or "excess capacity" of the pilot has an excellent potential of becoming an objective measure of pilot rating which can be related directly to pilot and system parameters.

The technique was not optimized, nor has any supporting theory been evolved. The results indicate that the application does have the potential of supplying the handling qualities community with a "pilot rating thermometer." It is therefore recommended that some additional work be carried out along the lines of optimization of the technique, and that some effort be directed at a theory connecting secondary loading score with primary effective time delay, channel capacity, maximum data rates, etc.

A negative conclusion can be drawn from the neuromuscular tension data. It was initially hypothesized that the task difficulty would also be reflected by the overall muscular tension level, which could even be a primary "cause" of decrement in rating. The data did not bear this out, however. The average tension level did increase with increased stick displacement, which is a rather trivial result, but also a result which confirms the accuracy of the measurement method.

The limited number of participating pilots (two) precluded the discovery of any "set" or "motivational" rules. The correlation results are thus really only applicable to <u>incremental</u> changes in rating. It is suggested that the problem will be extremely difficult to quantify. Therefore, another appeal will be made here to the engineer: thoroughly specify the task, including required performance. Publish the task specification along with the data. Only then can useful data be interchanged between experimenters and designers.

Finally, because of the vast amount of data accumulated during this study, the choice between correlations of parameters versus Cooper ratings or versus ψ ratings had to be made in many places for the sake of space and economy. Since so many previous Cooper rating correlations exist, and because such a wide audience has been exposed to them, the Cooper rating was usually selected. However, it has been shown here that the bad end of the Cooper scale can be misleading because of a pilot's lack of sensitivity at that end. It is therefore suggested that a scale similar to that shown in Fig. 46 be developed. Any averaging will then be legitimate, variabilities will be constant across the scale, and the number of necessary trials will be fixed across the scale.

REFERENCES

1. Cooper, G. E., "Understanding and Interpreting Pilot Opinion," Aeron. Eng. Rev., Vol. 16, No. 3, Mar. 1957, pp. 47-52.

2. Harper, Robert P., Jr., In-Flight Simulation of the Lateral-Directional Handling Qualities of Entry Vehicles, WADD TR 61-147, Nov. 1961.

3. Harper, R. P., Jr., and George E. Cooper, A Revised Pilot Rating Scale for the Evaluation of Handling Qualities, Cornell Aero. Labs. Rept. 153, Sept. 1966.

4. Weir, David H., Compilation and Analysis of Flight Control System Command Inputs, AFFDL TR 65-119, June 1965.

5. McRuer, Duane, Dunstan Graham, Ezra Krendel, and William Reisener, Jr., HUMAN PILOT DYNAMICS IN COMPENSATORY SYSTEMS — Theory, Models, and Experiments with Controlled Element and Forcing Function Variations, AFFDL TR 65-15, July 1965.

6. Sensory Communication, ed. Walter A. Rosenblith, Pub. jointly by MIT Press and John Wiley and Sons, Inc., New York, 1961.

7. Guilford, J. P., Psychometric Methods, McGraw-Hill Book Company, Inc., New York, 1954.

8. Torgenson, W. S., Theory and Methods of Scaling, John Wiley and Sons, Inc., New York, 1965.

9. Uhrbrock, Richard S., "Standardization of 724 Rating Scale Statements," Personnel Psychol., Vol. 3, 1950, pp. 285-316.

10. Ferguson, Leonard W., "The Development of a Method of Appraisal," Personnel Psychol., Vol. 24, 1947, pp. 127-136.

11. Thurstone, L. L., "Attitudes Can Be Measured," Amer. J. Sociology, Vol. 33, 1928, 529-554.

12. Uhrbrock, Richard S., "Attitudes of 4430 Employees, J. Social Psychol., Vol. 5, 1934, 365-377.

13. Churchman, C. W., and P. Ratoosh, Measurement: Definitions and Theories, John Wiley and Sons, Inc., New York, 1959.

14. Stevens, S. S., "On the Psychophysical Law," Psychol. Rev., Vol. 64, 1957, pp. 153-181.

15. Lefritz, N. M., The Application of Pilot Opinion Data in the Determination of Aircraft Handling Qualities, term paper, Industrial Eng. 571, USC, June 1962.

16. Diederich, G. W., S. J. Messick, and L. R. Tucker, "A General Least Squares Solution for Successive Intervals," Psychometrika, Vol. 22, No. 2, June 1957, pp. 159-173.

17. Cumrey, A. L., A Program for Successive Intervals and Multidimensional Scaling, Dept. of Psychol., UCLA, 1966.

18. Jex, H. R., and C. H. Cromwell, III, Theoretical and Experimental Investigation of Some New Longitudinal Handling Qualities Parameters, ASD TR 61-26, June 1962.

19. Durand, T. S., Theory and Simulation of Piloted Longitudinal Control in Carrier Approach, Systems Technology, Inc., Tech. Rept. 130-1, Mar. 1965.

20. Hald, A., Statistical Theory with Engineering Applications, John Wiley and Sons, Inc., New York, 1952.

21. Ashkenas, I. L., A Consolidation of Lateral-Directional Handling Qualities, AIAA Paper 65-314, July 1965.

22. Jex, H. R., J. D. McDonnell, and A. V. Phatak, A "Critical" Tracking Task for Man-Machine Research Related to the Operator's Effective Delay Time, Part I: Theory and Experiments with a First-Order Divergent Element, NASA CR-616, Nov. 1966.

23. Kelley, C. R., and M. J. Wargo, "Cross-Adaptive Operator Loading Tasks," Human Factors, Vol. 9, No. 5, Oct. 1967, pp. 395-405.

24. Stapleford, R. L., D. T. McRuer, and R. E. Magdaleno, Pilot Describing Function Measurements in a Multiloop Task, NASA CR-542, Aug. 1966.

25. Stapleford, R. L., and I. L. Ashkenas, Effects of Manual Altitude Control and Other Factors on Short-Period Handling Quality Requirements, AIAA Paper 67-579, Aug. 1967.

26. McRuer, D. T., I. L. Ashkenas, and C. L. Guerre, A Systems Analysis View of Longitudinal Flying Qualities, WADD TR 60-43, Jan. 1960.

27. McDonnell, J. D., Preliminary Experimental Investigations of a New Pilot Rating Scale and Pilot Parameter Tracker, Systems Technology, Inc., Working Paper 134-9, July 1964.

28. McRuer, Duane T., and Ezra S. Krendel, Dynamic Response of Human Operators, WADC TR 56-524, Oct. 1957.

29. McDonnell, J. D., and H. R. Jex, A "Critical" Tracking Task for Man-Machine Research Related to the Operator's Effective Delay Time, Part II: Experimental Effects of System Input Spectra, Control Stick Stiffness, and Controlled Element Order, NASA CR-674, July 1966.

30. Osgood, C. E., "The Nature and Measurement of Meaning," Psychol. Bull., Vol. 49, No. 3, May 1952, pp. 197-238.

31. Determination of Flight Characteristics of SST During the Landing Approach with a Large Jet Transport In-Flight Simulator, NASA TN D-3971, June 1967.

32. Meeker, J. I., Evaluation of Lateral-Directional Handling Qualities of Piloted Re-Entry Vehicles Utilizing Fixed-Base and In-Flight Evaluations, NASA CR-778, May 1967.

33. Bullard, E. C., F. E. Oglebay, W. H. Munk, and G. R. Miller, A Users' Guide to BOMM, UC LaJolla, Jan. 1966.

34. Swartz, W. F., Control Display Pilot Factors Program, Instrument Evaluation, USAF, IPIS, Randolph AFB, Proj. No. 62-1B, Sept. 1966.

35. Hall, Ian A. M., Effects of Controlled Element on the Human Pilot, WADC TR 57-509, Aug. 1958.

36. Krendel, Ezra S., and Duane T. McRuer, "A Servomechanisms Approach to Skill Development," J. Franklin Inst., Vol. 269, No. 1, Jan. 1960, pp. 24-42.

37. Ashkenas, I. L., A Study of Conventional Airplane Handling Qualities Requirements, Part I, Roll Handling Qualities; Part II, Lateral-Directional Oscillatory Handling Qualities, AFFDL TR 65-138, Oct. 1965.

38. Creer, B. Y., J. D. Stewart, R. B. Merrick, and F. J. Drinkwater, III, A Pilot Opinion Study of Lateral Control Requirements for Fighter-Type Aircraft, NASA Memo 1-29-59A, Mar. 1959.

39. Dander, Vernon A., "Predicting Pilot Ratings of Multi-Axis Control Tasks from Single-Axis Data," IEEE Trans. on Human Factors in Electronics, Vol. HFE-4, No. 1, Sept. 1963, p. 15.

40. Durand, T. S., and H. R. Jex, Handling Qualities in Single-Loop Roll Tracking Tasks: Theory and Simulator Experiments, ASD TDR 62-507, Nov. 1962.

41. Rolls, L. S., and F. J. Drinkwater, III, A Flight Determination of the Attitude Control Power and Damping Requirements for a Visual Hovering Task in the Variable Stability and Control X-14A Research Vehicle, NASA TN D-1328, May 1962.

42. Faye, A. E., Jr., Attitude Control Requirements for Hovering Control Through the Use of a Piloted Flight Simulator, NASA TN D-792, Apr. 1961.

43. Walton, R. P., Analytical Review of Military Helicopter Flying Qualities, Systems Technology, Inc., Tech. Rept. 143-1, Aug. 1967.

44. Stapleford, R. L., J. Wolkovitch, R. E. Magdaleno, C. P. Shortwell, and W. A. Johnson, An Analytical Study of V/STOL Handling Qualities in Hover and Transition, AFFDL TR 65-73, Oct. 1965.

45. Miller, David P., and James W. Clark, Research on Methods Presenting VTOL Aircraft Handling Qualities Criteria, AIAA Paper 64-618, Aug. 1964.

46. Daw, D. F., D. G. Gould, and D. M. McGregor, A Flight Investigation of the Effects of Weathercock Stability on V/STOL Aircraft Directional Handling Qualities, National Research Council of Canada Rept. LR-400, May 1964.

47. Teper, G. L., and H. R. Jex, Synthesis of Manned Booster Control Systems Using Mathematical Pilot Models, paper pres. at Sixth Annual Symposium of PGHFE, IEEE, Boston, Mass., May 6-8, 1965.

48. Kidd, Edwin A., and Robert P. Harper, Jr., Fixed-Base and In-Flight Simulations of Longitudinal and Lateral-Directional Handling Qualities for Piloted Re-Entry Vehicles, ASD TDR 61-362, Feb. 1964.

49. Magdaleno, R. E., Neuromuscular Subsystem — Data and Models, Systems Technology, Inc., Working Paper 154-1, Oct. 1965.

50. Stapleford, Robert L., Donald E. Johnston, Gary L. Teper, and David H. Weir, Development of Satisfactory Lateral-Directional Handling Qualities in the Landing Approach, NASA CR-239, July 1965.

51. Taylor, L. W., and K. W. Iliff, Recent Research Directed Toward the Prediction of Lateral-Directional Handling Qualities, AGARD Paper No. 531, May 1966.

52. Ashkenas, I. L., and D. T. McRuer, The Determination of Lateral Handling Quality Requirements from Airframe-Human Pilot System Studies, WADC TR 59-135, June 1959.

53. Versace, J., Measurement of Ride Comfort, SAE Paper 638A, Jan. 1963.

54. Creer, Brent Y., Gordon H. Hardy, and Dallas G. Denery, "Flight Simulation and Pilot Describing Function Techniques Applied to the Analysis of a Pilot Control System for a Large Flexible Launch Vehicle," Symposium on the Human Operator in Aircraft and Spacecraft Control (Ad Hoc Panel on Guidance and Control) of AGARD, Paris, France, Sept. 5-6, 1966.

55. Hewes, Donald E., Interim Report on Flight Evaluations of Lunar Landing Vehicle Attitude Control Systems, AIAA Paper 67-239, Feb. 1967.

56. Matranga, Gene J., Donald L. Mallick, and Emil E. Kluever, An Assessment of Ground and Flight Simulators for the Examination of Manned Lunar Landing, AIAA Paper 67-238, Feb. 1967.

57. Gaul, John W., Application of Pilot-Controller Integration Techniques to a Representative V/STOL Aircraft, AFFDL TR 65-200, Oct. 1965.

58. McRuer, D. T., and I. L. Ashkenas, "The Role of Man in Control Systems," Handbook of Astronautical Engineering, H. H. Koelle, ed., McGraw-Hill Book Co., Inc., New York, 1961.

APPENDIX A

RATING SCALE SURVEY AND RESULTING RAW DATA

A questionnaire designed to determine the semantic values of 64 handling qualities descriptive phrases was received from 67 professionals in the piloting, engineering, and human factors fields. Of those received, four were discarded because of grossly incorrect interpretation of the experiment. The instructions, experience form, phrases, and the first page of the response sheets are given here. The responses were read off the axes to the nearest tenth of a division and tabulated. The tabulations follow in Table A-I and present the raw data used in the successive interval digital program.

A QUESTIONNAIRE TO EXPERIMENTALLY DETERMINE THE PSYCHOLOGICAL INTERVAL BETWEEN SOME PHRASES COMMONLY FOUND IN THE HANDLING QUALITIES LITERATURE

INSTRUCTIONS

The purpose of this questionnaire is to evaluate the meanings of some words and phrases commonly used in handling qualities literature, flight test reports, and pilot rating scales. The need to make such an evaluation has become apparent from inconsistencies and ambiguities in present scales. Hopefully, the results of this questionnaire will allow some modifications to existing scales which will vastly improve their utility.

At the beginning of the questionnaire is a list of the words and phrases which we hope to evaluate. The list is presented at the beginning so that you can familiarize yourself with the types of phrases, the spread that each type covers, and the way in which they interrelate with each other. You will notice that the phrases refer to characteristics such as controllability, sensitivity, etc., in varying degrees. Perhaps a good way to become familiar with them would be to look for the extremes of each characteristic (i.e., the "best" and the "worst") in the list. In any modified scale we will probably combine some of these phrases if they seem to have similar psychological weights, or degrees of goodness, to you.

Most of the words and phrases used are expected to be completely familiar. However, the use of the term "primary and/or secondary responses" needs some explanation. This phrase is intended to make you think of two kinds of responses — the first, the direct (and desired) result of control actions, e.g., roll to a specified bank angle; the second, the indirect motions which also occur, e.g., sideslipping and yawing. In the vertical plane a pertinent example is the "secondary" altitude or speed perturbations following a "primary" change in pitch attitude. Notice that "secondary" responses <u>are</u> desirable (e.g., airspeed change or turn rate) when they are of the proper form.

In the questionnaire itself the phrases are presented individually in a random manner alongside a vertical bar graph. Imagine that you are reading the phrase in a handling qualities or flight test report, and that the test pilot is describing a vehicle which he has tested. When you have formed an impression of the vehicle, document your impression by placing an "X" on the vertical line in the appropriate spot. If you feel that the phrase describes a vehicle with the best imaginable handling qualities, your "X" would belong at the very top of the line. Conversely, the worst imaginable handling qualities should be rated at the very bottom edge of the scale. The marks along the scale are intended only to help you precisely place your "X" on the vertical line. The scale should be considered continuous. To carry out the experiment:

1. Please fill out the experience form (page 2).
2. Study the list of phrases (page 3-5) long enough to become familiar with them (rereading the second paragraph above may help you).
3. Then reread this entire page so that the purposes and instructions are clear.
4. Then turn to the questionnaire (page 6) and start working through the phrases. Please work through them in order, and do not turn back to the pages listing the phrases.

RATING SCALE EXPERIENCE

Date: ________

Name: ____ Age: ________

Occupation: ____

Location: ____

____

Experience relevant to rating scales obtained as:

- ☐ Pilot
- ☐ Test Pilot
- ☐ Handling Qualities Engineer
- ☐ Psychologist
- ☐ Human Factors
- ☐ Other: ______________________________

If pilot, total hours (approximately).................... ________ hr

- Military fighters.................................. ______ hr
- Heavy aircraft (bombers, transports)................ ______ hr
- Light aircraft..................................... ______ hr
- Helicopters.. ______ hr
- Instrument... ______ hr
- Simulator.. ______ hr

Rating scales with which you are familiar:

- ☐ Cooper's (NASA)
- ☐ Cornell Aeronautical Laboratory
- ☐ Other: ______________________________

Approximate time spent evaluating with rating scale....... ________ hr

- Fixed-base simulator............................... ______ hr
- Moving-base simulator.............................. ______ hr
- Aircraft... ______ hr

WORDS AND PHRASES TO BE EVALUATED

1. Fair, somewhat impure primary or secondary response characteristics.
2. Excellent, pure (i.e., no "accidental" excitation) primary and secondary response characteristics.
3. Moderately sensitive, sluggish or uncomfortable in primary or secondary responses.
4. Barely controllable.
5. Easy to control with fair precision.
6. Major improvements are needed.
7. Highly desirable handling qualities.
8. Controllable with fair but somewhat inadequate precision.
9. Moderately objectionable deficiencies.
10. Very objectionable deficiencies.
11. Extremely easy to control with excellent precision.
12. Difficult to control.
13. Requires maximum available pilot skill and attention to retain control.
14. Some minor but annoying deficiencies.
15. Marginally controllable.
16. Completely demanding of pilot attention, skill or effort.
17. Excellent handling qualities.
18. Controllable, but only very imprecisely.
19. Extremely sensitive, sluggish or uncomfortable in primary or secondary response.
20. Effect of deficiencies on performance is easily compensated for by pilot.
21. Largely undemanding of pilot; relaxed.
22. Nearly uncontrollable.
23. Some mildly unpleasant characteristics.
24. Fair handling qualities.

25. Controllable with somewhat inadequate precision.

26. Improvement is requested.

27. Quite sensitive, sluggish or uncomfortable in primary or secondary responses.

28. Uncontrollable.

29. Very demanding of pilot attention, skill or effort.

30. Completely undemanding of pilot; very relaxed and comfortable.

31. Very difficult to control.

32. Pilot compensation required for acceptable performance in mission is too high.

33. Mildly demanding of pilot attention, skill or effort.

34. Controllable with definitely inadequate precision.

35. Very sensitive, sluggish or uncomfortable in primary or secondary responses.

36. Very bad handling qualities.

37. Good, relatively pure, primary and secondary response characteristics.

38. Requires substantial pilot skill and attention to retain control and continue mission.

39. Somewhat undesirably demanding of pilot attention, skill or effort.

40. Improvement is needed.

41. Good handling qualities.

42. Mildly sensitive, sluggish or uncomfortable.

43. Very easy to control with good precision.

44. Requires best available pilot compensation to achieve minimum acceptable performance.

45. Controllable with fair, but somewhat inadequate precision.

46. Much too sensitive, sluggish or uncomfortable in primary or secondary responses.

47. Good enough for mission.

48. Very poor handling qualities.

49. Pleasant handling qualities.

50. Controllable with difficulty.

51. Definitely sensitive, sluggish or uncomfortable in primary or secondary responses.

52. Extremely demanding of pilot attention, skill or effort.

53. Major deficiencies.

54. Controllable with somewhat inadequate precision.

55. Objectionable deficiencies.

56. Definitely demanding of pilot attention, skill or effort.

57. Bad handling qualities.

58. Quite demanding of pilot attention, skill or effort.

59. Reasonable performance requires considerable pilot compensation.

60. Mandatory improvement required.

61. Controllable with poor precision.

62. Very objectionable deficiencies.

63. Demanding of pilot attention, skill or effort.

64. Poor handling qualities.

Most Favorable

1. Fair, somewhat impure primary or secondary response characteristics.

Least Favorable

Most Favorable

2. Excellent, pure (i.e., no "accidental" excitation) primary and secondary response characteristics.

Least Favorable

Most Favorable

3. Moderately sensitive, sluggish or uncomfortable in primary or secondary responses.

Least Favorable

TABLE A-I

RAW SCORES OF QUESTIONNAIRE SURVEY

PHRASE	HANDLING QUALITIES ENGINEERS (HQE)																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	3.8	5.0	7.5	2.4	3.6	4.8	3.9	3.0	2.9	3.5	5.5	4.5	2.3	3.7	3.6	3.4	3.0	2.5
2	1.2	1.0	3.5	.9	.6	.1	.1	1.5	.1	.1	.5	.6	.1	.5	.1	.5	.5	.5
3	4.0	6.0	4.5	4.4	4.6	4.5	6.9	3.0	2.2	5.7	5.5	5.5	5.0	2.5	5.0	6.5	3.5	3.5
4	8.0	8.5	6.5	8.1	3.2	8.8	7.0	8.5	7.2	8.6	8.4	8.6	9.4	6.5	9.0	8.5	9.1	8.5
5	3.0	2.0	3.5	2.9	3.6	4.0	2.0	1.5	.9	2.7	2.5	1.6	3.5	1.8	3.0	2.5	2.0	2.5
6	7.5	6.8	5.5	5.0	6.6	9.0	7.9	6.1	4.1	5.4	6.4	5.6	6.1	5.7	5.0	7.5	5.0	4.6
7	.9	2.0	2.6	1.4	1.6	3.4	1.5	1.5	.3	1.7	1.5	.6	.4	.9	.6	1.4	.7	1.5
8	5.3	6.4	4.6	4.5	5.5	6.7	4.2	3.0	3.1	4.7	4.5	3.5	3.9	3.4	3.6	5.5	5.0	3.5
9	5.2	5.6	4.6	4.4	4.6	6.0	5.3	3.5	2.2	4.7	5.5	4.6	4.5	3.5	5.6	3.5	3.0	4.5
10	7.7	7.0	5.5	6.0	7.5	8.0	7.4	6.0	5.9	6.1	7.5	6.5	6.0	6.0	7.6	8.5	8.0	5.5
11	.01	.5	2.4	.8	1.6	2.0	.5	.7	.3	.5	1.5	.6	.1	.6	.01	1.5	.7	.5
12	6.0	8.0	5.5	4.8	5.6	7.0	6.1	6.5	6.8	5.6	5.6	6.6	8.0	5.0	5.0	6.6	6.0	4.5
13	9.5	9.5	9.5	7.8	6.7	8.8	8.2	8.5	9.6	7.7	7.4	8.6	9.1	9.4	8.9	6.9	9.0	7.5
14	3.5	5.5	4.5	3.0	3.6	4.1	5.4	3.0	1.8	3.6	4.5	2.6	2.5	2.8	2.8	3.5	4.0	3.6
15	9.0	8.5	6.3	6.7	8.6	8.9	7.7	8.0	9.4	7.5	6.5	5.5	7.4	7.5	8.6	7.6	8.5	7.5
16	8.4	8.5	7.5	5.0	6.6	8.9	8.8	8.5	9.2	7.7	7.5	7.6	7.6	9.4	9.0	5.5	8.8	6.5
17	.7	.4	2.5	1.8	.6	1.0	.2	1.6	.7	.7	.4	.6	1.0	.5	.01	.5	.5	.5
18	7.1	6.6	8.5	4.1	6.8	7.2	4.7	7.5	6.0	4.5	6.5	5.6	6.3	4.9	7.6	6.5	7.2	4.5
19	6.0	7.8	5.5	4.8	7.7	6.4	3.7	6.0	7.3	6.7	5.5	2.6	9.0	8.7	7.6	4.5	6.5	4.5
20	3.2	4.5	4.5	3.9	3.6	4.9	3.4	2.9	2.3	3.6	4.4	2.5	2.3	2.0	4.0	3.5	2.9	3.6
21	1.2	2.0	3.5	1.7	2.6	3.2	.4	1.5	.9	2.5	2.5	1.6	1.8	1.1	1.0	1.6	1.6	1.5
22	9.4	7.9	6.7	6.5	8.5	9.0	9.5	7.5	8.7	8.5	7.5	8.5	9.0	8.0	9.5	8.5	8.5	7.5
23	1.7	6.0	5.5	2.9	2.6	3.5	3.7	3.6	1.8	3.5	4.5	3.6	3.2	3.0	2.6	2.6	3.4	3.5
24	4.5	3.0	4.6	3.5	3.7	5.0	2.7	3.5	3.0	3.8	4.5	3.4	3.0	3.6	3.0	2.6	3.5	2.5
25	5.0	5.0	4.5	4.0	4.7	7.5	5.4	4.5	3.5	4.6	4.5	3.5	3.6	3.6	7.5	3.7	7.5	4.5
26	3.1	4.7	2.5	4.5	3.7	6.0	3.7	3.5	4.1	4.8	3.6	2.4	4.2	2.6	3.0	4.5	7.5	3.5
27	5.5	6.6	2.6	4.4	5.6	7.0	5.4	5.5	5.4	4.8	6.6	4.5	6.0	5.7	5.0	4.5	5.5	4.5
28	10.0	10.0	9.6	8.7	7.7	10.0	10.0	9.5	10.0	9.6	8.6	8.5	10.0	8.5	10.0	9.5	9.7	9.5
29	7.1	8.0	8.5	4.5	6.6	7.0	6.7	8.5	3.6	9.5	6.5	7.5	7.5	6.7	9.6	6.5	8.0	5.5
30	1.0	2.5	2.7	.1	1.8	2.5	.5	1.5	.1	1.6	1.5	.5	1.1	.5	1.0	1.5	1.6	.5
31	8.0	8.0	9.5	6.4	7.6	8.5	8.0	8.5	8.9	7.5	8.5	8.5	8.7	6.5	7.6	8.8	7.9	6.5
32	4.8	6.5	5.6	4.5	6.6	7.5	5.0	5.5	5.7	6.9	6.5	5.6	4.3	4.5	4.5	6.5	7.7	6.6
33	3.4	4.7	3.6	2.4	4.6	4.6	4.2	3.0	2.3	3.4	3.5	2.6	3.0	3.6	3.5	2.5	3.6	4.6
34	6.0	6.0	5.4	5.5	6.7	7.7	5.5	6.0	5.5	6.6	4.5	3.6	6.5	5.0	6.6	4.5	7.6	5.8
35	6.0	6.5	4.5	5.5	6.7	6.5	6.9	4.5	5.8	6.6	6.5	5.4	4.9	5.7	6.6	4.6	7.7	5.6
36	7.1	7.6	5.3	8.7	7.6	8.1	7.7	8.5	7.9	6.7	7.5	7.4	6.3	7.0	7.6	7.5	8.5	6.7
37	2.5	2.5	3.9	2.5	2.6	4.1	2.3	1.5	.7	2.5	2.6	1.6	1.7	1.5	2.5	1.6	1.6	1.5
38	6.3	6.6	5.5	4.5	4.7	6.2	7.5	4.5	6.1	6.0	6.5	4.5	6.0	4.5	7.3	5.0	7.0	5.6
39	4.0	5.7	5.5	4.0	3.6	6.9	5.7	3.5	3.7	5.5	4.5	4.6	5.9	4.5	3.6	4.6	8.9	4.5
40	4.0	6.0	0	3.8	4.7	7.0	5.5	3.6	4.3	6.5	4.5	2.5	8.5	3.7	5.0	4.4	5.1	3.5
41	2.5	2.4	1.5	2.1	2.8	3.8	1.8	1.5	1.6	2.6	2.5	2.5	2.1	1.3	2.6	2.5	1.6	1.5
42	3.5	5.5	5.7	3.8	4.7	6.4	3.9	3.0	3.5	3.8	4.5	3.6	4.2	3.0	3.6	3.6	2.7	3.5
43	1.0	1.5	2.5	1.7	2.7	4.4	1.5	1.5	.4	1.9	1.5	1.5	1.1	1.7	1.0	1.5	.9	1.6
44	7.0	6.0	6.5	4.5	5.6	8.5	6.5	5.9	3.3	8.0	6.6	5.5	6.3	5.0	6.6	4.8	9.0	5.5
45	4.5	5.5	4.4	3.9	4.8	6.5	4.9	4.6	3.3	4.5	5.6	3.5	3.9	5.4	4.6	3.5	5.9	4.0
46	8.0	7.5	5.4	5.0	6.7	7.5	7.5	6.0	6.1	6.9	7.6	5.4	6.1	7.6	6.6	5.5	8.0	4.5
47	3.3	5.0	2.5	4.0	3.6	2.8	4.4	3.0	5.0	3.5	2.6	4.5	2.8	3.1	4.6	2.5	6.7	2.5
48	6.5	7.6	6.5	7.9	5.6	8.4	7.3	6.5	7.9	6.6	8.5	6.6	6.9	7.6	6.6	4.4	8.0	5.5
49	2.0	2.4	2.5	2.0	2.6	4.3	2.8	1.5	1.0	2.7	3.4	2.5	2.6	1.3	1.6	2.5	2.6	1.6
50	6.8	7.0	7.6	5.5	7.7	7.2	5.5	6.1	4.4	6.5	5.5	5.5	6.6	5.0	6.7	5.5	8.0	4.8
51	5.0	6.9	4.5	4.5	5.7	6.6	6.8	6.7	4.5	4.9	5.5	4.8	4.5	5.0	5.5	5.6	6.0	4.5
52	7.1	6.9	5.5	5.5	8.6	7.8	8.0	7.2	4.2	7.7	7.5	7.6	7.0	6.6	6.5	5.8	8.6	5.5
53	7.5	8.5	6.6	7.7	7.7	8.4	8.2	7.2	7.0	7.5	6.5	5.5	7.0	6.6	5.4	6.5	7.0	5.5
54	4.0	5.6	3.5	4.9	4.7	6.7	4.8	4.6	3.7	5.4	4.6	3.6	5.2	4.0	4.6	4.6	5.6	4.5
55	4.5	6.5	4.5	5.0	4.7	5.8	5.5	3.6	4.2	5.9	6.5	3.4	4.0	4.4	4.6	6.5	7.6	5.5
56	6.5	7.0	5.6	5.0	5.8	6.4	7.0	6.0	3.7	4.9	4.5	4.6	5.1	5.1	4.7	5.4	7.6	5.6
57	6.5	8.0	6.5	6.5	6.7	7.8	7.2	5.5	6.9	7.6	7.5	5.5	8.3	7.5	5.6	7.4	7.0	6.5
58	7.0	7.6	6.3	4.6	5.7	6.7	5.5	4.5	4.6	7.4	6.5	7.5	6.4	5.1	5.5	5.5	6.7	5.5
59	5.6	6.4	5.5	4.1	5.7	7.7	4.4	4.5	4.8	5.4	4.6	3.6	4.8	4.7	4.7	4.5	5.1	5.5
60	8.0	8.5	7.5	7.5	6.7	8.5	8.5	4.6	6.4	6.5	6.5	5.6	6.0	5.0	4.7	5.6	8.0	6.5
61	4.1	6.0	4.5	5.1	5.8	7.6	7.2	6.0	4.4	6.0	6.5	4.6	6.1	3.6	5.6	5.6	7.6	5.5
62	7.0	8.6	5.6	5.0	6.9	7.9	4.7	5.5	6.6	6.5	3.5	6.6	7.0	7.0	6.4	6.5	7.8	5.5
63	5.0	5.5	4.5	4.5	4.7	7.8	5.1	5.5	2.4	4.8	5.5	3.6	5.9	4.9	5.5	4.6	5.2	3.6
64	5.5	6.9	4.7	6.9	5.7	8.0	5.5	4.5	3.5	5.9	7.5	5.5	7.6	4.4	5.6	6.5	6.5	3.5

TABLE A-I (Continued)

	HANDLING QUALITIES ENGINEER (HQE)					
PHRASE	19	20	21	22	23*	24
1	2.5	2.5	3.0	3.6	6.1	4.6
2	1.5	1.6	1.0	.8	3.0	.6
3	3.5	5.6	4.0	4.0	6.2	3.7
4	7.5	8.6	9.0	6.4	8.0	8.0
5	2.5	3.7	3.1	2.5	4.5	3.0
6	6.5	7.5	6.0	5.9	8.0	7.0
7	1.5	1.5	2.1	1.4	1.8	1.5
8	3.5	5.6	7.1	3.6	6.7	4.1
9	3.5	4.6	5.1	4.6	7.5	5.3
10	6.5	8.5	6.1	5.6	9.0	8.3
11	.5	1.5	.01	1.1	2.5	.4
12	5.9	7.6	8.0	6.1	7.2	6.4
13	7.5	8.5	9.0	6.6	9.5	9.2
14	3.6	2.7	4.0	3.1	5.0	5.9
15	6.5	8.6	8.9	6.5	8.0	7.5
16	8.5	7.5	6.8	6.6	9.0	8.8
17	.5	.5	.9	1.0	1.0	.7
18	4.5	7.5	8.9	5.0	7.0	6.3
19	5.6	8.5	6.9	5.7	7.5	7.4
20	2.5	3.6	3.8	5.0	3.0	3.8
21	1.4	2.6	3.1	2.2	3.0	2.7
22	6.5	8.5	9.0	6.5	9.5	9.2
23	3.0	3.6	2.9	3.0	4.5	4.2
24	2.5	3.6	2.8	2.6	5.0	3.9
25	4.6	7.7	6.9	3.6	5.2	5.0
26	4.4	4.5	5.0	6.0	5.5	5.1
27	4.4	2.6	4.9	6.1	5.7	5.8
28	9.5	9.5	10.0	8.6	10.0	9.6
29	6.5	7.5	6.0	5.0	7.8	7.3
30	.5	1.6	2.0	2.2	6.5	1.4
31	6.5	7.7	8.9	6.5	8.5	6.9
32	6.5	4.6	6.9	6.0	6.0	5.3
33	3.5	2.6	3.0	3.4	4.5	3.7
34	6.5	6.6	8.0	4.6	7.5	4.7
35	6.5	7.5	7.0	5.5	7.8	7.1
36	6.5	5.6	7.0	8.0	8.5	7.7
37	1.4	2.7	2.0	3.0	2.5	3.6
38	5.5	6.5	6.0	4.9	8.2	6.0
39	4.5	4.6	5.9	4.4	6.5	5.2
40	4.5	5.7	5.0	5.9	6.0	6.3
41	1.5	2.5	2.0	2.5	2.8	3.6
42	3.5	5.6	3.0	4.0	5.8	5.0
43	1.4	1.6	2.0	1.4	3.7	2.2
44	5.5	7.5	6.0	5.0	8.5	7.9
45	4.5	5.6	6.9	4.1	6.2	4.1
46	5.5	7.5	6.9	0	7.8	7.1
47	4.5	4.7	4.1	3.6	3.0	5.1
48	6.5	8.5	6.9	6.2	8.5	6.1
49	1.5	2.5	2.9	2.6	2.8	1.8
50	5.6	5.5	9.0	6.1	8.5	5.5
51	4.5	6.6	5.0	5.1	6.5	5.7
52	6.5	7.7	9.0	6.0	8.0	8.8
53	6.5	7.5	7.0	6.6	8.0	7.5
54	4.5	4.7	6.0	4.5	6.2	4.9
55	4.5	6.7	4.7	5.5	7.8	5.6
56	3.5	5.4	4.8	6.0	8.0	5.4
57	4.5	5.6	5.5	7.1	8.0	6.5
58	3.5	6.6	4.9	5.0	7.5	5.1
59	3.6	6.5	5.0	5	6.0	5.3
60	6.6	7.5	7.9	7.1	8.0	7.1
61	4.5	5.6	8.0	3.8	8.5	5.6
62	5.5	8.5	6.0	6.6	8.0	7.6
63	3.5	6.6	5.0	5.3	7.0	5.2
64	4.5	7.5	5.1	6.0	8.2	6.0

* NOT INCLUDED IN ANY CALCULATIONS BECAUSE OF FAMILIARITY WITH SURVEY DESIGN

TABLE A-I (Continued)

PILOTS (PI)

PHRASE	1	2	3	4	5
1	3.7	2.0	2.0	5.6	2.9
2	.8	.1	.1	.6	.7
3	2.9	4.0	4.0	4.5	4.6
4	9.8	8.9	9.9	9.0	8.6
5	3.8	3.9	2.0	3.2	2.4
6	8.8	7.9	9.0	9.1	5.6
7	1.8	1.0	1.0	1.3	1.0
8	6.6	7.1	5.1	4.2	4.5
9	7.6	5.9	7.1	6.4	4.6
10	8.7	8.4	9.0	8.5	7.6
11	.8	.2	0	.1	.6
12	8.7	8.8	9.1	7.5	7.6
13	9.8	8.0	8.9	9.3	8.6
14	5.7	4.9	5.0	4.6	3.7
15	8.8	5.9	9.9	7.7	8.6
16	9.7	7.5	9.1	8.2	6.5
17	2.7	.9	.1	.7	.7
18	5.7	6.5	7.0	4.8	7.5
19	8.8	6.9	7.0	6.3	6.6
20	5.5	4.9	7.0	2.8	3.5
21	2.7	2.0	2.0	1.2	1.6
22	8.7	9.4	9.5	9.7	8.6
23	5.8	5.6	6.0	6.3	3.0
24	4.7	5.9	7.0	4.5	2.7
25	6.6	5.5	6.0	5.3	4.7
26	4.9	5.0	6.1	6.4	3.7
27	6.8	5.6	7.1	7.3	6.0
28	9.8	9.9	10.0	10.0	9.5
29	8.8	7.0	8.9	8.6	7.6
30	1.7	.9	1.0	2.2	.7
31	8.6	8.7	9.2	8.4	7.7
32	9.8	7.0	8.9	8.2	6.5
33	5.8	5.0	2.0	5.3	3.5
34	8.6	5.9	7.0	7.4	7.5
35	8.8	5.9	7.0	7.7	7.5
36	8.7	9.0	10.0	8.4	8.5
37	1.6	2.1	1.1	2.3	2.0
38	8.9	5.0	9.1	8.8	5.7
39	7.8	3.1	7.0	5.5	8.6
40	5.8	4.0	5.9	5.2	5.6
41	3.6	2.0	.9	2.4	1.7
42	5.8	4.8	5.9	7.9	3.9
43	2.6	.8	0	1.7	1.6
44	9.7	8.3	9.9	9.0	5.6
45	4.8	5.9	7.0	4.5	3.5
46	8.5	7.7	7.8	8.1	7.6
47	6.7	4.0	5.0	5.0	3.6
48	8.8	7.8	9.2	8.7	8.1
49	3.7	1.9	2.9	2.2	1.9
50	8.9	7.1	8.0	6.0	8.0
51	8.7	7.0	7.0	6.1	4.7
52	9.7	7.0	9.1	6.5	8.5
53	8.8	7.9	9.1	9.0	6.5
54	6.9	5.5	7.0	6.7	3.5
55	9.7	6.9	9.1	8.0	6.5
56	8.8	5.8	7.9	7.9	7.6
57	9.7	9.0	8.9	8.0	7.5
58	8.6	5.4	9.9	8.2	6.6
59	8.6	6.0	7.7	7.6	6.6
60	9.7	8.0	9.9	9.4	6.6
61	4.7	5.9	6.9	6.5	6.6
62	8.8	7.6	8.0	4.7	9.7
63	7.6	5.5	8.0	5.8	5.5
64	6.6	9.0	7.9	7.6	6.7

TABLE A-I (Continued)

TEST PILOTS (TP1)																
PHRASE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	4.0	3.4	5.0	3.0	5.5	3.0	2.0	5.7	3.5	3.7	6.0	5.6	7.6	8.0	3.6	3.5
2	.01	.9	1.0	1.1	.5	1.0	.5	1.6	.01	.6	.01	.1	.6	.1	1.5	.3
3	6.1	4.0	6.1	5.0	6.5	4.0	.7	4.7	4.0	4.7	2.0	6.9	5.6	6.0	8.5	4.5
4	9.6	9.4	10.0	7.9	8.6	7.1	8.0	8.6	6.3	8.5	9.0	9.0	8.5	9.0	8.5	8.7
5	2.5	2.7	3.0	3.0	4.5	2.0	.8	3.6	2.9	2.5	2.0	2.9	2.5	4.0	2.5	2.4
6	6.9	5.0	7.2	7.0	5.5	5.0	5.0	6.7	0	6.4	9.0	8.0	6.5	6.0	5.6	6.9
7	1.0	.9	.2	2.1	1.5	1.7	.7	.6	0	1.3	1.0	1.0	1.5	1.0	1.7	.5
8	5.7	4.0	3.0	4.1	5.4	4.5	4.5	3.6	4.2	4.4	7.0	6.0	4.5	5.0	4.7	4.6
9	6.1	5.0	5.0	5.0	4.5	5.0	4.5	4.6	3.5	4.5	8.0	2.6	6.5	6.0	3.6	4.9
10	9.0	7.0	8.0	8.0	7.5	7.6	7.0	7.6	0	6.4	9.0	9.3	8.6	7.0	6.6	5.5
11	.6	1.0	.9	.9	.5	.7	.2	.7	.01	1.0	.1	.01	.5	1.1	1.6	.5
12	9.0	4.5	8.7	7.0	8.6	6.1	7.0	7.5	4.7	4.5	8.0	2.6	8.6	6.0	4.7	5.6
13	9.0	8.0	10.0	7.0	8.6	8.0	8.0	8.5	5.2	8.7	9.0	10	9.6	9.0	8.5	8.5
14	4.0	4.0	5.9	3.0	3.6	5.0	3.5	3.7	2.9	3.4	6.0	5.9	6.5	3.0	3.5	3.4
15	9.0	8.0	7.0	8.0	8.6	6.7	8.7	7.5	6.2	7.6	9.4	8.1	6.5	8.0	7.5	8.4
16	10.0	8.1	9.0	9.8	7.4	7.0	8.0	9.1	0	7.8	7.7	8.2	8.5	8.0	8.7	7.6
17	.5	1.0	.1	.9	.7	1.0	1.0	.6	.01	.3	1.9	.01	.6	1.0	1.5	.5
18	6.9	6.0	6.0	2.0	5.5	5.0	5.5	4.6	6.3	5.9	6.0	6.9	5.7	5.0	5.4	5.5
19	8.7	4.0	5.9	7.0	6.5	6.1	6.0	7.6	0	5.2	9.7	9.3	8.5	7.0	4.6	5.6
20	3.9	3.5	3.0	4.0	3.6	3.1	2.0	3.5	3.0	4.0	5.0	5.8	4.5	3.0	3.6	2.5
21	1.0	2.0	2.0	1.6	1.6	1.7	1.2	1.5	1.0	1.4	2.0	3.0	2.5	3.0	2.7	1.5
22	9.6	8.7	8.1	8.0	8.5	6.9	8.2	8.6	6.9	8.6	7.2	9.3	8.5	9.0	7.7	8.5
23	3.1	3.0	5.1	2.0	4.5	3.0	3.4	5.4	3.0	2.9	2.5	5.9	5.5	5.0	4.6	2.6
24	2.1	4.1	4.0	3.0	4.5	3.5	1.5	4.6	2.9	3.2	5.1	5.0	3.5	3.6	3.5	2.5
25	6.0	3.5	5.0	6.0	7.6	5.0	4.5	5.6	0	5.8	7.5	5.9	4.6	4.0	6.5	4.5
26	4.9	3.9	5.9	7.0	5.5	3.3	3.5	5.6	0	4.2	6.0	5.0	4.5	5.0	5.6	3.5
27	6.9	4.5	6.9	6.0	4.5	5.4	5.0	6.6	4.5	6.5	8.6	9.7	6.5	6.0	4.7	4.5
28	10.0	9.4	10.0	10.0	9.6	9.0	10.0	9.6	9.0	9.5	10.0	10.0	9.9	10.0	9.7	9.5
29	9.0	7.0	9.1	7.0	7.4	6.6	6.5	7.6	5.0	4.7	2.0	2.1	8.6	5.0	0	4.5
30	.9	2.0	3.0	.9	1.5	.1	7.3	.5	.1	.9	1.0	1.0	3.5	3.0	0	.5
31	9.5	7.0	9.0	7.0	8.6	6.9	8.0	6.6	6.0	7.3	9.0	7.5	9.5	8.0	6.6	7.5
32	8.5	5.0	9.0	7.0	6.7	3.6	3.5	8.7	7.0	6.4	7.0	6.8	8.4	6.0	3.5	4.5
33	5.1	4.0	4.0	2.0	2.6	3.1	2.6	4.6	4.0	3.5	4.0	5.8	4.5	4.0	3.6	3.5
34	6.5	5.0	6.0	7.0	4.5	5.7	5.0	6.6	0	6.4	7.5	6.6	5.6	6.0	5.6	6.5
35	7.0	4.0	8.0	8.0	6.5	6.2	5.0	7.7	0	4.7	8.7	7.8	4.5	7.0	8.6	4.5
36	9.0	6.0	9.0	6.0	7.4	7.1	7.0	8.6	5.0	7.1	9.3	8.7	4.5	8.0	6.5	6.5
37	1.5	2.0	5.0	2.1	1.6	1.5	1.5	2.8	1.0	1.7	2.0	1.4	3.5	2.0	3.7	1.5
38	6.1	6.0	8.0	7.0	2.6	4.0	6.0	6.7	0	4.7	7.9	8.0	7.5	6.0	5.5	7.5
39	7.5	5.0	6.0	3.1	5.1	5.0	5.0	6.5	0	3.5	6.0	6.2	5.6	5.0	4.6	4.9
40	6.5	4.0	6.0	3.0	4.5	0	5.0	5.7	0	5.4	6.0	9.0	4.5	4.0	5.6	5.6
41	2.0	2.0	1.0	1.0	1.6	0	1.2	2.7	1.7	1.7	4.0	1.7	3.5	2.0	2.5	1.4
42	6.0	5.4	6.0	3.0	3.6	0	2.6	5.5	0	3.5	6.0	5.4	4.5	4.0	4.5	3.4
43	1.5	1.4	1.0	.9	.5	1.9	.5	2.8	0	1.7	.6	1.0	1.6	2.0	1.5	1.4
44	5.9	7.0	6.0	5.9	8.4	6.5	6.0	7.7	0	6.5	10.0	8.4	8.5	5.0	6.5	5.5
45	5.4	4.1	4.0	3.0	5.4	4.0	3.3	5.7	0	4.6	6.9	5.4	5.6	5.0	4.5	4.5
46	8.0	4.4	7.0	7.0	5.5	8.6	6.3	7.7	6.0	9.5	9.0	9.4	8.5	9.0	4.7	5.9
47	5.0	4.0	6.0	3.0	2.5	8.4	3.0	3.6	4.9	5.4	4.9	4.1	2.5	4.0	4.4	2.4
48	9.1	6.5	9.1	4.9	7.7	5.9	5.5	7.6	4.6	7.7	7.4	8.7	8.5	8.0	6.6	6.6
49	1.5	2.5	2.1	1.1	1.6	1.5	2.0	2.7	1.6	1.8	2.5	2.7	3.5	3.0	2.7	1.5
50	9.6	6.0	6.0	5.0	6.6	5.3	7.0	7.6	6.0	6.9	3.5	6.1	6.5	8.0	3.6	9.2
51	9.5	4.0	7.9	6.0	6.5	5.2	4.0	7.6	0	5.0	7.0	7.1	5.6	6.0	4.6	5.5
52	9.7	6.5	8.9	8.0	6.5	5.9	7.6	7.5	6.0	6.0	10.0	7.9	8.9	9.5	6.5	6.5
53	7.5	5.1	6.3	5.1	7.5	5.2	7.0	8.7	0	6.5	10.0	9.0	7.5	6.1	6.4	6.5
54	5.5	3.5	5.0	2.9	4.6	4.4	4.5	6.6	3	4.5	6.0	5.9	6.5	5.0	4.5	5.5
55	6.0	5.1	6.1	7.2	5.6	4.6	4.0	6.6	3.8	3.7	2.0	7.5	7.5	6.0	5.5	5.5
56	6.3	5.5	7.0	5.0	4.7	4.3	5.0	7.6	5.2	4.6	7.0	7.7	7.5	6.0	4.6	5.6
57	7.0	5.5	10.0	7.0	6.5	6.0	7.0	9.5	3.7	7.6	8.0	9.0	8.5	7.0	6.7	6.8
58	6.5	5.1	8.0	9.0	4.5	5.4	6.0	6.6	7.6	5.0	10	6.7	7.5	6.0	4.7	4.7
59	6.0	4.5	8.0	9.0	5.6	4.0	5.0	6.5	4.2	4.8	8.1	5.6	8.5	4.0	4.6	4.5
60	9.0	5.6	8.1	6.0	7.5	5.1	7.5	5.7	0	6.5	9.0	8.1	7.5	6.4	6.7	6.5
61	6.5	4.6	5.1	7.0	6.5	4.2	6.5	7.6	5.9	2.4	7.0	2.5	6.5	6.0	5.9	5.5
62	9.6	5.5	7.0	8.0	7.5	6.4	7.0	6.6	0	4.5	7.9	8.2	7.5	7.0	7.5	5.5
63	7.0	5.0	7.0	6.0	3.5	4.5	6.0	6.4	0	5.1	5.7	5.4	6.5	5.0	6.5	4.6
64	7.0	6.6	7.0	6.0	4.5	5.0	4.0	7.7	3.8	4.8	7.2	7.5	7.5	5.0	0	6.5

TABLE A-I (Continued)

	TEST PILOTS (TPI)											
PHRASE	17	18	19	20	21	22	23	24	25	26	27	28
1	3.1	7.1	6.0	6.0	5.6	7.0	3.0	3.5	6.0	6.0	8.0	6.5
2	.01	1.2	.6	.01	.6	.5	.4	.5	1.0	6.0	.3	5.4
3	6.0	5.8	5.7	8.2	8.6	6.7	4.0	4.5	2.1	7.0	5.0	6.6
4	9.0	9.0	8.7	9.1	9.5	9.6	8.8	8.4	9.2	9.0	9.2	9.0
5	3.0	4.9	3.4	2.0	2.8	5.4	1.7	2.7	2.5	6.1	2.4	5.1
6	6.0	6.6	8.0	10.0	8.6	9.7	7.4	4.7	7.6	8.0	6.4	8.0
7	.01	.5	1.0	.4	.4	2.0	.3	.5	1.1	1.0	.01	3.0
8	7.0	7.0	5.3	5.1	6.6	4.7	6.4	4.2	5.2	7.0	5.1	7.0
9	5.1	6.4	5.0	6.1	6.5	6.5	5.5	4.8	6.0	8.0	5.0	8.0
10	7.0	8.7	9.0	10.0	8.6	8.7	7.6	5.5	9.1	9.0	8.0	9.0
11	.01	.5	.01	.01	.6	.6	.6	.5	1.1	.5	1.0	3.0
12	8.1	8.0	5.9	9.0	7.4	9.5	7.5	5.8	7.0	9.1	8.0	8.6
13	9.1	9.4	8.0	9.0	9.5	9.3	9.0	8.2	8.4	10.0	9.0	9.1
14	4.0	6.8	4.8	4.0	4.6	5.6	4.0	2.8	5.1	6.0	2.7	5.7
15	3.0	9.2	7.0	9.0	9.5	7.5	8.4	8.3	9.1	10.0	3.0	8.5
16	10.0	8.6	9.0	9.0	9.5	9.4	9.2	5.3	8.5	10.0	9.0	9.0
17	.01	.3	.6	.01	.5	.6	.4	.4	.2	1.0	1.0	2.8
18	8.0	7.5	5.6	6.0	7.6	6.4	6.8	5.3	7.5	8.0	7.9	7.6
19	8.1	7.8	7.0	9.0	7.5	7.5	7.2	5.6	9.1	8.0	8.0	7.8
20	4.0	5.2	4.1	4.0	2.7	5.0	4.3	2.5	4.5	5.1	4.0	4.0
21	1.1	4.2	2.0	2.1	3.5	1.5	1.7	1.4	3.0	4.0	3.0	3.0
22	9.0	9.5	9.2	9.0	9.5	7.5	9.4	5.4	9.5	9.9	9.2	9.1
23	3.0	6.5	4.7	7.2	5.6	5.5	3.1	2.4	6.0	6.0	3.0	6.5
24	2.5	5.9	3.4	5.1	4.7	3.5	3.3	3.6	5.0	6.0	4.0	5.8
25	7.0	7.4	5.6	6.1	6.5	5.0	5.1	4.3	6.7	6.0	5.5	5.8
26	5.1	6.5	5.6	7.3	8.6	7.4	4.9	3.7	5.5	7.0	5.0	5.1
27	6.0	8.4	7.7	8.0	5.6	7.5	5.2	4.8	8.0	6.9	7.0	6.2
28	10.0	9.8	10.0	10.0	9.7	9.5	10.0	9.6	10.0	10.0	10.0	10.0
29	7.1	8.6	9.3	9.1	9.6	7.5	9.0	7.8	8.0	9.6	7.0	8.0
30	.1	2.2	.3	2.0	1.5	1.4	2.0	.5	.9	.1	1.0	2.5
31	8.0	8.7	7.0	9.1	8.5	8.5	8.5	8.1	8.7	9.5	9.0	7.8
32	7.2	7.6	6.8	8.4	7.5	7.5	6.8	5.7	8.1	8.0	6.5	6.8
33	4.9	5.9	4.3	4.2	3.5	6.5	5.1	2.3	5.5	6.5	2.9	5.0
34	7.0	7.9	5.8	8.0	8.6	6.5	6.2	5.2	8.5	8.6	8.0	6.9
35	6.0	8.8	6.7	9.0	5.6	7.5	6.4	5.5	8.6	8.0	8.0	8.1
36	8.2	9.0	9.8	10.0	8.5	9.5	8.1	5.0	9.1	9.0	9.9	9.8
37	2.0	4.4	1.7	2.0	2.5	2.6	2.4	1.5	2.2	3.0	1.9	4.3
38	8.0	8.0	8.7	9.1	9.6	8.4	9.7	6.2	7.0	9.6	9.5	7.9
39	6.0	6.4	6.0	9.2	6.6	7.5	6.3	4.5	6.1	6.0	5.9	7.0
40	5.1	5.4	7.0	8.0	8.6	7.5	4.5	4.3	6.1	6.0	4.6	5.8
41	2.0	3.5	1.0	1.9	2.5	2.5	2.0	1.5	2.0	3.0	2.5	4.0
42	5.0	5.8	5.3	4.0	6.5	6.5	4.1	3.6	5.8	6.0	2.5	6.0
43	1.0	2.7	.2	1.4	.5	.6	2.7	.5	2.0	2.4	1.6	3.0
44	6.0	8.1	9.2	9.2	9.6	8.5	7.9	5.0	8.5	8.4	7.0	8.1
45	7.0	5.9	5.7	7.0	8.5	6.4	6.2	4.2	6.5	6.1	5.0	6.0
46	8.0	7.3	9.6	8.0	8.5	8.4	7.9	5.0	8.0	8.0	6.0	7.1
47	4.0	5.1	3.5	5.0	7.4	5.5	4.6	2.6	5.0	5.0	4.6	5.6
48	7.0	8.9	9.3	9.2	8.5	9.5	8.9	5.3	8.6	8.5	7.0	8.3
49	2.0	4.0	2.0	1.1	1.6	2.5	2.4	1.5	4.1	4.0	1.3	3.6
50	9.0	8.1	8.4	8.0	9.6	8.3	8.4	6.4	9.4	10.0	7.0	7.0
51	6.0	6.8	9.0	7.2	5.6	7.4	6.7	3.5	6.1	8.0	7.0	7.0
52	7.1	7.1	9.3	7.0	9.5	8.5	6.9	5.9	8.0	10.0	8.5	8.2
53	7.0	8.1	9.0	9.2	9.5	9.6	6.8	4.6	8.0	9.0	6.9	8.1
54	7.1	6.1	6.0	5.3	7.5	3.5	6.7	4.0	7.5	7.0	5.0	6.6
55	5.1	7.6	7.8	9.0	8.5	8.7	6.5	6.4	6.6	6.0	5.8	6.2
56	6.0	8.0	9.0	8.1	8.6	7.6	6.0	5.3	7.0	7.0	5.9	7.5
57	7.0	8.7	9.8	10.0	8.5	9.5	6.0	4.5	8.5	8.1	9.0	8.3
58	6.0	7.0	7.3	9.0	8.5	7.5	7.1	5.6	6.6	8.0	7.9	7.4
59	5.1	8.0	6.0	6.0	8.5	7.7	6.0	5.6	5.2	6.5	7.0	6.3
60	7.1	8.5	8.3	10.0	9.5	9.6	6.4	5.7	7.2	8.0	6.0	7.8
61	6.1	2.4	6.8	9.0	8.6	7.4	6.3	4.5	7.8	8.1	5.0	7.0
62	6.0	9.8	9.2	10.0	8.5	8.7	7.1	5.8	8.6	9.6	7.0	8.1
63	6.0	7.0	6.3	8.2	7.5	9.4	5.5	5.2	6.5	7.5	5.0	6.0
64	5.0	8.4	9.6	9.9	8.6	9.6	8.6	5.0	8.0	7.6	7.9	8.1

TABLE A-I (Continued)

PSYCHOLOGIST (PSY)

PHRASE	1	2	3	4
1	6.5	5.6	2.8	6.3
2	1.0	.6	.2	2.3
3	5.7	3.8	4.0	6.7
4	9.0	7.6	9.0	8.2
5	3.0	1.7	2.0	3.7
6	8.2	8.6	9.4	6.3
7	2.0	1.7	.7	1.0
8	6.0	3.6	4.4	4.8
9	6.9	6.6	6.6	5.0
10	9.2	8.6	9.9	7.8
11	0	.7	.1	1.6
12	7.2	6.5	6.6	6.5
13	9.1	8.5	9.6	6.4
14	2.1	1.6	5.4	4.3
15	6.2	4.6	9.0	8.3
16	9.0	8.7	9.2	6.7
17	.2	.1	1.0	.2
18	7.0	7.5	9.1	5.8
19	7.0	0	8.4	6.8
20	6.1	5.6	2.2	4.2
21	4.0	2.5	.9	1.8
22	9.3	8.6	9.5	9.3
23	7.1	6.6	6.0	5.7
24	4.0	4.6	2.6	6.2
25	6.0	6.4	4.0	4.6
26	6.1	5.7	2.1	6.2
27	6.9	0	6.0	6.9
28	10.0	9.6	10.0	10.0
29	8.2	2.5	7.0	4.8
30	2.0	1.6	.01	1.6
31	9.2	8.6	8.7	8.9
32	7.0	8.6	9.6	4.7
33	6.0	3.6	5.0	2.7
34	8.0	8.6	7.8	5.3
35	7.0	0	9.0	6.2
36	9.1	8.6	8.9	9.2
37	2.2	1.7	1.4	2.7
38	8.0	7.6	6.8	5.5
39	7.0	5.7	7.4	3.6
40	6.3	7.7	8.0	5.7
41	3.0	2.8	3.1	2.5
42	6.2	4.5	6.0	2.9
43	1.3	1.7	.7	1.6
44	8.0	8.8	9.8	5.9
45	6.0	5.7	7.3	5.0
46	9.1	8.5	9.2	6.4
47	5.0	6.6	4.1	4.0
48	9.0	8.6	9.0	8.3
49	5.2	1.6	1.1	1.6
50	6.1	8.6	3.8	5.7
51	7.2	0	7.3	4.7
52	6.3	7.5	9.7	3.6
53	8.2	9.3	7.9	6.3
54	5.2	4.7	6.2	3.6
55	7.0	8.5	8.2	8.0
56	3.5	7.7	8.0	5.6
57	7.0	9.5	9.0	8.5
58	7.8	6.5	9.2	4.7
59	4.6	8.5	7.3	4.2
60	7.1	5.7	9.8	8.2
61	9.1	8.6	9.2	5.6
62	8.2	9.6	8.3	7.6
63	3.2	6.6	5.0	3.5
64	7.5	8.7	9.0	5.0

TABLE A-I (Concluded)

HUMAN FACTORS (HE)

PHRASE	1	2	3
1	3.5	6.0	3.5
2	1.3	1.1	1.5
3	5.0	5.9	4.5
4	2.1	9.5	8.6
5	4.1	5.0	2.6
6	6.0	7.9	5.9
7	2.0	2.5	2.7
8	5.0	5.0	3.8
9	6.1	2.0	5.7
10	7.2	8.2	8.0
11	7.0	1.9	1.5
12	6.8	8.0	7.5
13	8.0	9.6	8.7
14	5.0	6.1	6.6
15	8.0	8.1	7.7
16	7.0	7.8	7.6
17	2.4	1.0	1.6
18	5.1	7.5	3.6
19	6.1	7.9	4.9
20	4.1	3.9	2.7
21	1.5	3.9	1.6
22	8.3	9.6	9.8
23	3.0	6.5	6.7
24	3.1	5.1	7.5
25	4.5	5.3	3.5
26	4.6	5.9	6.5
27	3.9	6.1	3.7
28	9.0	8.5	9.6
29	6.9	8.8	2.7
30	7.0	4.0	1.5
31	8.1	2.0	7.5
32	5.2	6.9	7.9
33	3.1	5.5	6.5
34	5.4	6.0	6.3
35	4.1	9.9	2.7
36	6.0	8.6	2.4
37	2.1	3.0	2.5
38	5.2	9.0	7.6
39	5.1	5.6	7.5
40	4.0	6.0	9.5
41	2.0	2.5	1.5
42	3.3	6.5	4.4
43	2.1	2.5	1.7
44	4.5	8.6	8.4
45	3.9	6.0	6.7
46	3.5	8.5	9.5
47	4.0	5.0	2.7
48	5.0	7.5	2.3
49	2.1	4.5	7.5
50	7.2	6.0	2.5
51	4.0	6.0	2.4
52	7.2	8.0	8.6
53	7.0	7.9	1.9
54	5.5	5.5	6.3
55	5.2	7.5	7.6
56	6.5	5.5	7.6
57	6.0	2.5	9.5
58	5.1	8.6	7.5
59	4.5	7.9	6.5
60	6.7	9.6	[illegible]
61	3.0	6.9	[illegible]
62	6.1	8.8	2.6
63	5.4	9.0	7.5
64	5.8	7.0	8.6

APPENDIX B

REDUCED SURVEY DATA

The raw data given in Appendix A was processed in several ways to obtain the desired relationships. The results are given in Table B-I, where the column numbers correspond to the following calculations:

①, ② The grand means and variances were computed for all of the items. The items were then rank-ordered by mean.

③, ④ The scores for each rater were transformed as described in Section III-C-3b. The grand means and variances were then computed for the transformed scores.

⑤, ⑥ The scale values and discriminal dispersions were computed for the 63 items (No. 28, uncontrollable, was not included for reasons noted in Section III) with a digital program [Cumrey (17)].

⑦ The scale values were recomputed after the high variability items were excluded, leaving 31 items to be scaled. The retained items are given in Table B-II.

TABLE B-I

SUMMARY OF RESULTS OF PROCESSING THE QUESTIONNAIRE DATA OF APPENDIX A

ITEM	PHRASE	OBSERVED ITEMS		TRANSFORMED ITEMS		COMPLETE MATRIX (63) x (63)		PARTIAL MATRIX (31) x (63)
		MEAN ① ($\bar{R}_O$)	VARIANCE ② ($\sigma^2_{R_O}$)	MEAN ③ ($\bar{R}_T$)	VARIANCE ④ ($\sigma^2_{R_T}$)	ψ-SCALE VALUE ⑤ (m_i)	DISCRIMINAL DISPERSION ⑥ (s_i)	ψ-SCALE VALUE ⑦ (m_i)
11	Extremely easy to control with excellent precision	0.79	0.53	0.97	0.44	0.74	1.76	0.74
2	Excellent, pure (i.e., no accidental excitation) primary and secondary response characteristics	0.81	0.54	0.99	0.49	0.94	1.74	0.93
17	Excellent handling qualities	0.80	0.47	1.00	0.00	1.00	1.56	1.00
7	Highly desirable handling qualities	1.28	0.63	1.47	0.45	2.25	1.11	2.16
30	Completely undemanding of pilots; very relaxed and comfortable	1.45	1.04	1.65	0.94	2.36	1.46	2.26
43	Very easy to control with good precision	1.57	0.76	1.76	0.63	2.59	1.16	2.48
21	Largely undemanding of pilots; relaxed	2.15	1.13	2.36	0.98	3.56	1.13	3.40
37	Good, relatively pure, primary and secondary response characteristics	2.29	0.78	2.47	0.88	3.78	1.04	3.60
41	Good handling qualities	2.35	1.41	2.58	1.51	3.70	0.92	3.53
49	Pleasant handling qualities	2.43	1.37	2.65	1.42	3.71	1.12	3.52
5	Easy to control with fair precision	3.00	1.22	3.21	1.15	4.35	1.20	4.15
20	Effect of deficiencies on performance is easily compensated for by pilot	3.85	1.12	4.04	1.33	5.31	0.95	5.06
24	Fair handling qualities	3.91	1.52	4.13	1.39	5.34	0.96	5.09
33	Mildly demanding of pilot attention, skill or effort	4.02	1.36	4.22	1.39	5.45	0.97	5.19
14	Some minor but annoying deficiencies	4.29	1.60	4.50	1.59	5.64	1.01	5.38
47	Good enough for mission	4.32	1.84	4.54	2.15	5.67	0.99	
23	Some mildly unpleasant characteristics	4.31	2.58	4.54	2.36	5.66	1.22	
1	Fair, somewhat impure primary or secondary response characteristics	4.42	2.29	4.62	2.43	5.76	1.27	5.50
42	Mildly sensitive, sluggish or uncomfortable	4.57	1.81	4.78	1.94	5.89	0.96	
26	Improvement is requested	4.87	2.12	5.08	2.27	6.04	1.03	
3	Moderately sensitive, sluggish or uncomfortable in primary or secondary responses	4.94	2.03	5.12	2.49	6.20	1.07	
8	Controllable with fair but somewhat inadequate precision	4.93	1.35	5.13	1.44	6.21	1.02	
54	Controllable with somewhat inadequate precision	5.22	1.24	5.43	1.28	6.37	0.83	6.10
25	Controllable with somewhat inadequate precision	5.27	1.35	5.45	1.40	6.44	0.90	
45	Controllable with fair, but somewhat inadequate precision	5.29	1.40	5.48	1.44	6.41	0.84	
9	Moderately objectionable deficiencies	5.34	1.60	5.57	1.48	6.51	0.93	6.26

TABLE B-I (Continued)

ITEM	PHRASE	OBSERVED ITEMS MEAN ① (R_O)	OBSERVED ITEMS VARIANCE ② ($\sigma^2_{R_O}$)	TRANSFORMED ITEMS MEAN ③ (R_T)	TRANSFORMED ITEMS VARIANCE ④ ($\sigma^2_{R_T}$)	COMPLETE MATRIX (63) x (63) ψ-SCALE VALUE ⑤ (m_i)	COMPLETE MATRIX (63) x (63) DISCRIMINAL DISPERSION ⑥ (s_i)	PARTIAL MATRIX (31) x (63) ψ-SCALE VALUE ⑦ (m_i)
39	Somewhat undesirably demanding of pilot attention, skill or effort	5.40	1.67	5.61	1.60	6.51	0.96	
40	Improvement is needed	5.48	2.03	5.71	2.15	6.59	0.96	
63	Demanding of pilot attention, skill or effort	5.68	1.57	5.88	1.70	6.72	0.91	6.47
59	Reasonable performance, requires considerable pilot compensation	5.70	1.95	5.92	1.98	6.82	0.93	
27	Quite sensitive, sluggish or uncomfortable in primary or secondary responses	5.83	1.74	6.00	2.49	6.81	0.91	6.56
51	Definitely sensitive, sluggish or uncomfortable in primary or secondary responses	5.94	2.06	6.10	2.84	7.03	0.90	
55	Objectionable deficiencies	6.10	2.37	6.34	2.35	7.03	1.08	
56	Definitely demanding of pilot attention, skill or effort	6.13	1.77	6.35	2.00	7.06	0.89	
61	Controllable with poor precision	6.26	2.07	6.44	2.40	7.06	1.03	
32	Pilot compensation required for acceptable performance in mission is too high	6.32	2.44	6.51	2.99	7.08	1.04	
34	Controllable with definitely inadequate precision	6.33	1.68	6.51	2.17	7.21	0.80	
58	Quite demanding of pilot attention, skill or effort	6.34	1.99	6.57	1.86	7.19	0.93	
18	Controllable, but only very imprecisely	6.40	1.68	6.65	1.59	7.15	0.82	6.92
35	Very sensitive, sluggish or uncomfortable in primary or secondary responses	6.51	2.24	6.67	2.87	7.26	0.96	
38	Requires substantial pilot skill and attention to retain control and continue mission	6.75	2.21	6.91	2.54	7.48	1.01	
64	Poor handling qualities	6.73	2.40	6.98	2.28	7.39	1.04	
46	Much too sensitive, sluggish or uncomfortable in primary or secondary responses	6.90	1.98	7.07	2.56	7.61	0.92	
19	Extremely sensitive, sluggish or uncomfortable in primary or secondary response	6.87	1.94	7.10	1.94	7.50	0.96	7.30
50	Controllable with difficulty	6.87	2.01	7.10	1.93	7.62	0.95	
6	Major improvements are needed	6.87	2.16	7.11	2.02	7.58	0.92	
12	Difficult to control	6.94	1.80	7.18	1.67	7.56	0.99	7.35
44	Requires best available pilot compensation to achieve minimum acceptable performance	7.05	2.91	7.26	3.16	7.74	1.09	
29	Very demanding of pilot attention, skill or effort	7.26	1.88	7.50	1.86	7.78	0.98	7.61
53	Major deficiencies	7.35	1.67	7.60	1.57	7.87	0.83	7.69

TABLE B-I (Concluded)

ITEM	PHRASE	OBSERVED ITEMS		TRANSFORMED ITEMS		COMPLETE MATRIX (63) x (63)		PARTIAL MATRIX (31) x (63)
		MEAN ① (R_o)	VARIANCE ② ($\sigma^2_{R_o}$)	MEAN ③ (R_T)	VARIANCE ④ ($\sigma^2_{R_T}$)	ψ-SCALE VALUE ⑤ (m_i)	DISCRIMINAL DISPERSION ⑥ (s_i)	ψ-SCALE VALUE ⑦ (m_i)
62	Very objectionable deficiencies	7.41	1.71	7.65	1.64	7.85	0.82	7.67
52	Extremely demanding of pilot attention, skill or effort	7.50	1.77	7.71	2.12	7.95	0.85	
48	Very poor handling qualities	7.48	1.64	7.74	1.58	7.87	0.88	
57	Bad handling qualities	7.49	1.97	7.74	1.81	7.97	1.05	7.81
60	Mandatory improvement required	7.51	2.15	7.76	2.04	8.00	0.99	
10	Very objectionable deficiencies	7.60	1.56	7.80	1.76	8.12	0.91	
31	Very difficult to control	7.91	1.00	8.15	1.18	8.23	0.65	8.09
15	Marginally controllable	7.93	1.36	8.15	1.51	8.29	0.86	
36	Very bad handling qualities	7.99	1.50	8.22	1.61	8.33	0.85	8.21
16	Completely demanding of pilot attention, skill or effort	8.14	1.40	8.36	1.41	8.49	0.95	8.39
13	Requires maximum available pilot skill and attention to retain control	8.54	1.00	8.79	1.00	8.87	0.89	
4	Barely controllable	8.56	0.78	8.80	0.92	8.79	0.71	
22	Nearly uncontrollable	8.67	0.78	8.91	0.59	9.00	1.05	9.00
28	Uncontrollable	9.71	0.11	10.00	0.00			

TABLE B-II

ITEMS INCLUDED IN LOW-VARIABILITY CALCULATION (COL. 7) OF TABLE B-I

ITEM NO.	ITEM	TRANSFORMED SCORE	
		MEAN	VARIANCE
	Handling Qualities		
17	Excellent handling qualities	1.00	0
7	Highly desirable handling qualities	1.47	0.45
41	Good handling qualities	2.58	1.51
49	Pleasant handling qualities	2.65	1.42
24	Fair handling qualities	4.13	1.39
57	Bad handling qualities	7.74	1.81
36	Very bad handling qualities	8.22	1.61
	Control		
11	Extremely easy to control with excellent precision	0.97	0.44
43	Very easy to control with good precision	1.76	0.63
5	Easy to control with fair precision	3.21	1.15
54	Controllable with somewhat inadequate precision	5.43	1.28
18	Controllable, but only very imprecisely	6.65	1.59
12	Difficult to control	7.18	1.67
31	Very difficult to control	8.15	1.18
22	Nearly uncontrollable	8.91	0.59
	Precision		
11	Extremely easy to control with excellent precision	0.97	0.44
43	Very easy to control with good precision	1.76	0.63
5	Easy to control with fair precision	3.21	1.15
25	Controllable with somewhat inadequate precision	5.45	1.40
18	Controllable, but only very imprecisely	6.65	1.59

TABLE B-II (Continued)

ITEM NO.	ITEM	TRANSFORMED SCORE	
		MEAN	VARIANCE
	Response Characteristics		
2	Excellent, pure (i.e., no accidental excitation) primary and secondary response characteristics	0.99	0.49
37	Good, relatively pure, primary and secondary response characteristics	2.47	0.88
1	Fair, somewhat impure primary or secondary response characteristics	4.62	2.43
27	Quite sensitive, sluggish or uncontrollable in primary or secondary responses	6.00	2.49
19	Extremely sensitive, sluggish or uncontrollable in primary or secondary responses	7.10	1.94
	Effects of Deficiencies		
20	Effects of deficiencies on performance is easily compensated for by pilot	4.04	1.33
14	Some minor but annoying deficiencies	4.50	1.59
9	Moderately objectionable deficiencies	5.57	1.48
53	Major, very objectionable deficiencies	7.65	1.64
	Demands on Pilot		
30	Completely undemanding of pilots, very relaxed and comfortable	1.65	0.94
21	Largely undemanding of pilots, relaxed	2.36	0.98
33	Mildly demanding of pilot attention, skill or effort	4.22	1.39

TABLE B-II (Concluded)

ITEM NO.	ITEM	TRANSFORMED SCORE	
		MEAN	VARIANCE
63	Demanding of pilot attention, skill or effort	5.88	1.70
29	Very demanding of pilot attention, skill or effort	7.50	1.86
16	Completely demanding of pilot attention, skill or effort	8.36	1.41

BLANK PAGE

APPENDIX C

TABULATION OF EXPERIMENTAL MEASURES AND DESCRIBING FUNCTIONS

This appendix contains the describing function plots which were computed for pilot JDM, and a tabulation (Table C-I) of the experimental measures made during the trials. The curve fits for the describing functions are shown on the figures themselves. The describing function figures are identified by run number and controlled element, and are in chronological order. (The run number gives the year, month, day, and number of run on that day. Thus, 671002-3 was the third run on October 2, 1967.)

ω_1 = 1.88 rad/sec

ω_2 = 2.89 rad/sec

ω_3 = 4.78 rad/sec

σ_1 = 0.5 cm/sec

σ_2 = 1.0 cm/sec

σ_3 = 1.5 cm/sec

TABLE C-I

RUN LOG AND SUMMARY OF PARAMETERS MEASURED DURING THE EXPERIMENTS

Single Loop: JDM

RUN NO.	Y_c/K_c	i ω,σ	Y_p τ_e (sec)	Y_p T_I (sec)	Y_p T_L (sec)	Y_pY_c ω_c	Y_pY_c ϕ_m (deg)	PERFORMANCE $\bar{e}/\bar{i}$	PERFORMANCE λ_0	PERFORMANCE $\lvert e \rvert$ (cm)	PERFORMANCE $\bar{\sigma}^2/\bar{i}^2$	PERFORMANCE ρ_a^2	RATINGS CR	RATINGS HQ	RATINGS RC	RATINGS C	RATINGS DP	RATINGS ED	RATINGS C-H
671005-11	1/s	1,3	.192	0	0	4	28	1.11	4.3	.553	.207	.798	4.5	6	6.2	6.5	6.7	6.7	A5
-13	$1/s^2$	1,3	.394	≫1	0	3	28	1.88	0	.937	.545	.62	6	7.5	6.8	7.9	7.8	7.3	A6
-15	.5/s	1,3	.217	0	0	3	38	1.07	6.3	.537	.190	.818	3	4	5.5	4.2	6.4	4.4	A3
-17	.1/s(s+4)	1,2	.40	0	0	2	40	1.05	4	.507	.367	.809	4	5	4.5	4.2	6.7	6.7	A4.5
-19	$.1/s^2$	1,1	.435	≫1	1/4.5	2	36	.926	3.5	.44	1.30	.676	7	8	7.5	7.5	7.5	7.6	U7
-21	.1/s	1,3	.45			2.75	54	1.43	1.8	.67	.272	.846	8	8.5	8	8.6	8.5	8.1	U8
671006-1	.1/s(s+1)	1,2		1.0		3.4	12	.954	2.9	.464	.336	.624	4.5	6	6.2	5.2	5.6	6.3	A5
-3	$.5/s^2$	1,2	.357	>3.3	0	3.3	12	1.22	1.1	.591	.528	.493	4.5	6	6.3	6.5	6.7	6.7	A4.5
-5	1/.7, 7.8	1,2	.213	0	>1/.3	4.5	52	.856	4	.417	.269	.927	3	4	4.5	5.6	6	5.5	A3
-7	1/.7, 16	1,2	.302	0	~1.0	2	96	1.00	2.8	.483	.348	.9	7	8.3	6.8	6.8	6.7	7.3	U7.5
-9	10/.7, 16	1,2	.200			3.3	44	.92	3	.443	.306	.94	4.5	5.5	6.5	6.5	4.7	6.3	A5
-11	10/s	1,3	.194	0	0	4.6	30	1.11	1.8	.55	.214	.702	5.5	6	6.8	6.8	6.7	7	A5.5
-13	10/s	1,1	.192	0	0	4.5	28	.657	4	.31	.588	.521	4	5	7.5	4.7	6.4	7	A5
-15	10/s	1,2	.167	0	0	4.5	28	.914	3.4	.442	.306	.584	4	5	7.2	5.6	5.2	6	A4.5
-17	5/.7, 16	1,2		>3.3	0	3.2	60	.84	3.4	.406	.257	.968	5	6	7.2	6	5.6	6.3	A5
-19	$10/s^2$	1,1	.238			5.0	20	1.43	1	.597	2.46	.256	7	8.3	8	7.9	7.8	8.1	U7
-21	$10/s^2$	1,3	.263	>1	0	3.3	28	1.42	0			.206	9	9	8.7	9.3	8.8	8.6	U9.5
-23	$10/s^2$	1,2	.244	~.3	0	4.8	14	1.42	0	.687	.392	.521	8.5	9	8.4	9.3	8.8	8.6	U9
-25	1/(s-2)	1,2	.147	~0	0	4.7	20	1.53	4	.744	.452	.741	8.5	9	8	9	9.2	8.6	U9
-27	10/.7, 7.8	1,2	.189	0		3.3	50	.915	4	.443	.159	.948	4.5	5	7.5	5.2	6.7	6.7	A5
-29	5/s	1,2	.208	0	0	4.3	30	.885	4	.428	.281	.6	4	4.8	6.2	4.2	4.7	6	A4
-31	$5/s^2$	1,2	.295	>1	0	3.8	18	1.41	0	.685	.766	.506	7.5	8.2	7.2	7.5	7.8	8.1	U7.5
671009-1	.1/.7, 7.8	1,2	.385	0	1.0	2.2	86	.955	2.5	.459	.318	.888	8	8.3	7.5	7.2	6.7	8.1	U8
-3	1/.7, 16	1,2	.185	0	1/1.6	3.1	56	.732	3.8	.35	.186	.962	3	4	4.5	5.2	6	6	A3
-5	10/s(s+4)	1,2	.222	2.27	0	5	16	1.06	2.0	.51	.345	.439	7.5	8.5	8.7	8.6	7.5	7.9	U7.5
-7	1/s(s+1)	1,2	.294	~1	0	3.4	16	1.09	2.6	.518	.44	.604	5	6	6.2	6.5	5.2	7	A5.5
-9	1/s(s-1)	1,2	NA →					3.44	0	.757			9	9	8.7	9.3	8.8	8.6	10
-11	.5/.7, 16	1,2	.172	0	1/1.8	3.3	64	.72	4.6	.348	.187	.96	3	4	4	4.2	4.7	3.2	A2.5

TABLE C-I (Continued)

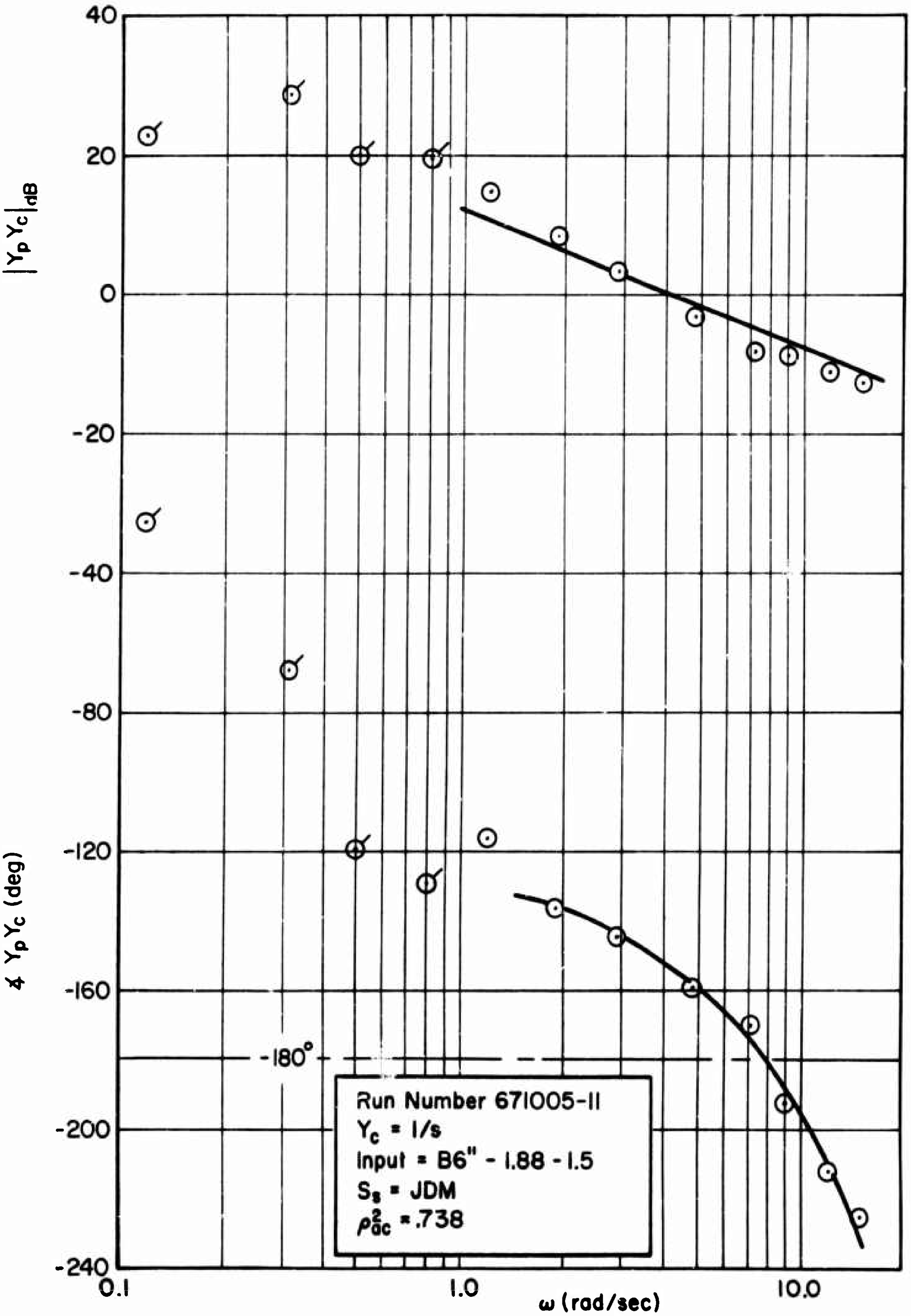
Single Loop: JDM

RUN NO.	Y_c/K_c	i ω,σ	Y_p τ_p (sec)	τ_L (sec)	τ_I (sec)	Y_pY_c ω_c ([illegible])	ϕ_m (deg)	PERFORMANCE $\lvert\bar{e}\rvert/\lvert\bar{i}\rvert$	λ_c ([illegible])	$\lvert\bar{e}\rvert$ (cm)	$\bar{e}^2/\bar{i}^2$	ρ_c^2	RATINGS CR	HQ	RC	C	DP	ED	C-H
671011 – 1	10/s(s+1)	1,2	.172	1	0	2.8	50	1.25	0	.608	.580	.0973	8	9	8.7	9	8.5	8.4	U9
– 3	.5/s	1,2	.133	0	0	3.6	48	.732	3.9	.355	.196	.81	3	4	4	3.6	4.2	6	A3
– 5	1/s	1,2	.200	0	0	4.0	40	.706	4.2	.342	.184	.781	2.5	3.7	4	3.6	3.7	5.5	A2.5
– 7	1/s	2,2	.192	0	0	4.2	40	.835	3.6	.403	.26	.819	3	4	4.5	4.2	4.2	6	A3
– 9	1/s	3,2	.147	0	0	4.5	60	1.08	3.5	.515	.428	.768	5	6	6.5	6.8	5.2	6.3	A5
– 11	1/s	1,1	.193	0	0	3.4	48	.413	5.5	.191	.252	.66	2	3.3	3.5	3	4.2	4.4	A2
– 13	1/s	1,3	.200	0	0	3.6	48	1.03	3.9	.508	.362	.702	4	5	6	5.2	5.2	6	A4
– 15	1/s	2,1	.189	0	0	4.3	40	4.85	5.3	.277	.322	.79	2	3	2.5	3	4.2	NA	A1.5
– 17	1/s	2,3	.182	0	0	5.0	36	1.25	2.8	.618	.270	.674	5	6.5	6.5	6.8	5.6	6.7	A5
– 19	$1/s^2$	1,2	.333	>1	0	2.9	20	1.41	2.1	.672	.735	.512	7	8	6.8	7.9	6.4	7.6	U7
– 21	$1/s^2$	2,2	.286	>1	0	3.0	24	1.56	1.0	.743	.886	.618	8.5	8.5	7.2	8.6	8.5	8.6	U8.5
– 23	$1/s^2$	3,2	.250	>3.3	0	2.4	40	2.03	0	.97	1.53	.744	8.5	8.5	6.8	8.2	8.2	8.4	U8.5
– 25	$1/s^2$	1,1	.344	>1	0	3.2	16	.682	3.5	.317	.636	.564	3.4	4	4.5	4.2	4.2	4.4	A3.4
– 27	$1/s^2$	2,1	.278	>3.3	0	4.0	20	.877	3.1	.417	1.124	.567	4	4.8	4.5	5.2	4.7	6	A4
– 29	$1/s^2$	2,3	.400	>1.5	0	3.2	8	2.04	0	1.01	.680	.658	8.0	9	8.7	7.9	7.8	8.6	U8.5
671129 – 1	1/s(s+2)	1,2	.264	.5	0	4.2	24	.965	3	.442	.307	.421	4	5	6.2	6	5.6	6.3	A4.5
– 3	1/s(s+4)	1,2	.250	.250	0	4.2	26	.845	3.8	.38	.214	.613	4	5	6.2	6	6	6.7	A4.5
– 5	$.1/s^2$	1,2	.333	>1	0	1.5	40	2.13	0.5	.933	2.62	.387	8	8.5	7.2	8.2	8.2	8.4	U8
– 7	.1/s(s+1)	1,2	.384	1	0	2.8	28	.981	0	.448	.155	.601	5.5	7	6.5	7.2	7	7	A5.5
– 9	.1/s	1,2	.270	0	0	3.1	44	.813	3.2	.373	1.07	.859	7	8	7.2	7.5	7.8	7.6	U7
– 11	$1/s^2$	1,2	.330	>1	0	4.0	11	1.23	1.8	.563	.279	.457	6	7.5	6.8	7.2	7.5	7	A6
– 13	1/s(s+1)	1,2	.244	1	0	3.8	40	1.08	1.7	.498	.206	.478	5.5	7	6.8	6.8	6.7	7	A5.5
– 15	$1/s^2$	1,2	.3	>1	0	3.3	20	1.23	1.5	.563	.251	.517	6	7.5	7.2	7.5	7.5	7.3	A6

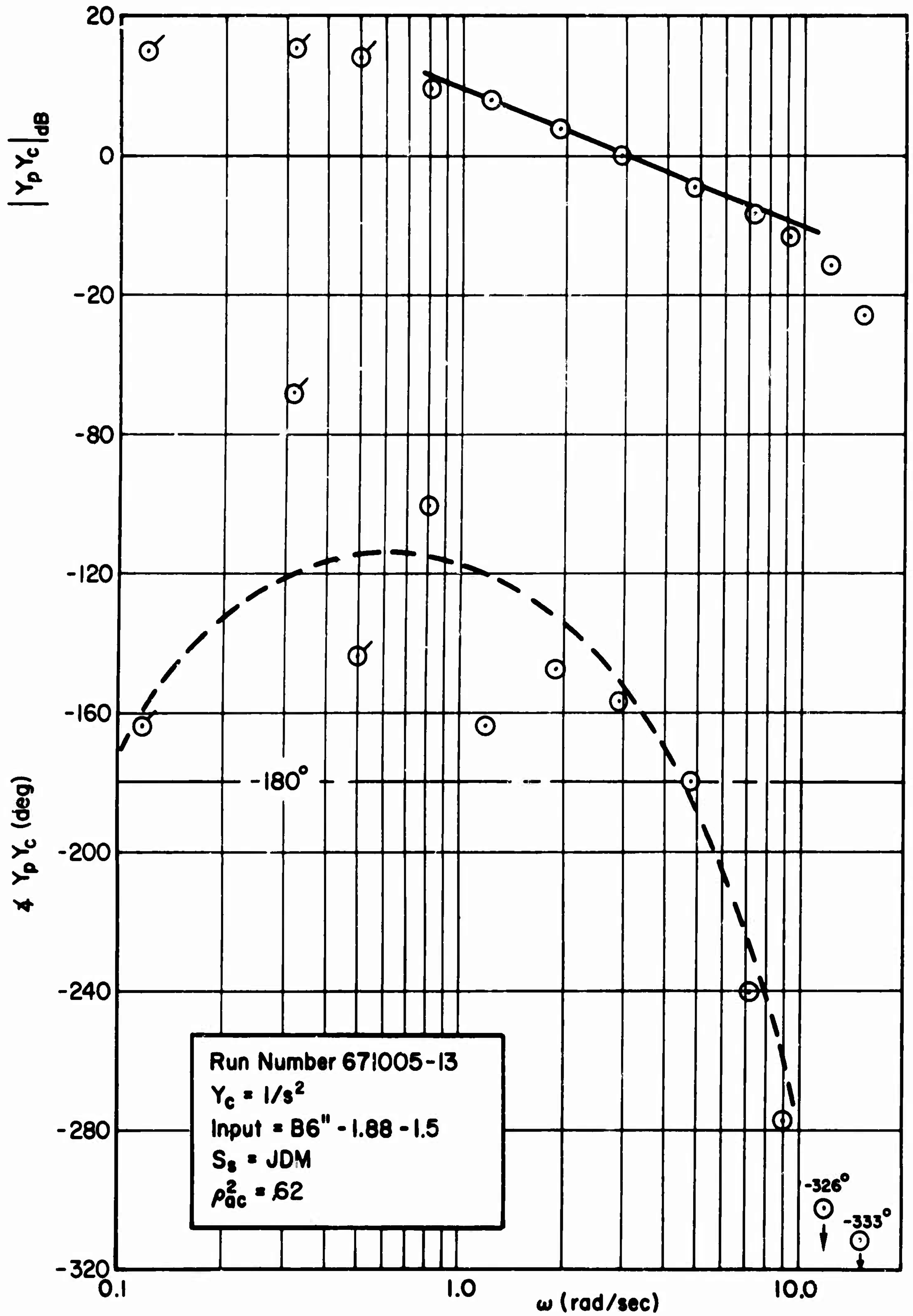
TABLE C-I (Concluded)

Single Loop: MDK

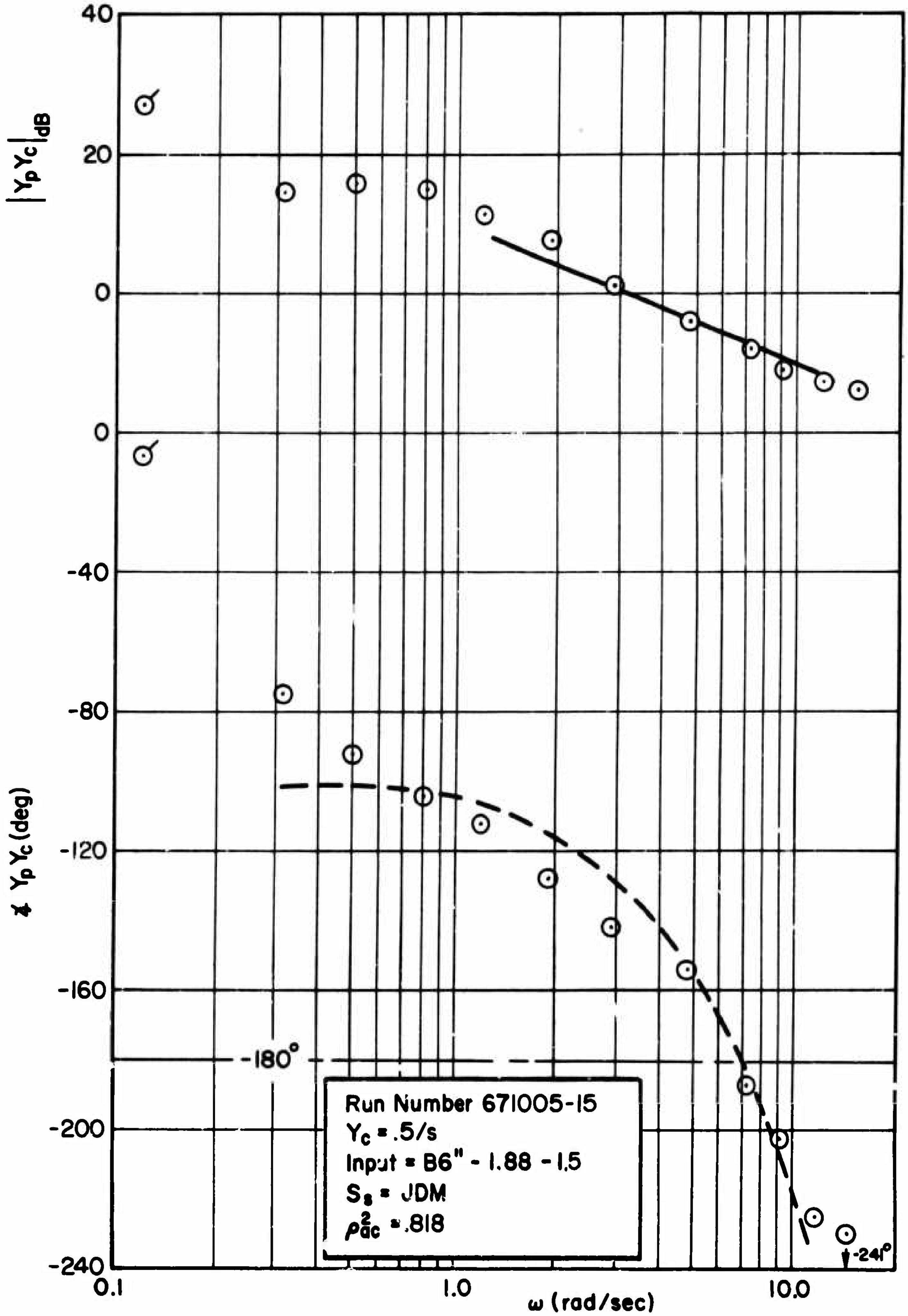
RUN NO.	Y_c/K_s	i ω,σ	PERFORMANCE			RATINGS						
			|ε|/|i|	λ_0([redacted])	|ε| (ω)	CR	HQ	RC	C	DP	ED	C-H
671030 - 01	1/s(s+2)	1,2	1.13	2	.527	8	9	8.7	9	8.8	8.6	U9
- 03	1/s(s+4)	1,2	1.18	2.4	.547	6	8	6.8	6.8	6.7	7	U7
- 05	1/s²	1,2	1.77	.2	.825	7	8.3	7.5	7.9	7.8	8.1	U8
- 07	1/s²	3,2		0		9	UNC	UNC	UNC	UNC	UNC	UNC
- 09	1/s	1,3	1.54	1.2	.739	6	7	6.8	7.5	6.7	7	U8
- 11	1/s²	1,3		0		9	UNC	UNC	UNC	UNC	UNC	UNC
- 13	.5/s	1,2	1.2	3.6	.558	4	5	6	6.5	6.7	6.3	A5
- 15	.1/s(s+4)	1,2	1.58	1	.732	7	8	6.8	7.5	7.5	8.1	U8
671031 - 91	.1/s²	1,1	1.18	1.1	.549	3	5	5	4.7	5.2	6.3	A4
- 03	.1/s	1,3	1.92	.2	.978	7	8	7.5	7.5	8.2	8.1	U8
- 05	.1/s(s+1)	1,2	1.53	2	.722	5	7	6.8	7.2	6.7	7	U7
- 07	.5/s²	1,2	1.7	.65	.802	7	8.3	7.5	7.9	8.2	8.1	U8
- 09	1/.7, 7.8	1,2	1.24	3.2	.588	3	5	5.5	6	5.6	6.3	A5
- 11	.1/.7, 16	1,2	1.15		.544	6	8	8	7.2	6.7	8.1	U7
- 13	10/.7, 16	1,2	.883	2	.418	6	7	7.5	5.2	6	7.6	A5
- 15	10/s	1,3	1.32	1.6	.643	6	8	6.8	6.8	7.5	7.6	U8
- 17	10/s	1,1	.647	1.8	.299	5	5	6.8	4.2	6.7	7	A5
- 19	10/s	1,2	.988	2.5	.467	7	8	7.5	7.5	7.8	8.1	U8
- 21	5/.7, 16	1,2	1.07		.504	6	7	6.8	7.5	7.5	7.3	U8
- 23	10/s²	1,1				9	UNC	UNC	UNC	UNC	UNC	UNC
- 25	10/s²	1,3				9	UNC	UNC	UNC	UNC	UNC	UNC
- 27	10/s²	1,2				9	UNC	UNC	UNC	UNC	UNC	UNC
- 29	1/(s-2)	1,2				9	UNC	UNC	UNC	UNC	UNC	UNC
- 31	10/.7, 7.8	1,2	.963	2.4	.457	7	8	6.8	7.5	7.5	7.6	U8
- 33	5/s	1,2	.878	4.2	.414	6	8	6.8	7.2	7.5	7.6	U7
671101 - 01	5/s²	1,2		0		9	UNC	UNC	UNC	UNC	UNC	UNC
- 03	.1/.7, 7.8	1,2	1.12	3.5	.529	6	8.3	8	6.8	5.2	8.1	U7
- 05	1/.7, 16	1,2	.973	3.2	.459	6	8.3	6.8	6.8	5.2	8.1	U7
- 07	10/s(s+4)	1,2	1.41	1.3	.668	7	8.3	6.8	7.5	7.5	8.1	U8
- 09	1/s(s+1)	1,2	1.5	2.5	.708	6	8	6.8	6.8	6.7	7.6	U7
- 11	1/s(s-1)	1,2		0		9	UNC	UNC	UNC	UNC	UNC	UNC
- 13	.5/[.7, 16]	1,2	1.11	3.5	.527	6	8.3	8	7.5	5.2	8.1	U7
- 15	10/s(s+1)	1,3				9	UNC	UNC	UNC	UNC	UNC	UNC
- 17	1/s	1,2	.87	3.9	.411	6	8.3	8	6.8	6	8.1	U7
- 17'	1/s	2,2		1.9								
- 23	1/s	3,2	1.44	3.5	.687	5	7	6.8	6.8	5.2	7.6	U7
- 25	1/s	1,1	.519	4.5	.241	3	4	4.5	4.2	5.2	6	A4
- 27	1/s	2,1	.724		.337	3	5	4.5	5.2	4.2	6.3	A4
671102 - 01	1/s	2,3	1.65		.832	4	5	5.5	6	5.6	6.3	A5
- 03	1/s²	2,2	1.96	0	.953	8	9	8.7	9	8.2	8.1	U9
- 05	1/s²	1,1	1.08	3	.51	7	8	7.5	6	6.7	8.1	U7
- 07	1/s²	2,1	1.52	3.2	.718	7	8.3	7.5	7.5	7.5	8.1	U8
- 09	1/s²	2,3				9	UNC	UNC	UNC	UNC	UNC	UNC
- 11	1/s	1,2	1.09	3.2	.521	6	8	7.5	6.8	5.2	8.1	U7
- 13	1/s²	1,2	2.79	0	1.31	7	8.3	7.5	7.5	7.8	8.1	U8
- 15	1/s	1,1	.705	4.1	.333	6	8	7.5	6.5	5.2	8.1	U7
- 17	1/s²	1,3				9	UNC	UNC	UNC	UNC	UNC	UNC

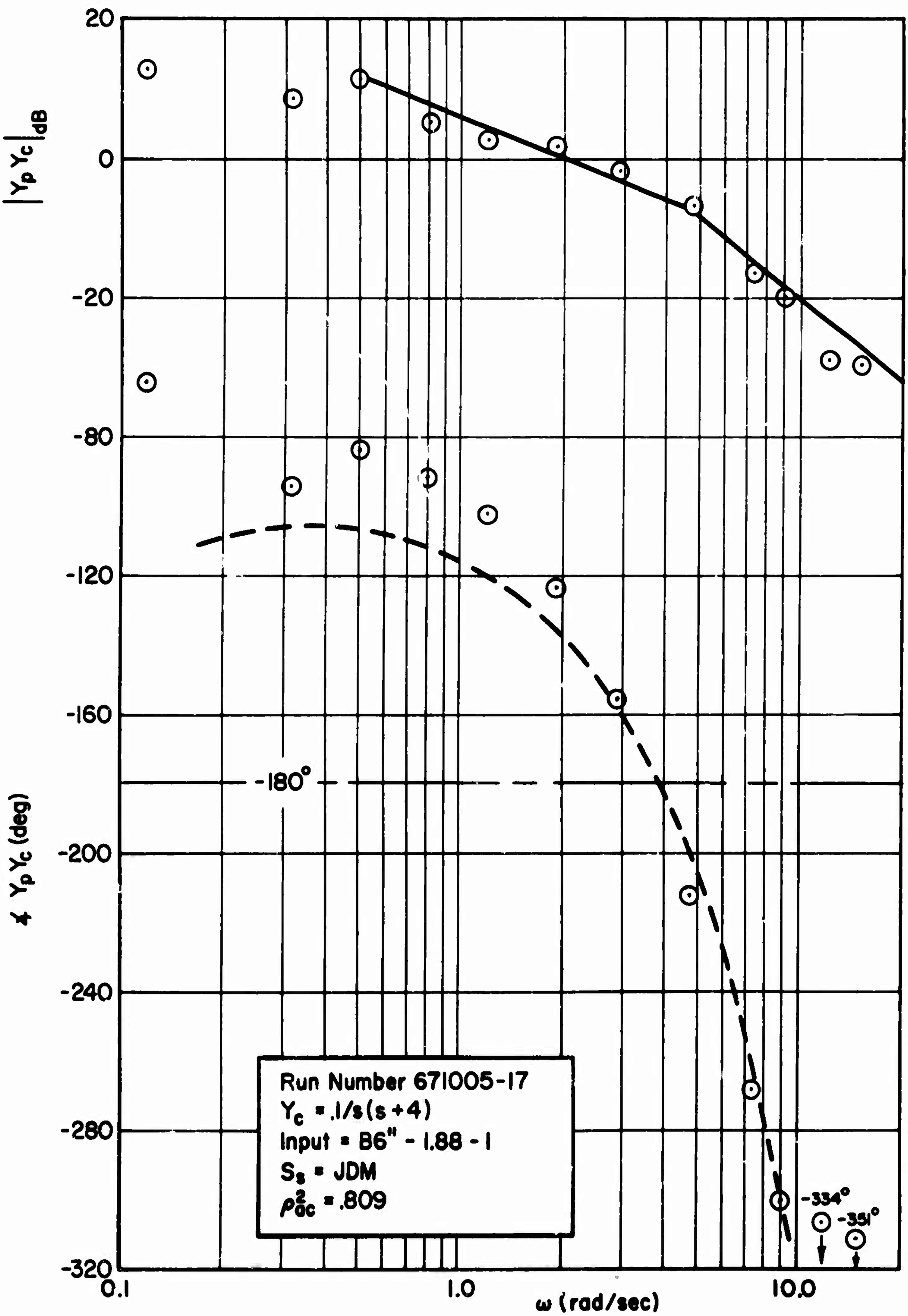


40
20
0
-20
-40
-80
-120
-160
-200
-240
$|Y_p Y_c|_{dB}$
∡ $Y_p Y_c$ (deg)
-180°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671005-11
Y_c = 1/s
Input = B6" - 1.88 - 1.5
S_s = JDM
ρ^2_{ac} = .738

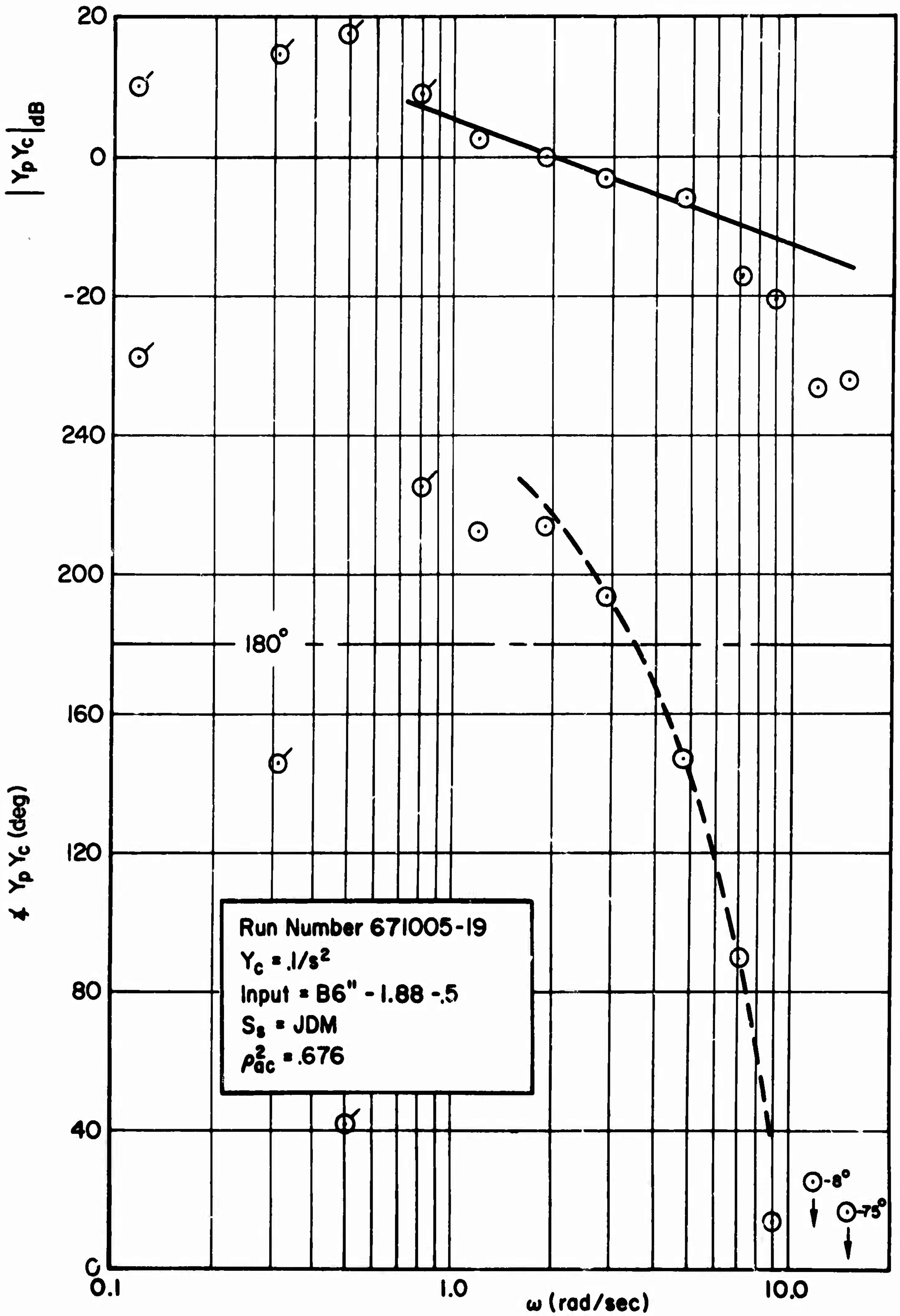


20
0
-20
-80
-120
-160
-180°
-200
-240
-280
-320
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
0.1
1.0
10.0
ω (rad/sec)
-326°
-333°
Run Number 671005-13
$Y_c = 1/s^2$
Input = B6" - 1.88 - 1.5
S_s = JDM
ρ^2_{ac} = .62

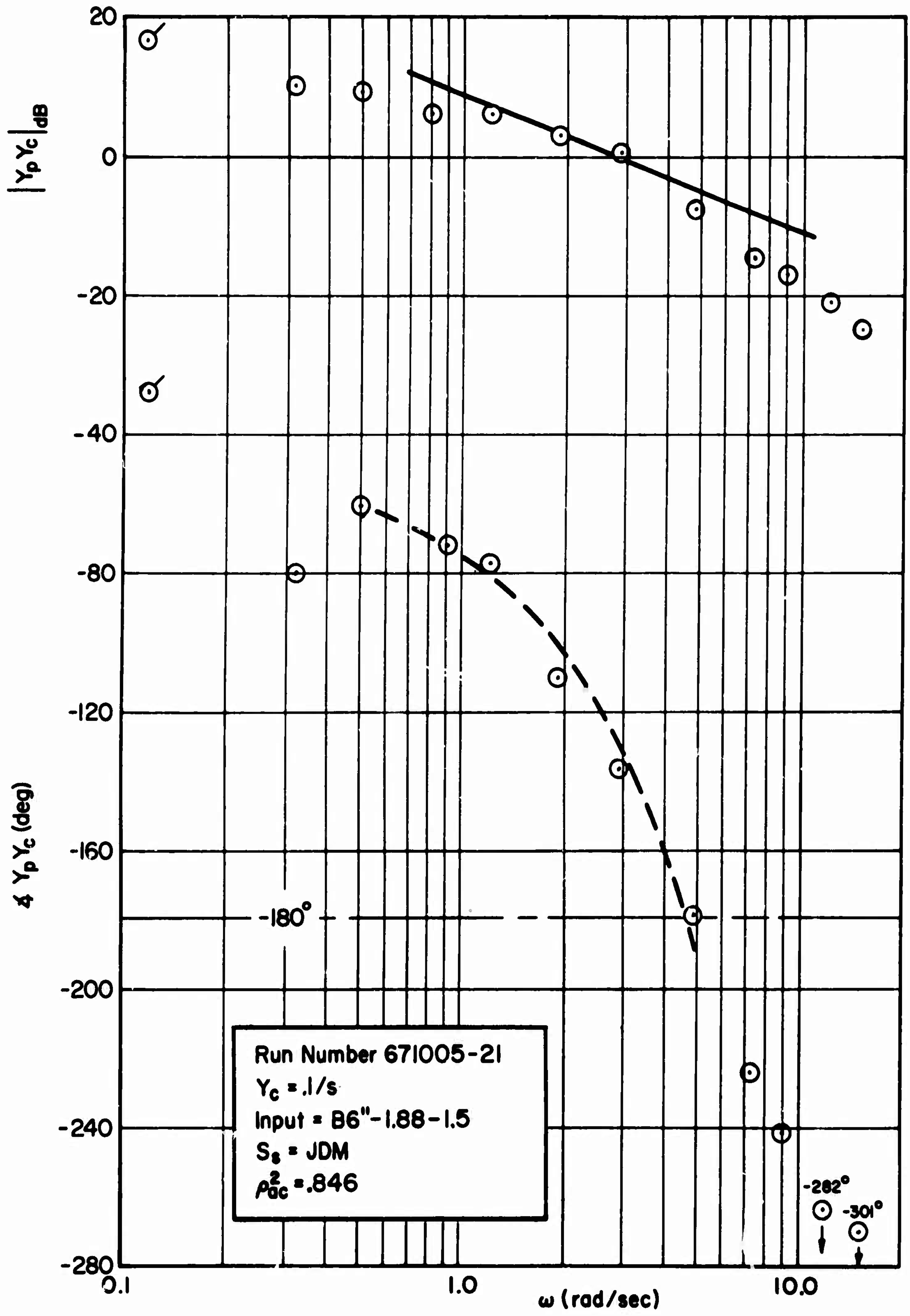

40
20
0
0
-40
-80
-120
-160
-200
-240
$|Y_p Y_c|_{dB}$
∡ $Y_p Y_c$ (deg)
-180°
-241°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671005-15
Y_c = .5/s
Input = B6" - 1.88 - 1.5
S_s = JDM
ρ^2_{ac} = .818



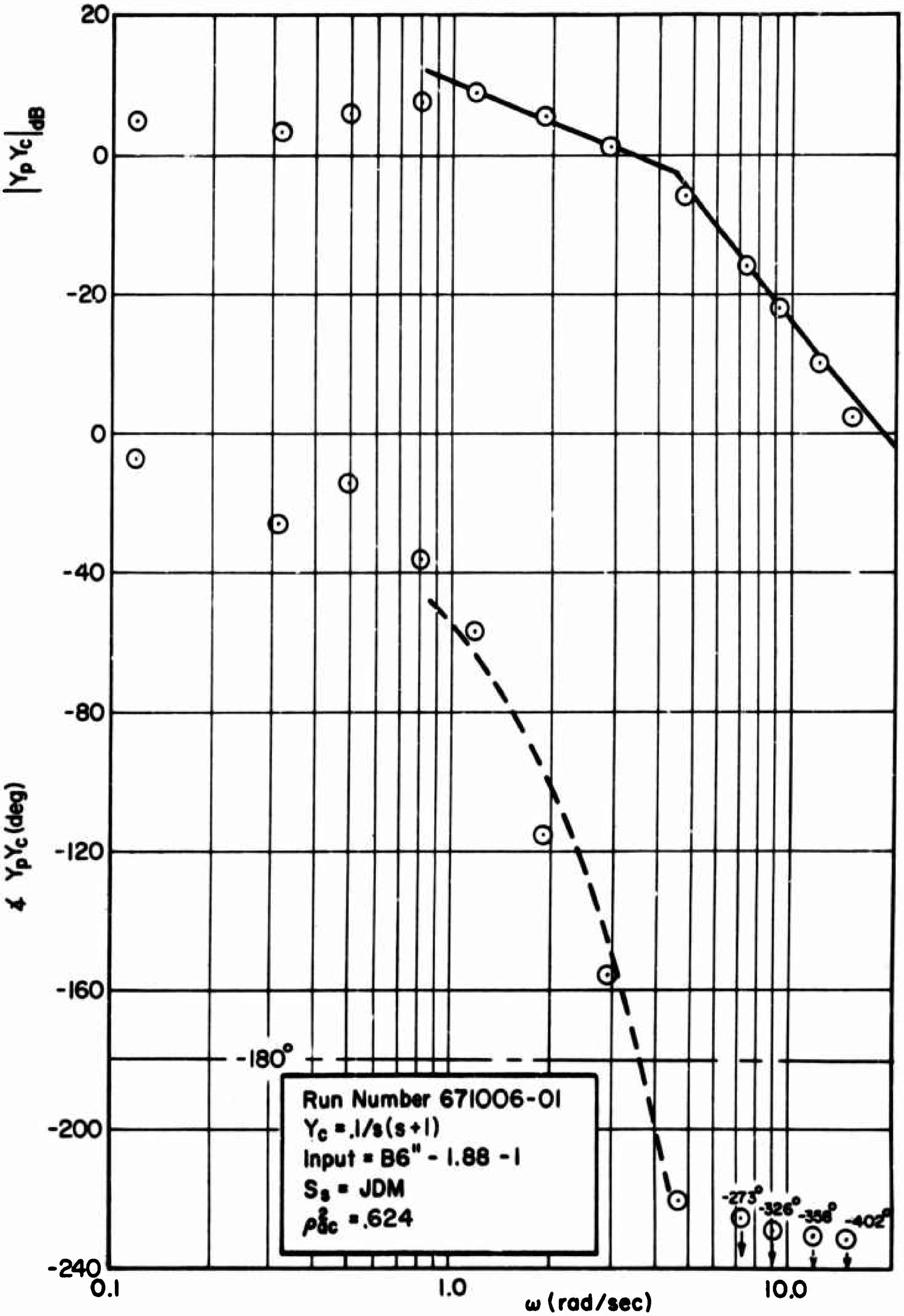
20
0
-20
$|Y_p Y_c|_{dB}$
-80
-120
-160
-180°
-200
$\measuredangle\ Y_p Y_c$ (deg)
-240
-280
-320
-334°
-351°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671005-17
Y_c = .1/s(s+4)
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} = .809



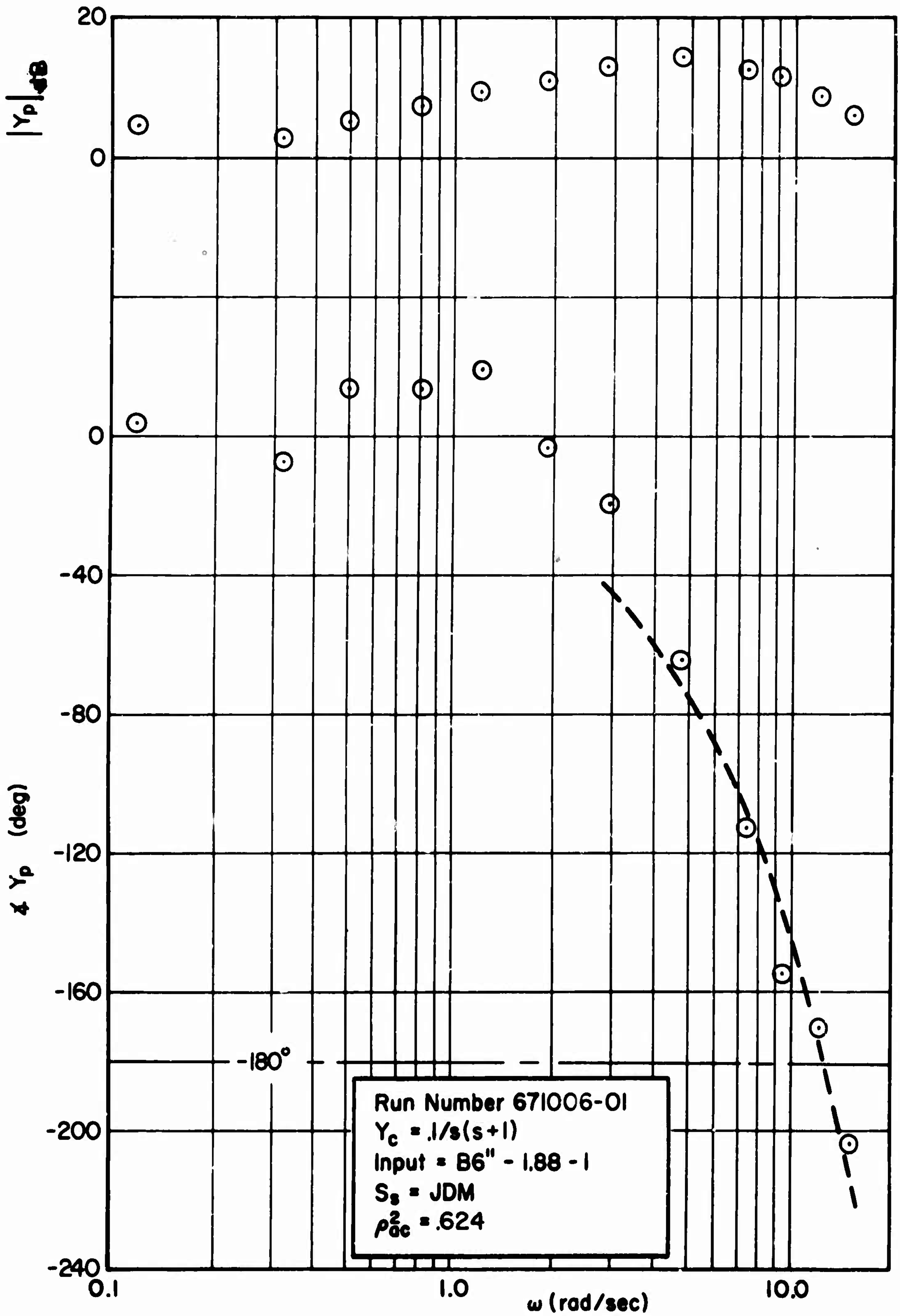
20
0
-20
240
200
180°
160
120
80
40
0
0.1
1.0
10.0
ω (rad/sec)
|Yp Yc|dB
∠ Yp Yc (deg)
Run Number 671005-19
Yc = .1/s²
Input = B6" - 1.88 -.5
Ss = JDM
ρ²ac = .676
-8°
-75°



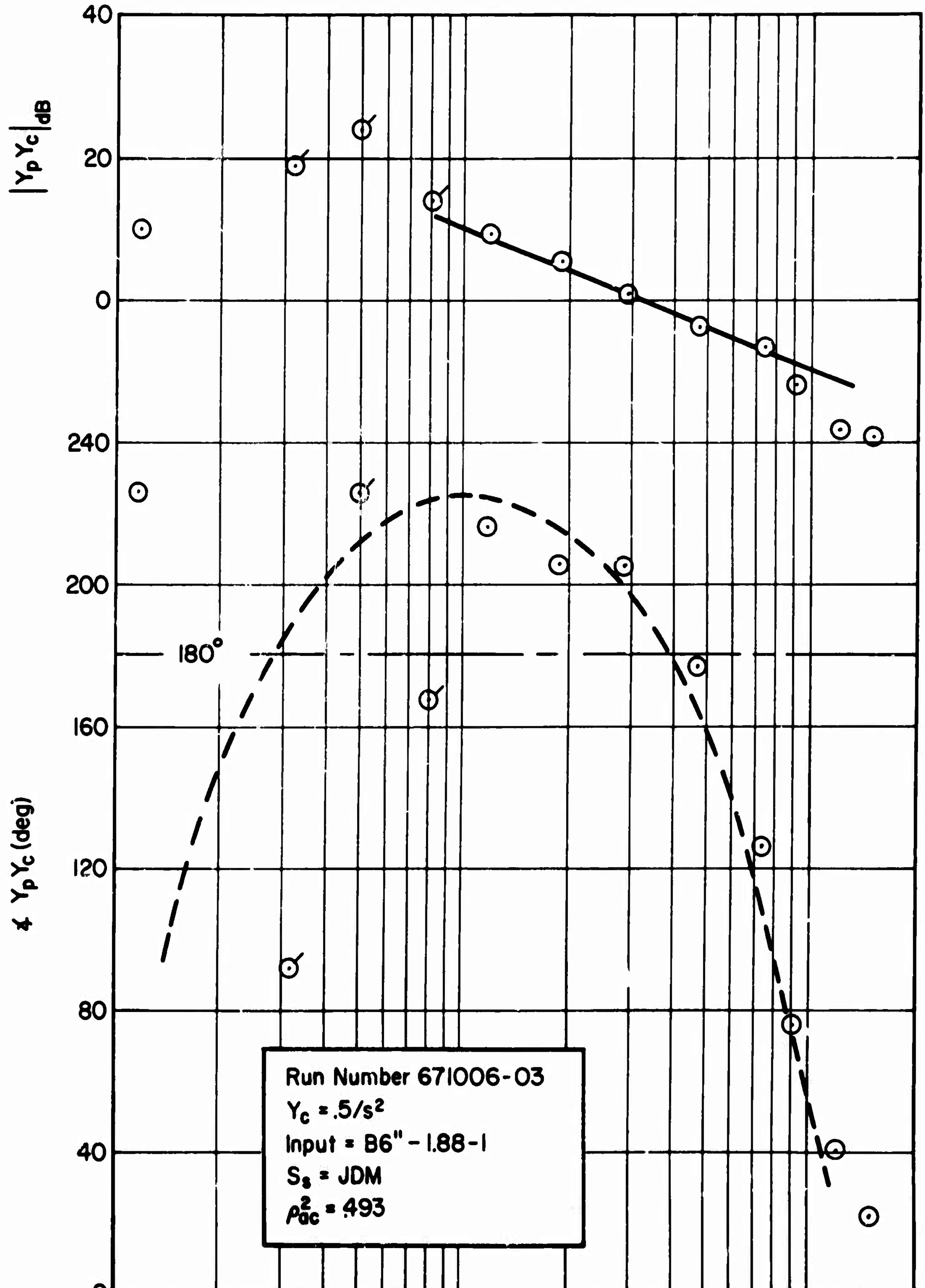
20
0
-20
-40
-80
-120
-160
-200
-240
-280
$|Y_p Y_c|_{dB}$
∡ $Y_p Y_c$ (deg)
-180°
Run Number 671005-21
Y_c = .1/s
Input = B6"-1.88-1.5
S_s = JDM
ρ^2_{ac} = .846
-282°
-301°
0.1
1.0
10.0
ω (rad/sec)



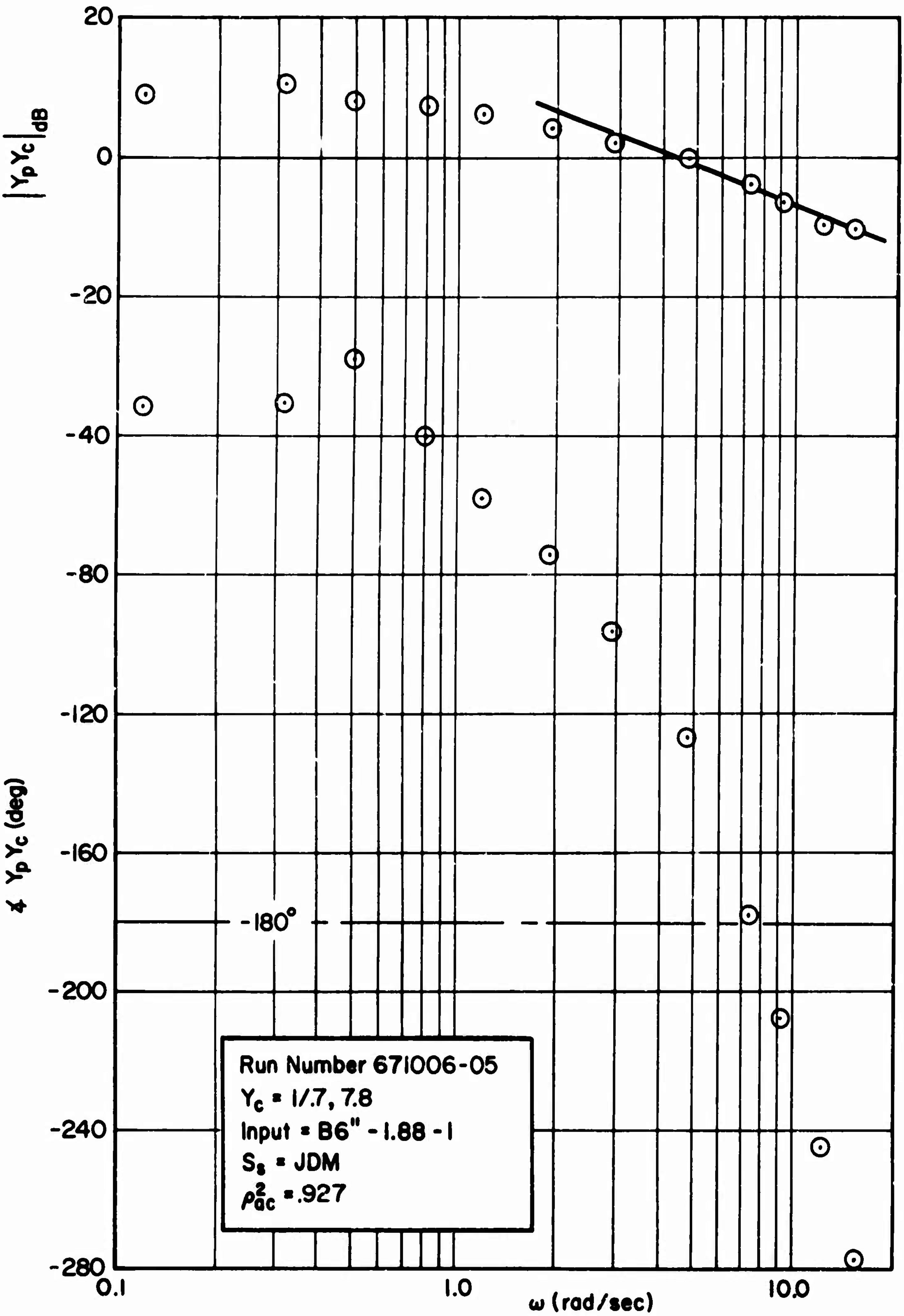
20
0
-20
0
-40
-80
-120
-160
-200
-240
0.1
1.0
10.0
|Yp Yc|dB
∡ Yp Yc (deg)
ω (rad/sec)
-180°
-273°
-326°
-356°
-402°
Run Number 671006-01
Yc = .1/s(s+1)
Input = B6" - 1.88 - 1
Ss = JDM
ρ²ac = .624

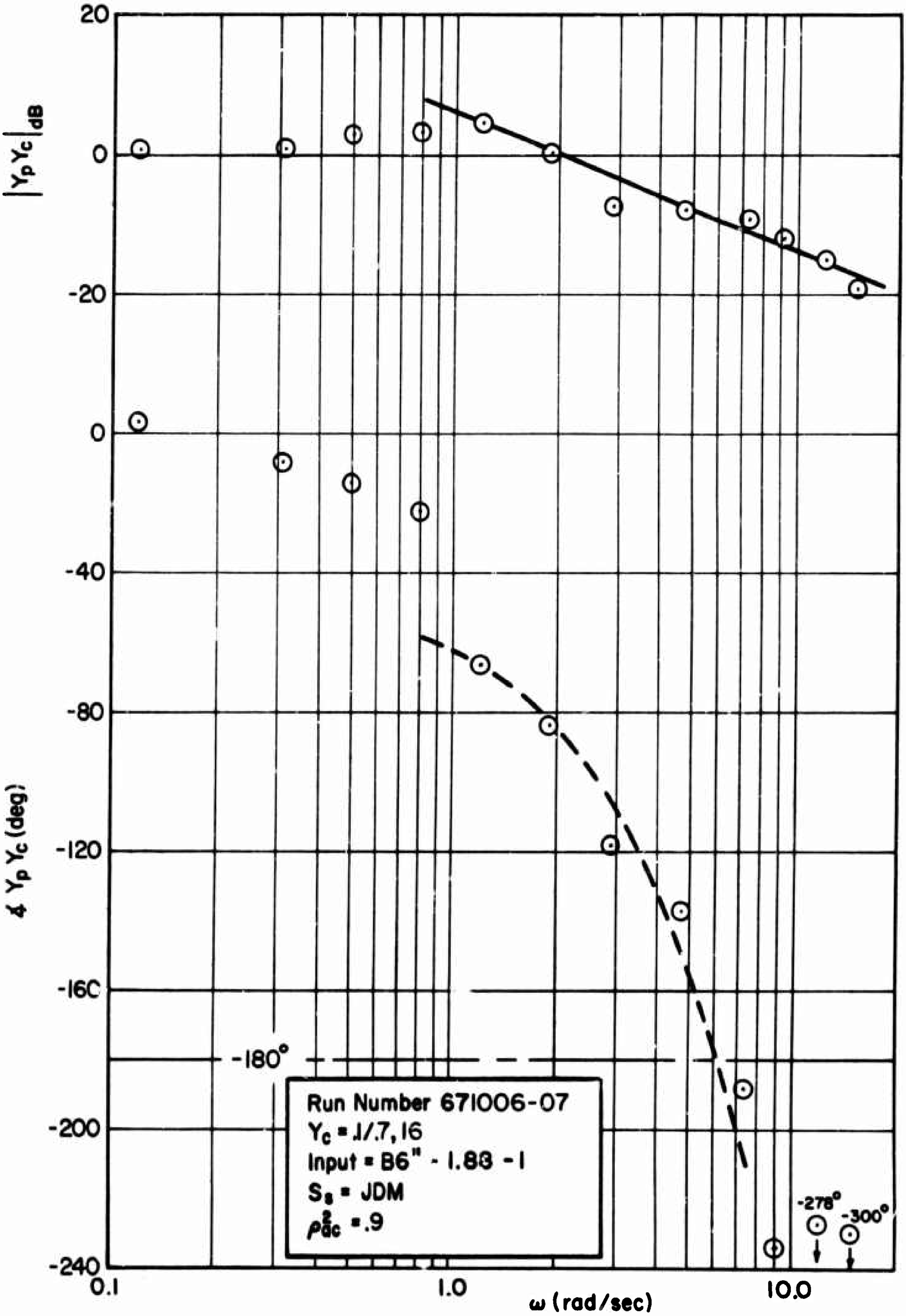


20
0
$|Y_p|_{dB}$
0
-40
-80
-120
-160
-200
-240
$\measuredangle Y_p$ (deg)
-180°
Run Number 671006-01
Y_c = .1/s(s+1)
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} = .624
0.1
1.0
10.0
ω (rad/sec)

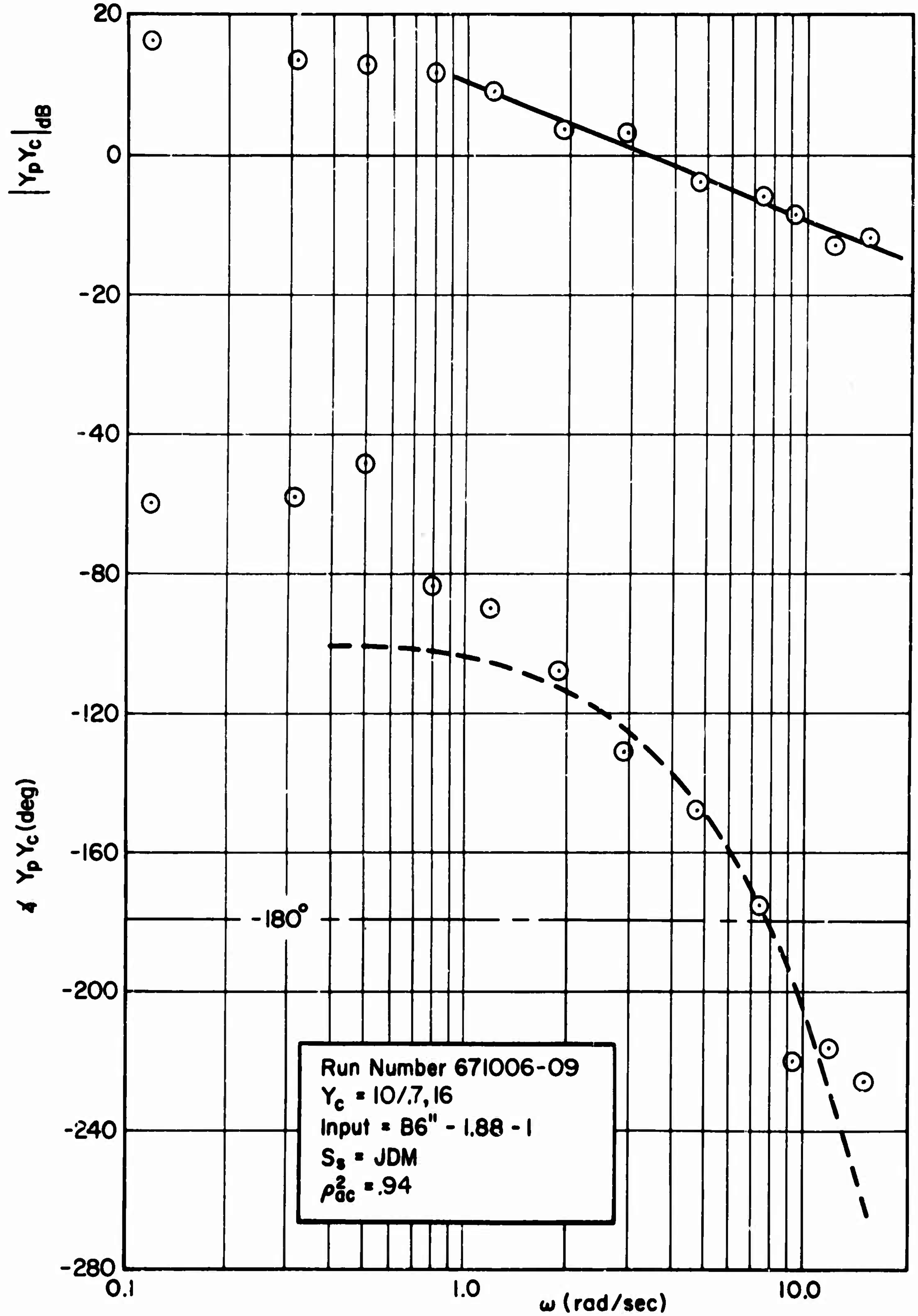


40
20
0
240
200
180°
160
120
80
40
0
0.1
1.0
10.0
ω (rad/sec)
$|Y_p Y_c|_{dB}$
∡ $Y_p Y_c$ (deg)
Run Number 671006-03
$Y_c = .5/s^2$
Input = B6" - 1.88 - 1
S_s = JDM
ρ_{ac}^2 = .493

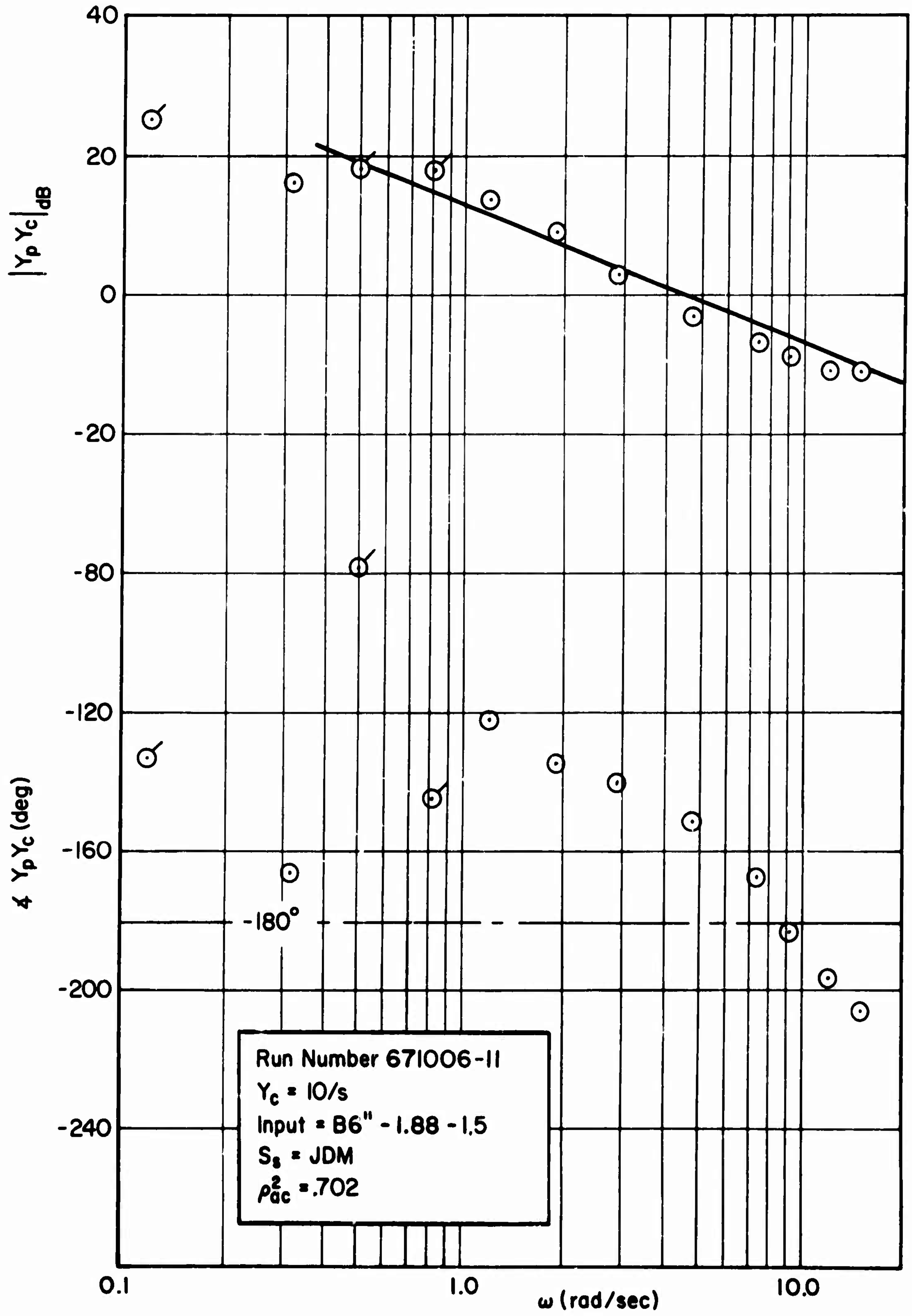

20
0
-20
-40
-80
-120
-160
-200
-240
-280
$|Y_p Y_c|_{dB}$
$\measuredangle\ Y_p Y_c$ (deg)
-180°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671006-05
Y_c = 1/7.7, 7.8
Input = B6" - 1.88 - 1
S_s = JDM
ρ_{ac}^2 = .927

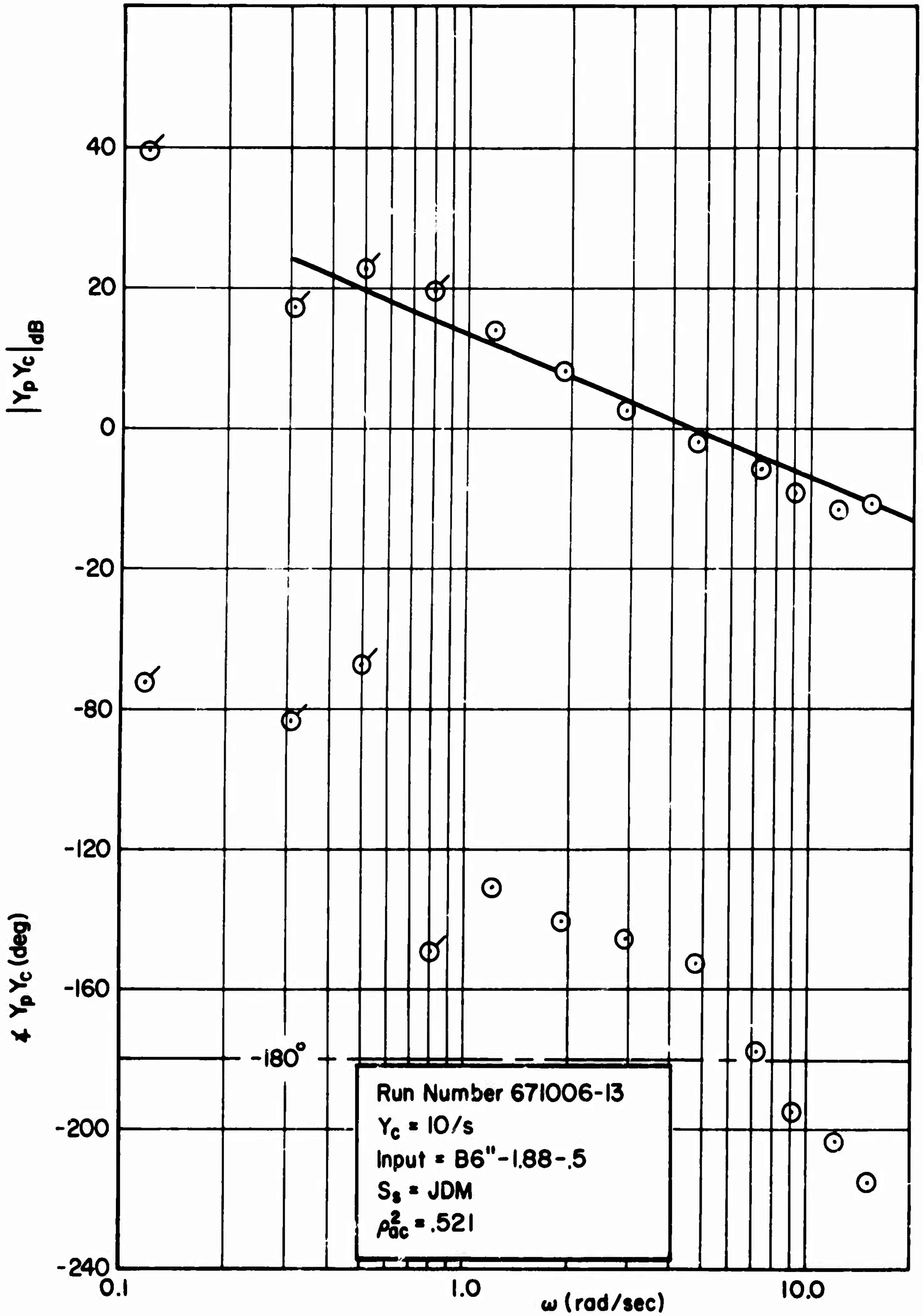


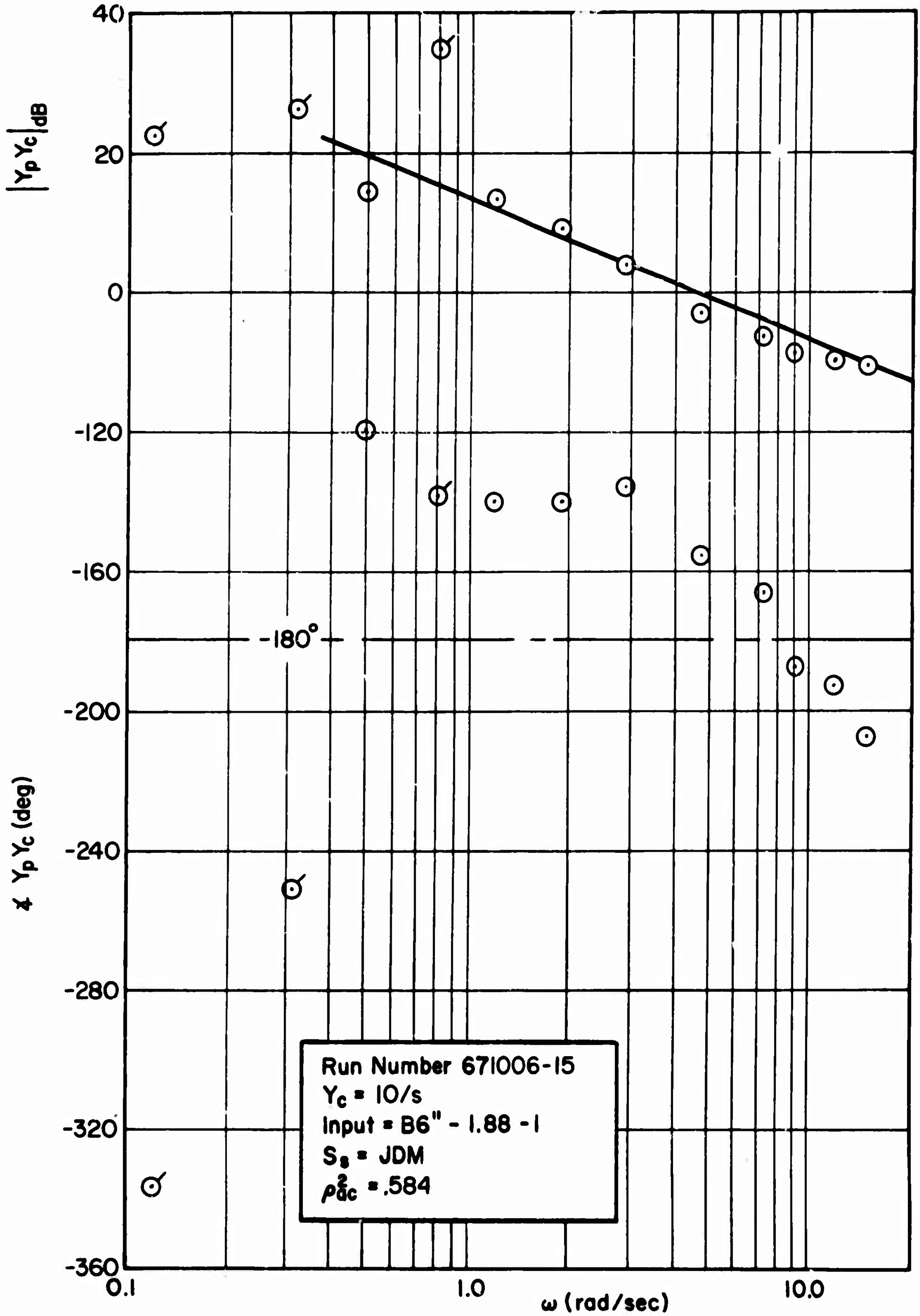
20
0
-20
0
-40
-80
-120
-160
-180°
-200
-240
0.1
1.0
10.0
ω (rad/sec)
-278°
-300°
Run Number 671006-07
Yc = 1/.7,16
Input = B6" - 1.83 - 1
Ss = JDM
ρ²ac = .9

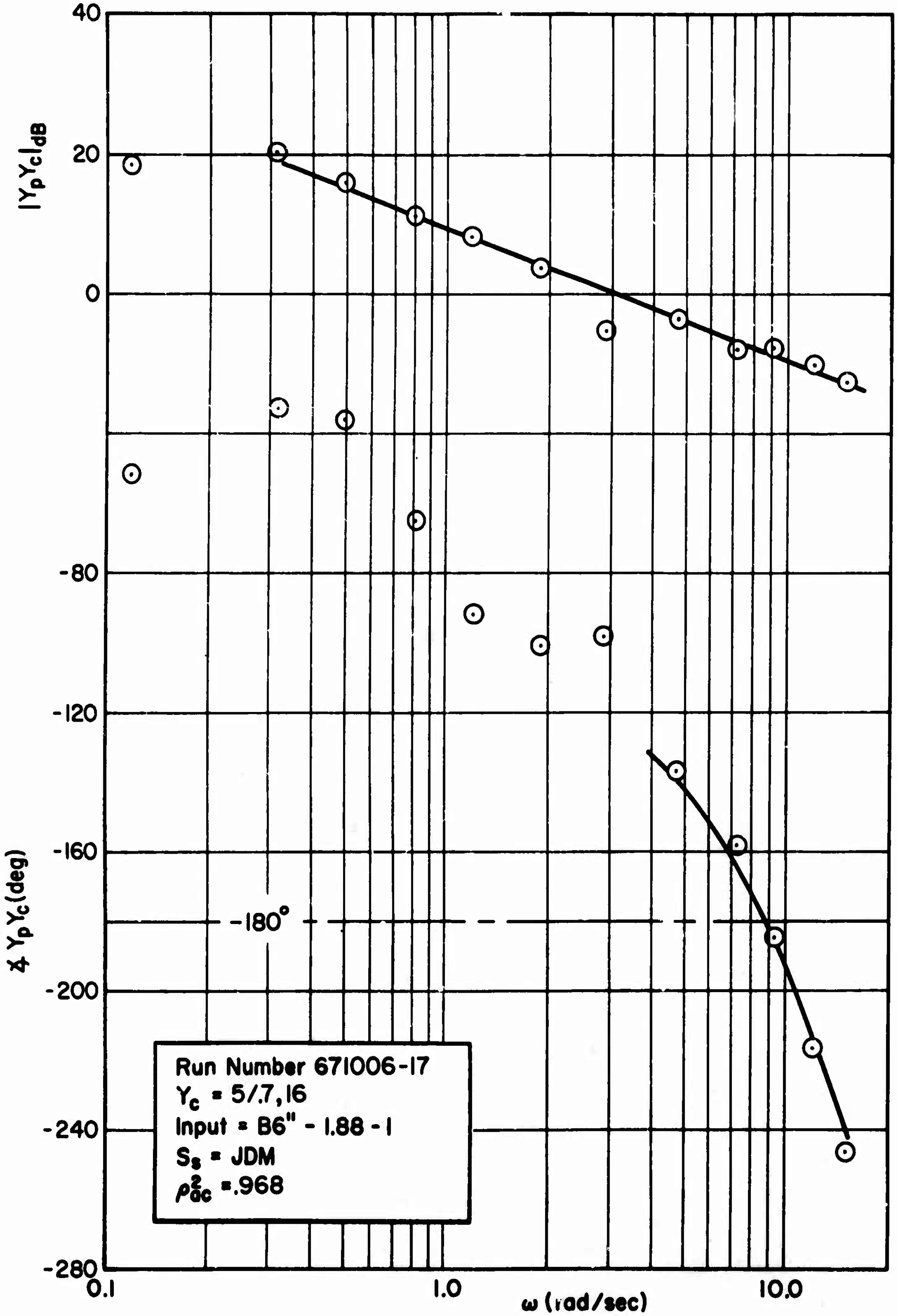


20
0
-20
-40
-80
-120
-160
-200
-240
-280
|Yp Yc|dB
∢ Yp Yc (deg)
-180°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671006-09
Yc = 10/.7,16
Input = B6" - 1.88 - 1
Ss = JDM
ρ²ac = .94

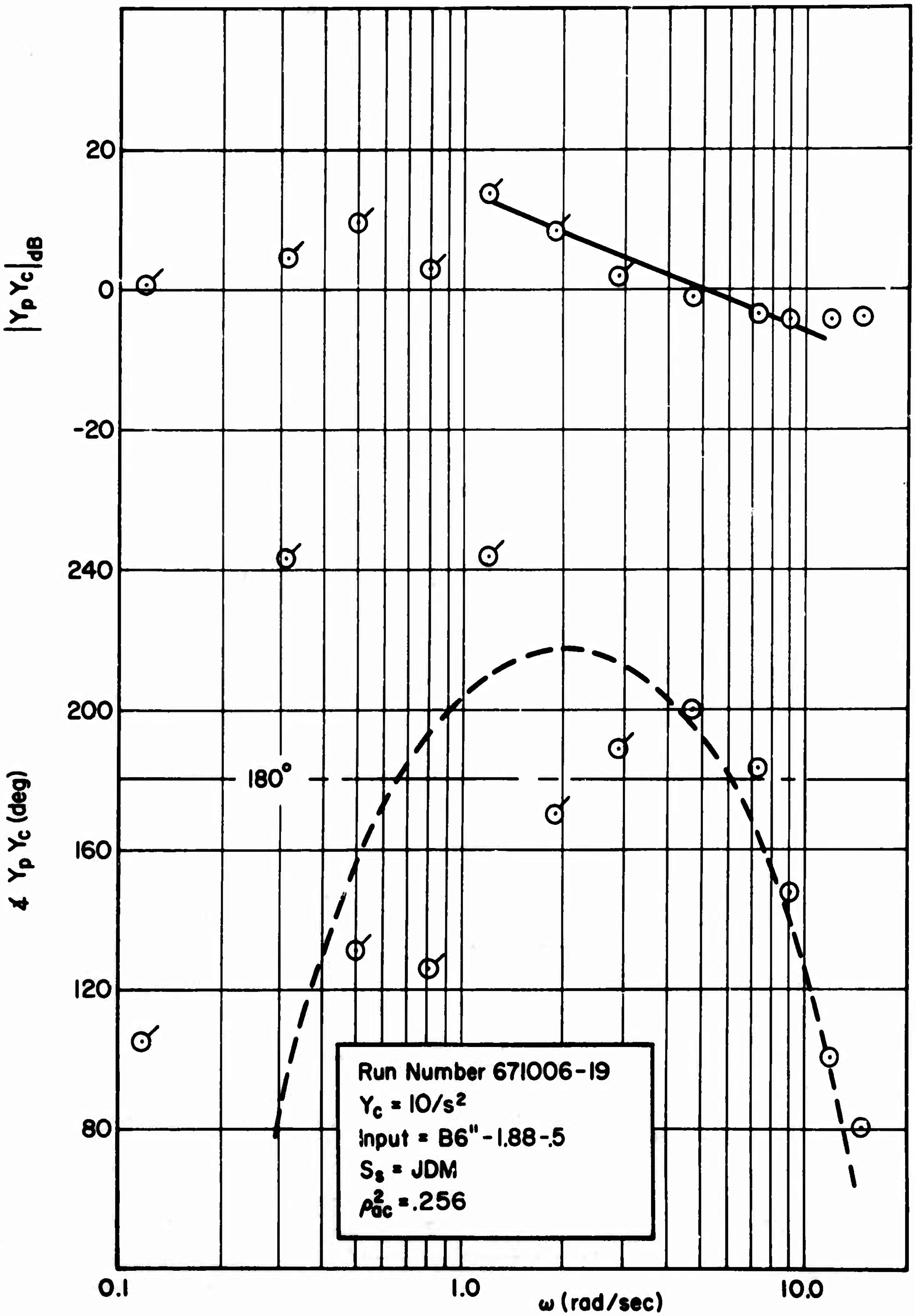

40
20
0
-20
-80
-120
-160
-200
-240
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
-180°
Run Number 671006-11
Y_c = 10/s
Input = B6" - 1.88 - 1.5
S_s = JDM
ρ^2_{ac} = .702
0.1
1.0
10.0
ω (rad/sec)


40
20
0
-20
-80
-120
-160
-200
-240
|Yp Yc|dB
∡ Yp Yc (deg)
-180°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671006-13
Yc = 10/s
Input = B6"-1.88-.5
Ss = JDM
ρ²ac = .521

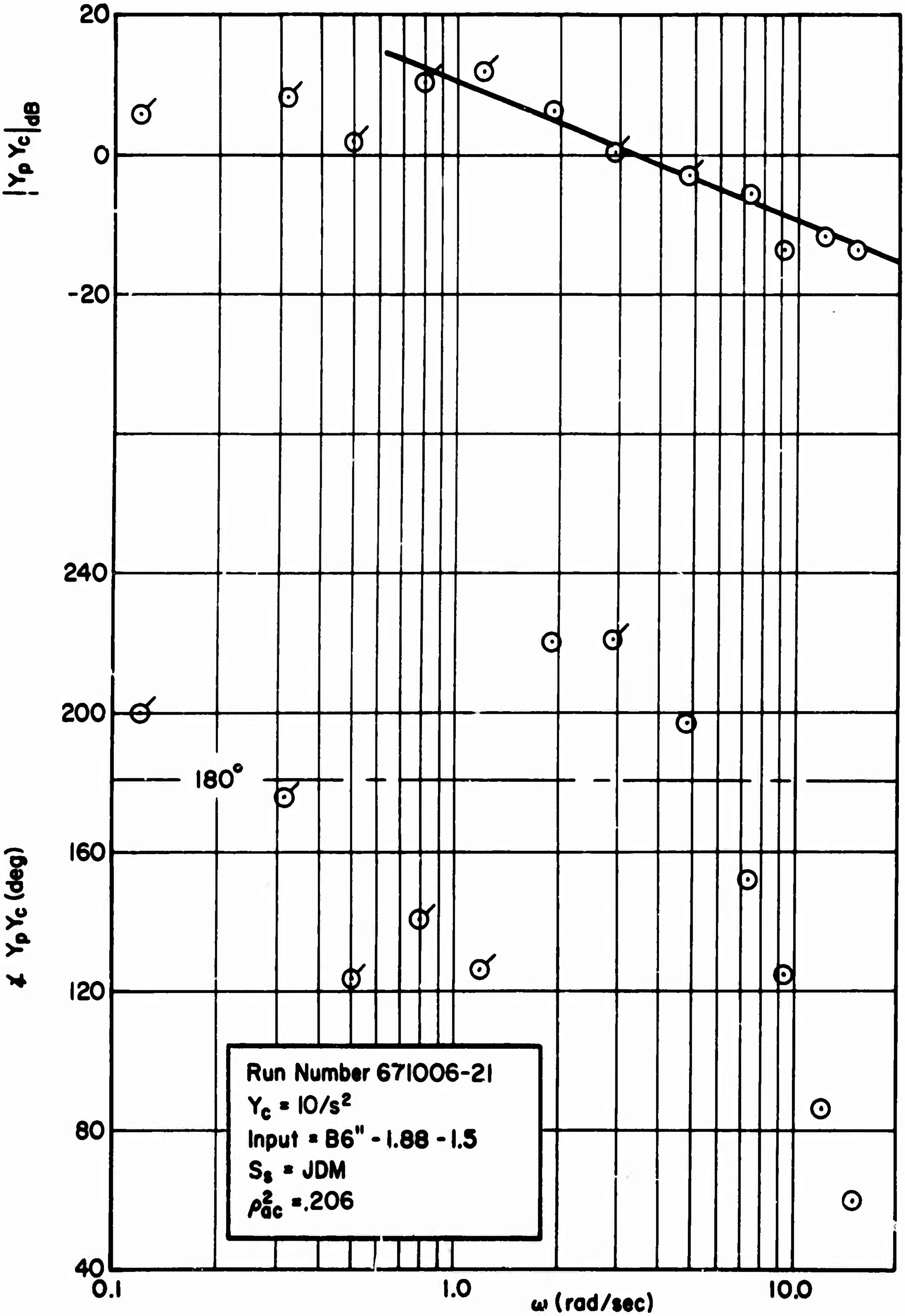

40
20
0
-120
-160
-180°
-200
-240
-280
-320
-360
0.1
1.0
10.0
ω (rad/sec)
|Yp Yc|dB
∡ Yp Yc (deg)
Run Number 671006-15
Yc = 10/s
Input = B6" - 1.88 - 1
Ss = JDM
ρ²ac = .584



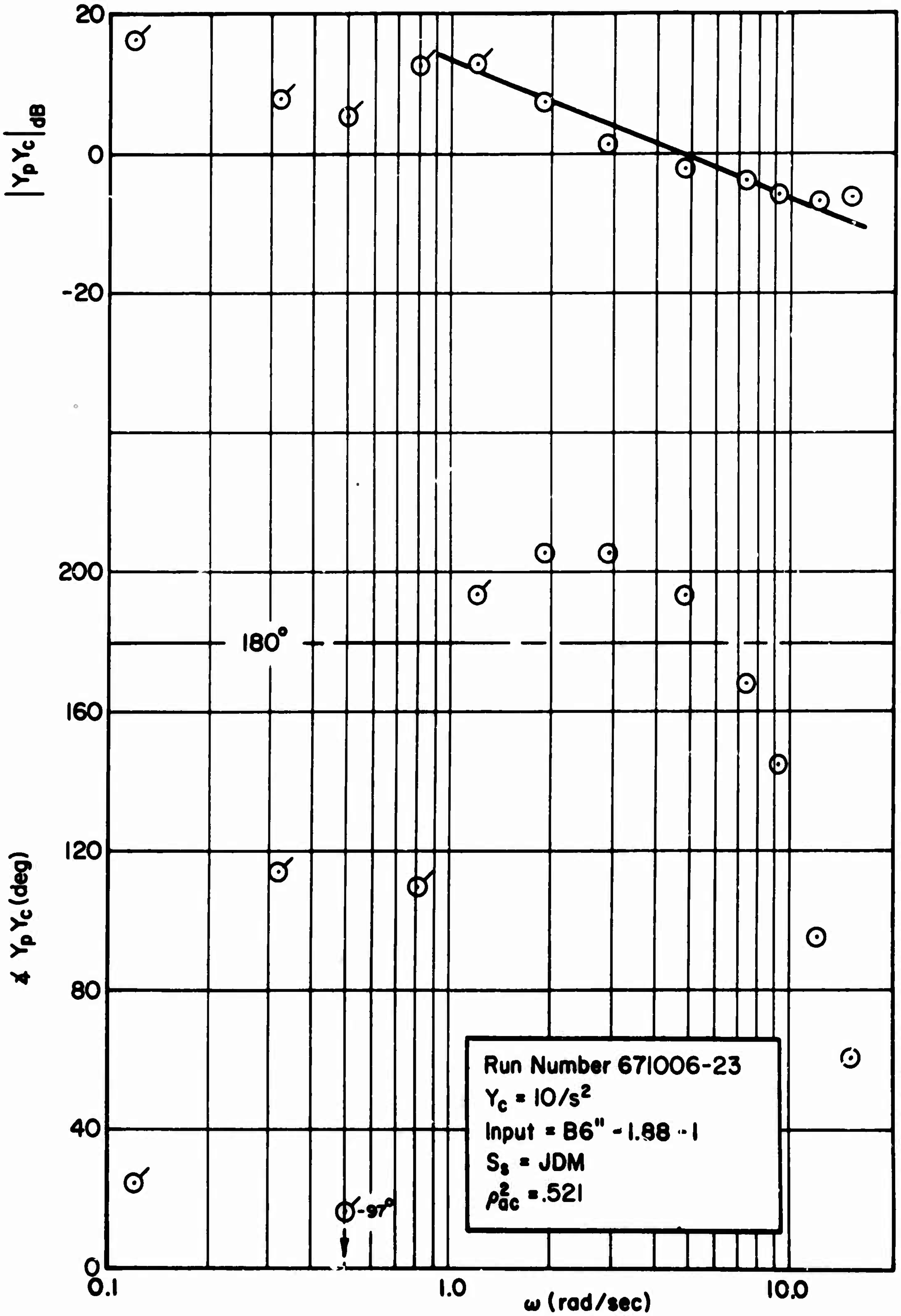
40
20
0
-80
-120
-160
-200
-240
-280
$|Y_pY_c|_{dB}$
$\measuredangle Y_pY_c$ (deg)
-180°
Run Number 671006-17
Y_c = 5/.7,16
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} =.968
0.1
1.0
10.0
ω (rad/sec)



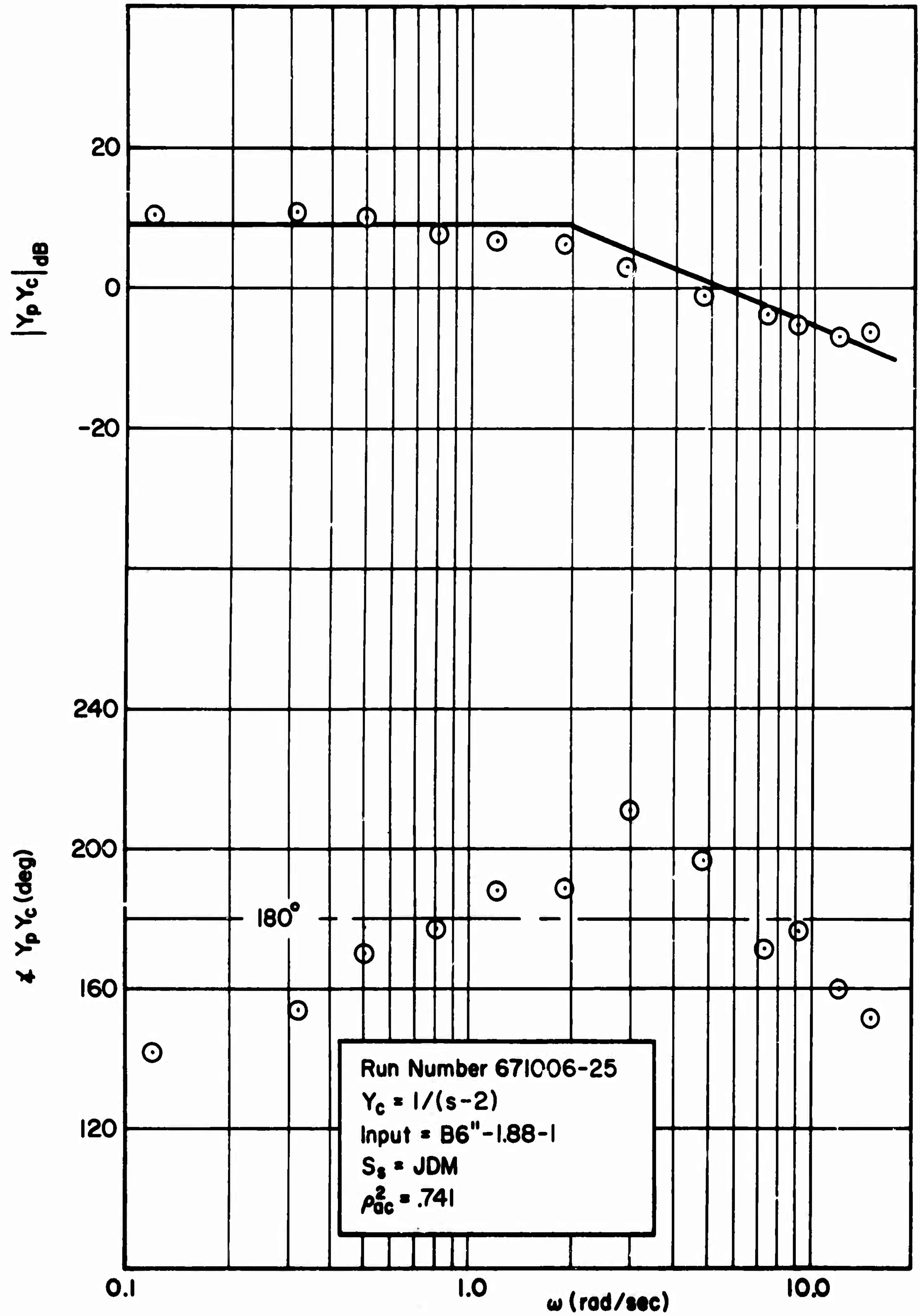
20
0
-20
240
200
180°
160
120
80
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
Run Number 671006-19
$Y_c = 10/s^2$
Input = B6"-1.88-.5
S_s = JDM
ρ_{ac}^2 = .256
0.1
1.0
10.0
ω (rad/sec)



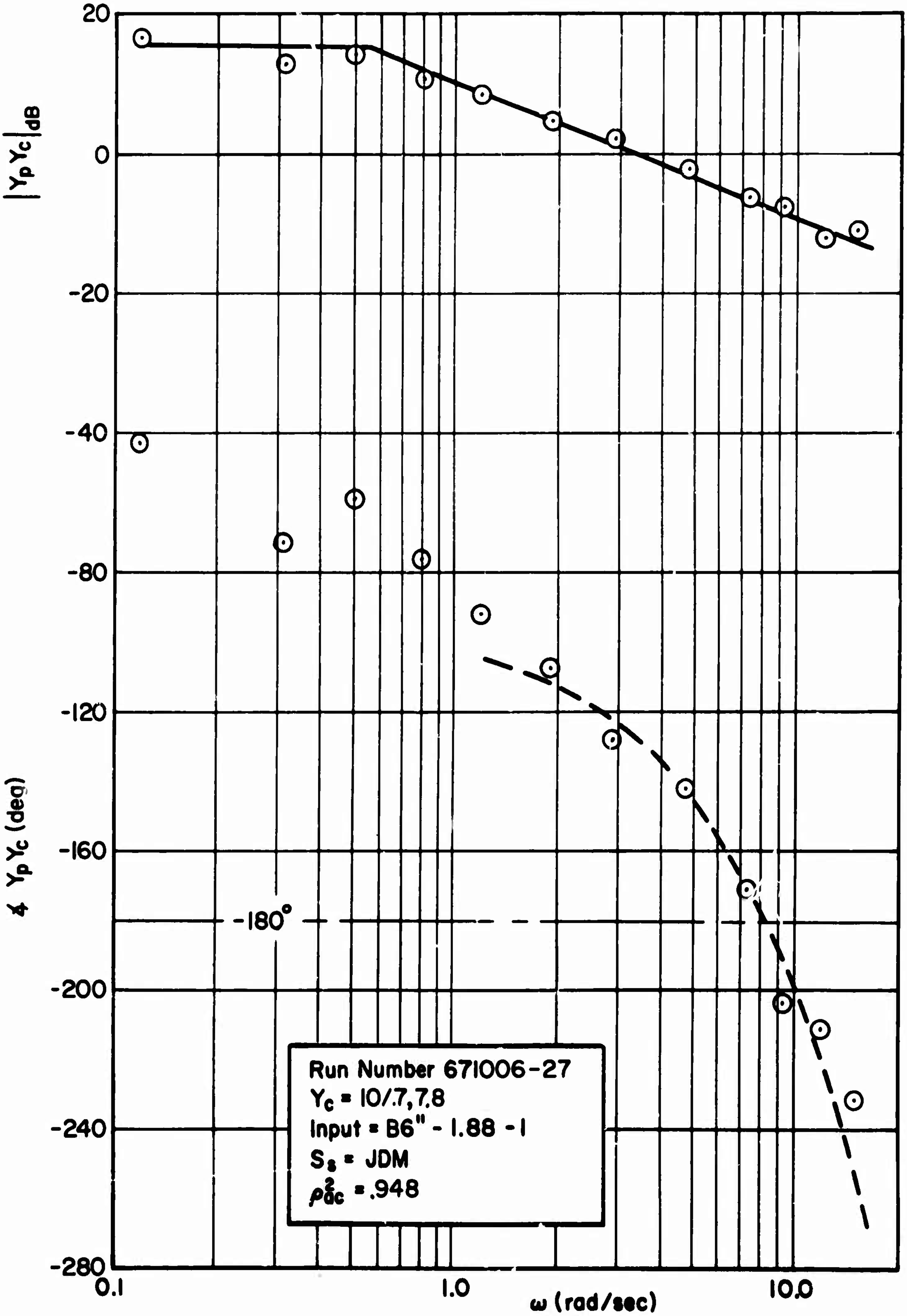
20
0
-20
240
200
180°
160
120
80
40
$|Y_p Y_c|_{dB}$
∡ $Y_p Y_c$ (deg)
0.1
1.0
10.0
ω (rad/sec)
Run Number 671006-21
$Y_c = 10/s^2$
Input = B6" - 1.88 - 1.5
S_s = JDM
ρ^2_{ac} = .206



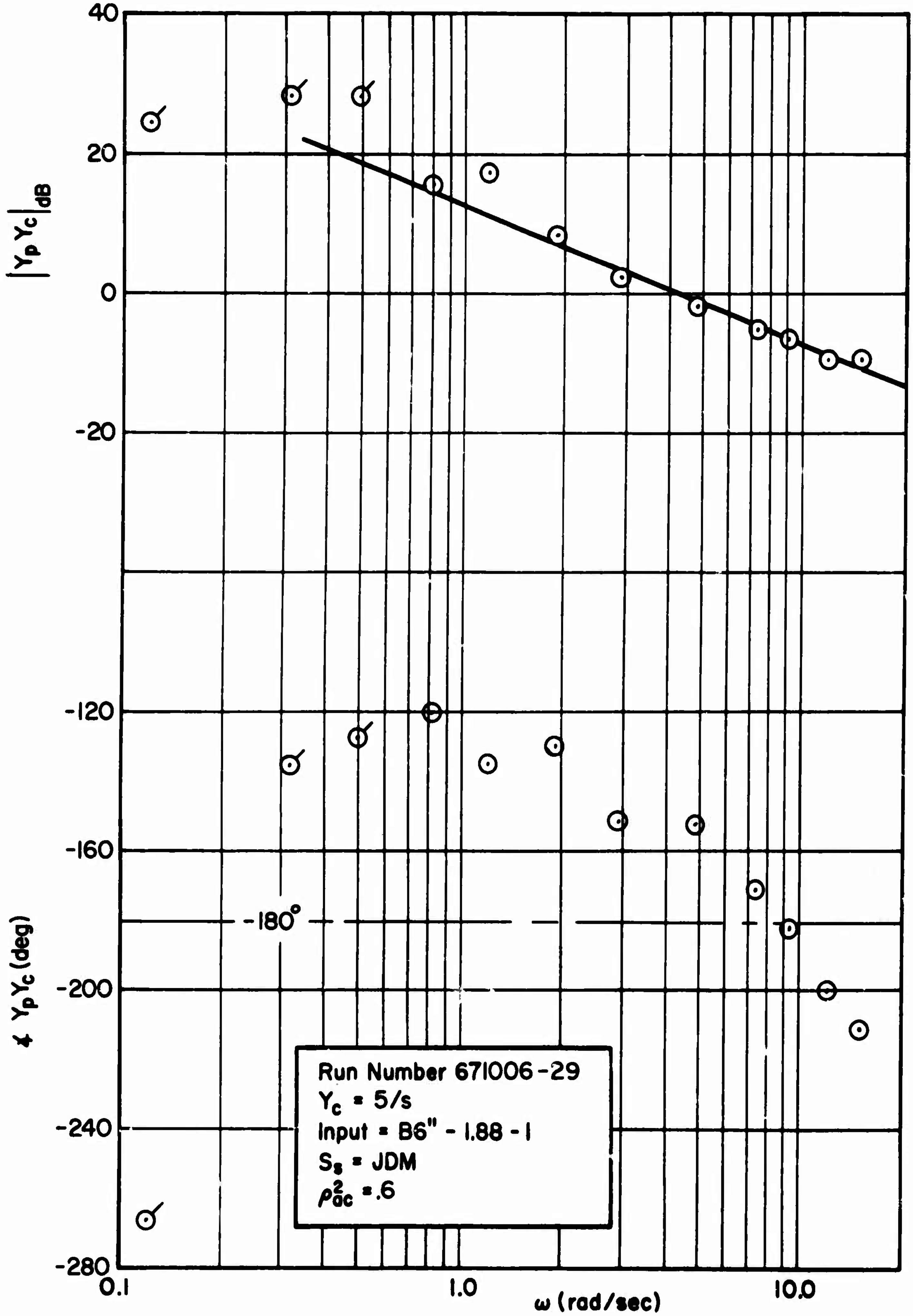
20
0
-20
200
180°
160
120
80
40
0
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
-97°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671006-23
$Y_c = 10/s^2$
Input = B6" - 1.88 - 1
S_s = JDM
ρ_{ac}^2 = .521



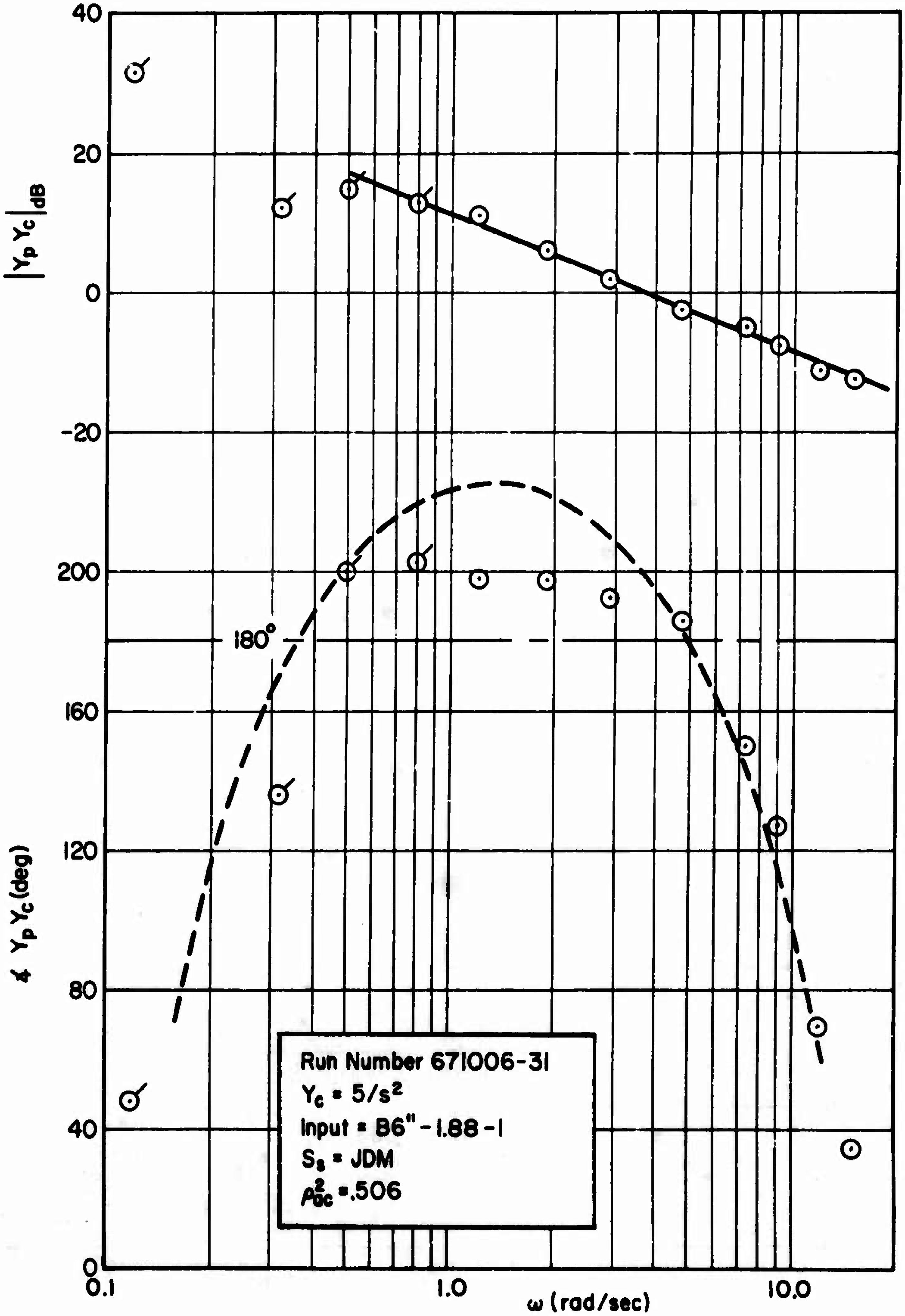
20
0
-20
240
200
180°
160
120
$|Y_p Y_c|_{dB}$
$\measuredangle\ Y_p Y_c$ (deg)
Run Number 671006-25
$Y_c = 1/(s-2)$
Input = B6"-1.88-1
S_s = JDM
$\rho^2_{ac} = .741$
0.1
1.0
10.0
ω (rad/sec)

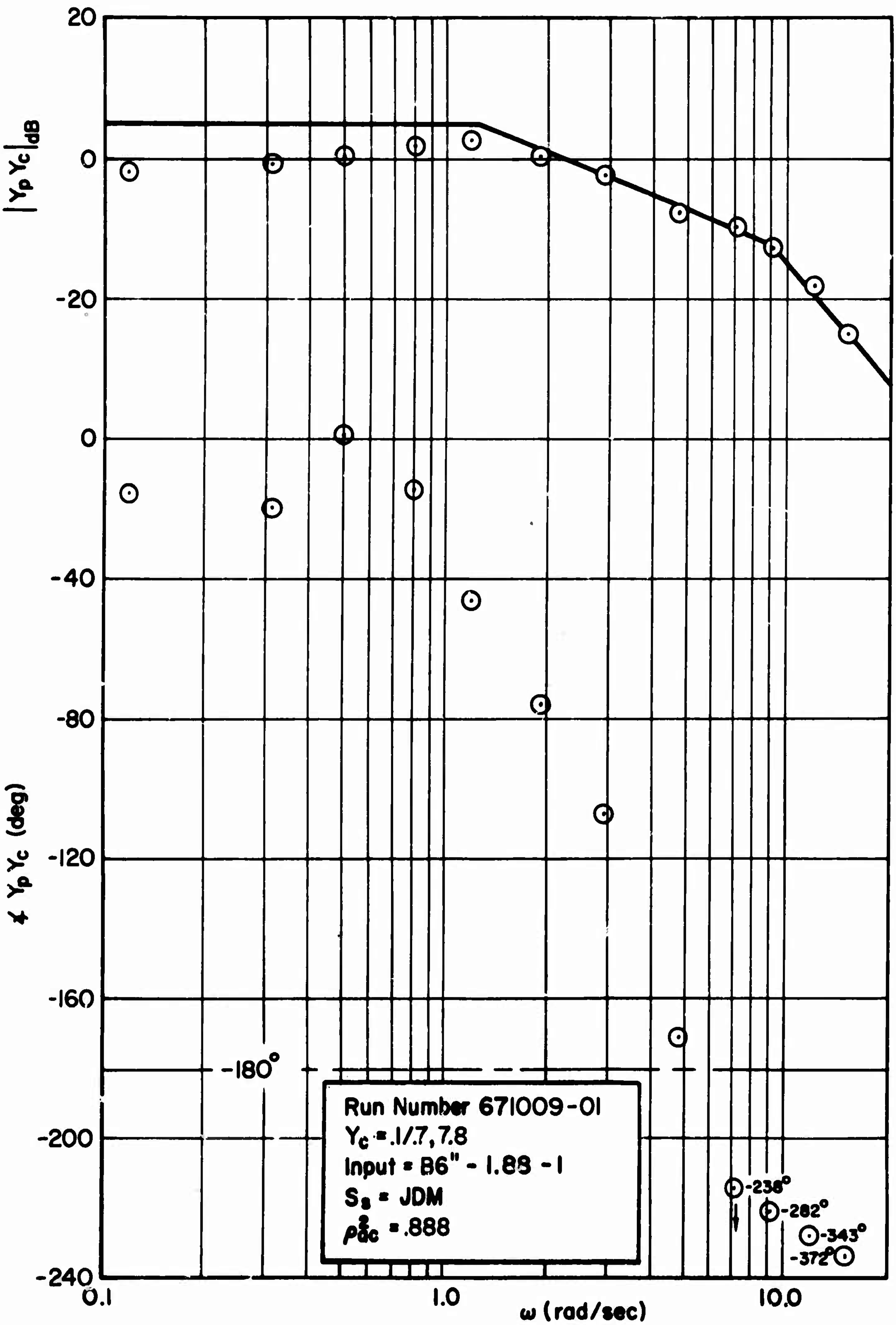


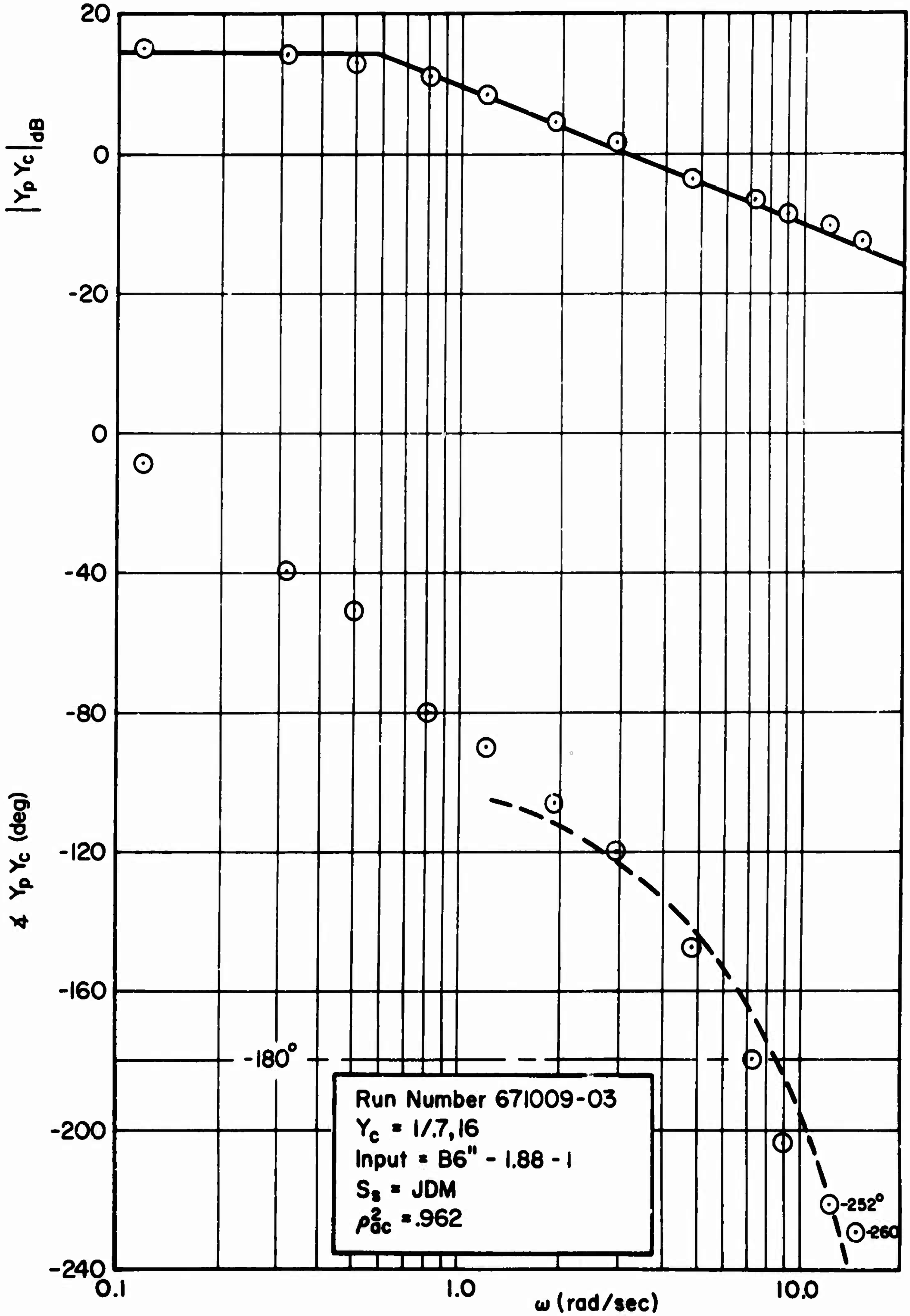
20
0
-20
-40
-80
-120
-160
-200
-240
-280
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
-180°
Run Number 671006-27
Y_c = 10/.7,7.8
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} = .948
0.1
1.0
10.0
ω (rad/sec)

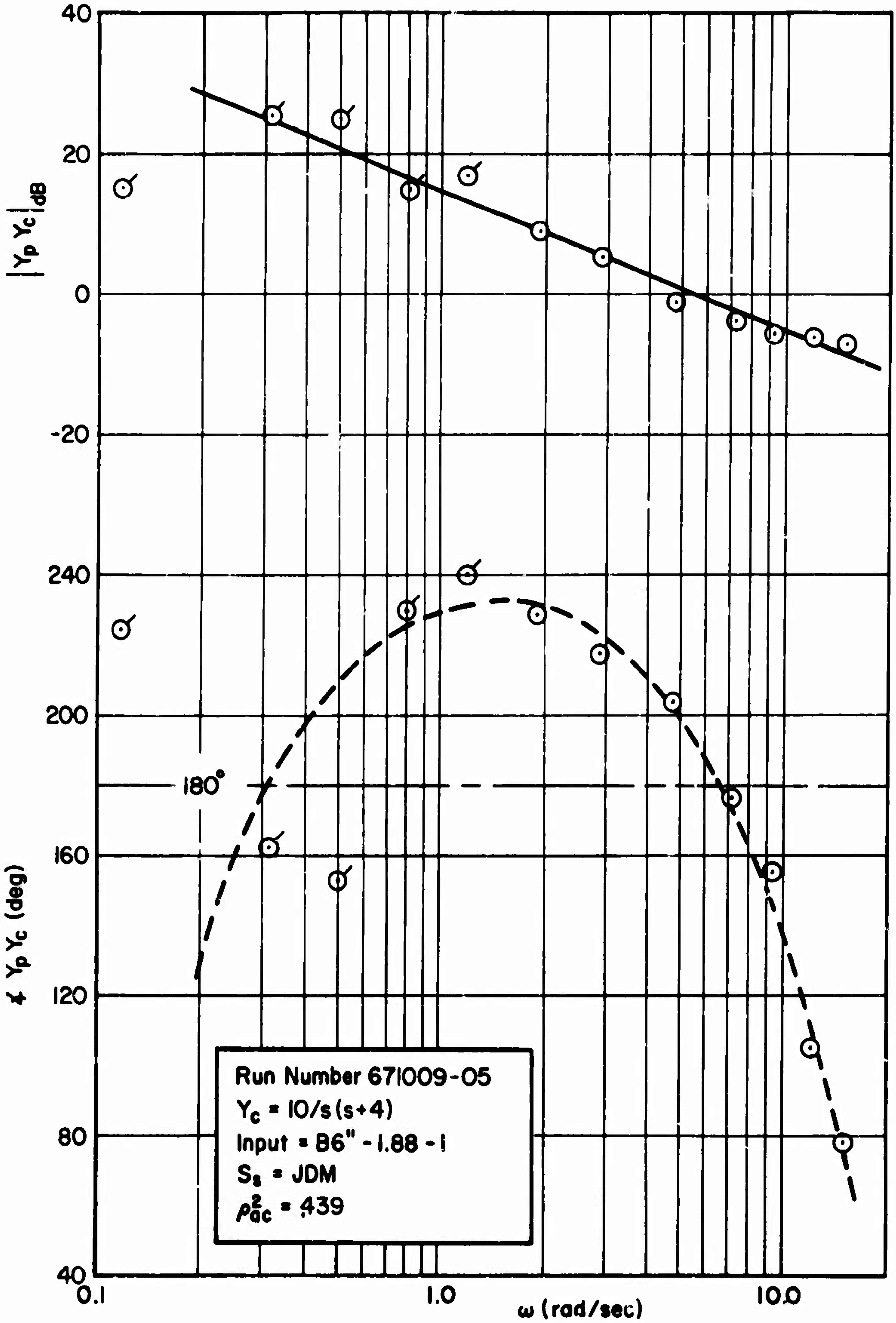


40
20
0
-20
-120
-160
-200
-240
-280
|Yp Yc|dB
∡ Yp Yc (deg)
-180°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671006-29
Yc = 5/s
Input = B6" - 1.88 - 1
Ss = JDM
ρ²ac = .6

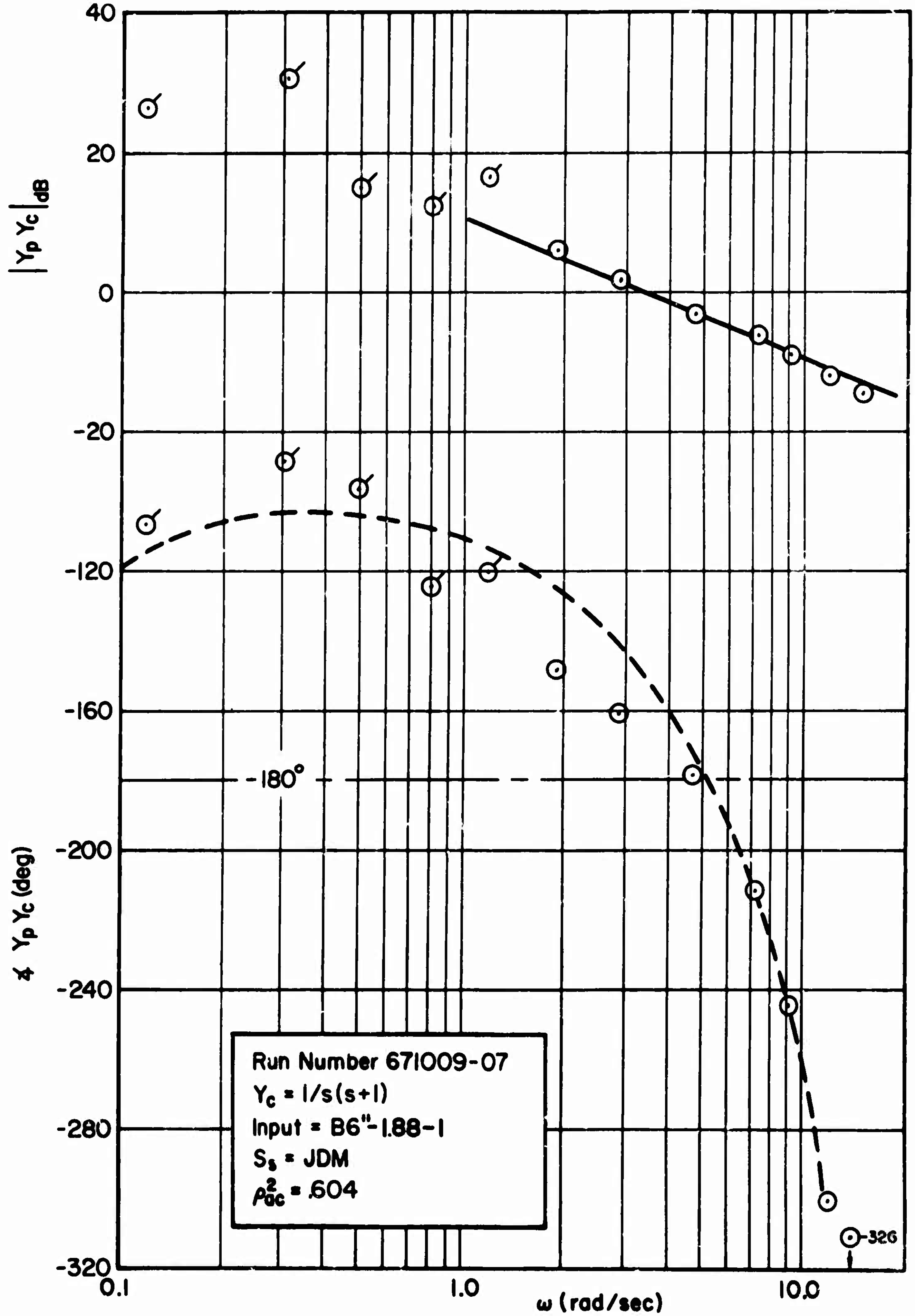

40
20
0
-20
200
180°
160
120
80
40
0
$|Y_p Y_c|_{dB}$
∢ $Y_p Y_c$ (deg)
0.1
1.0
10.0
ω (rad/sec)
Run Number 671006-31
$Y_c = 5/s^2$
Input = B6"-1.88-1
S_s = JDM
ρ^2_{ac} =.506

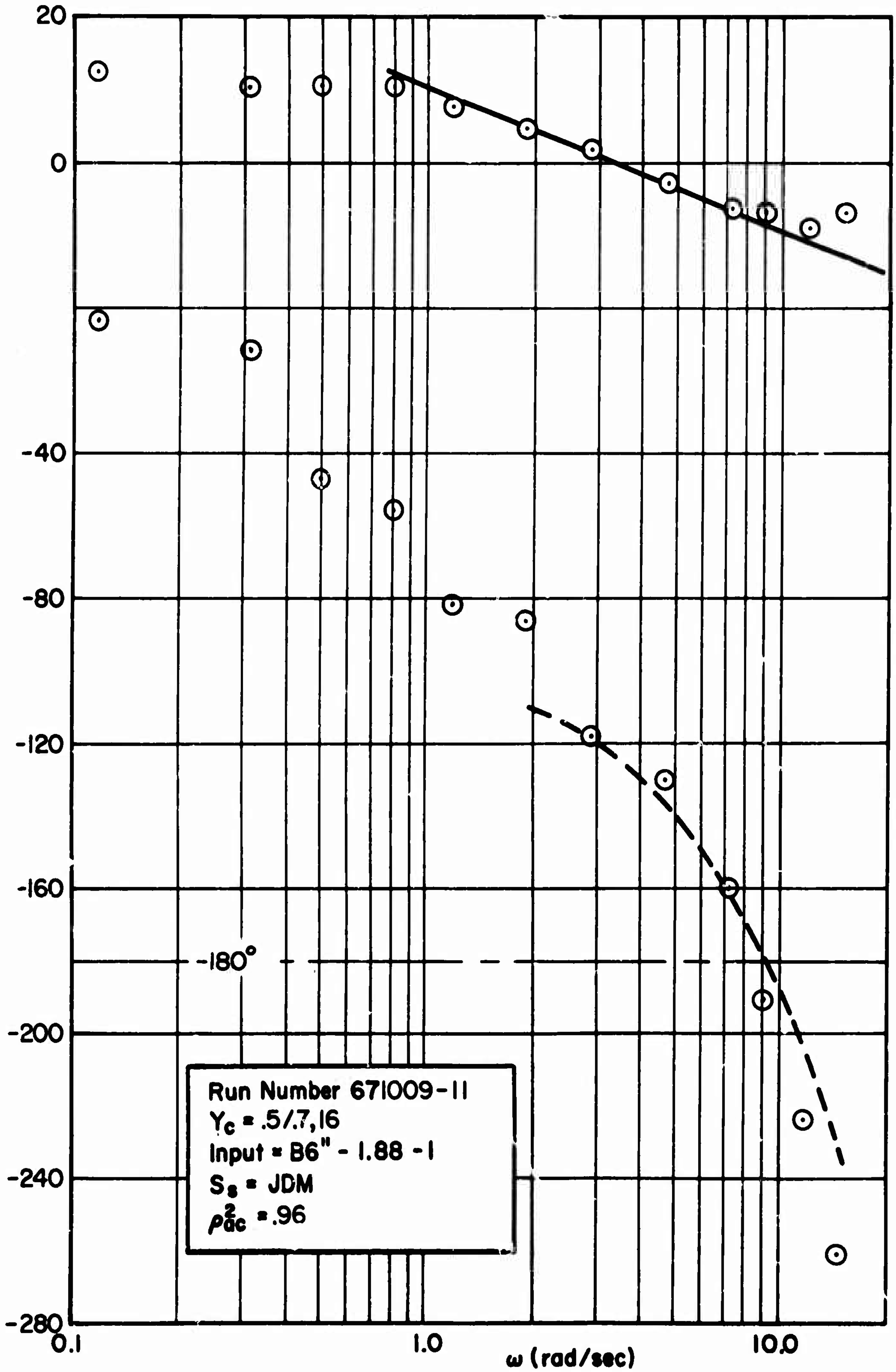

20
0
-20
0
-40
-80
-120
-160
-200
-240
|Yp Yc|dB
∡ Yp Yc (deg)
-180°
Run Number 671009-01
Yc = .1/7.7, 7.8
Input = B6" - 1.88 - 1
S8 = JDM
ρ²ac = .888
-236°
-282°
-343°
-372°
0.1
1.0
10.0
ω (rad/sec)


20
0
-20
$|Y_p Y_c|_{dB}$
0
-40
-80
-120
-160
-200
-240
$\measuredangle\ Y_p Y_c$ (deg)
-180°
-252°
-260
0.1
1.0
10.0
ω (rad/sec)
Run Number 671009-03
Y_c = 1/.7,16
Input = B6" - 1.88 - 1
S_s = JDM
ρ_{ac}^2 = .962

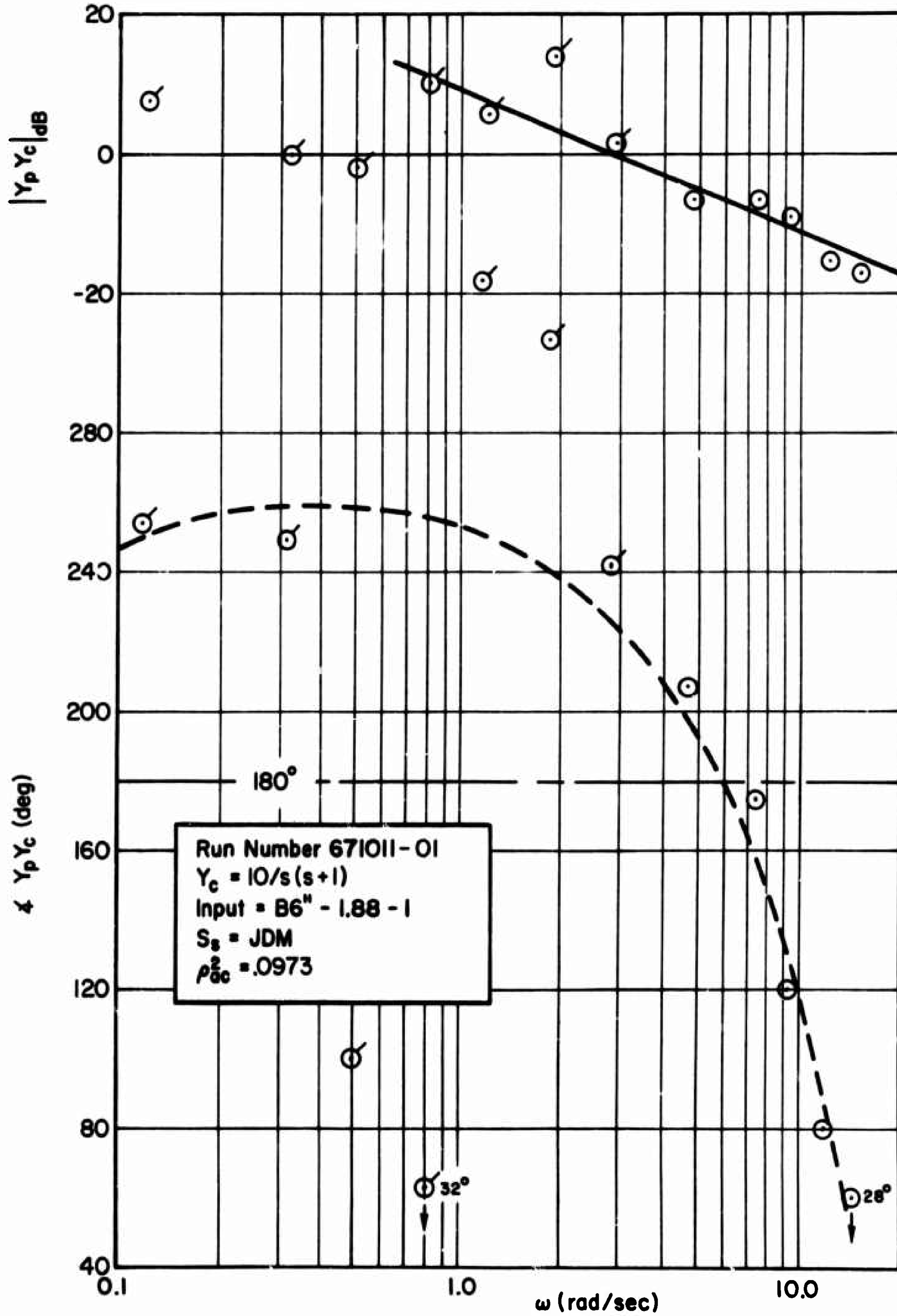


40
20
0
-20
240
200
180°
160
120
80
40
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
0.1
1.0
10.0
ω (rad/sec)
Run Number 671009-05
Y_c = 10/s(s+4)
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} = .439

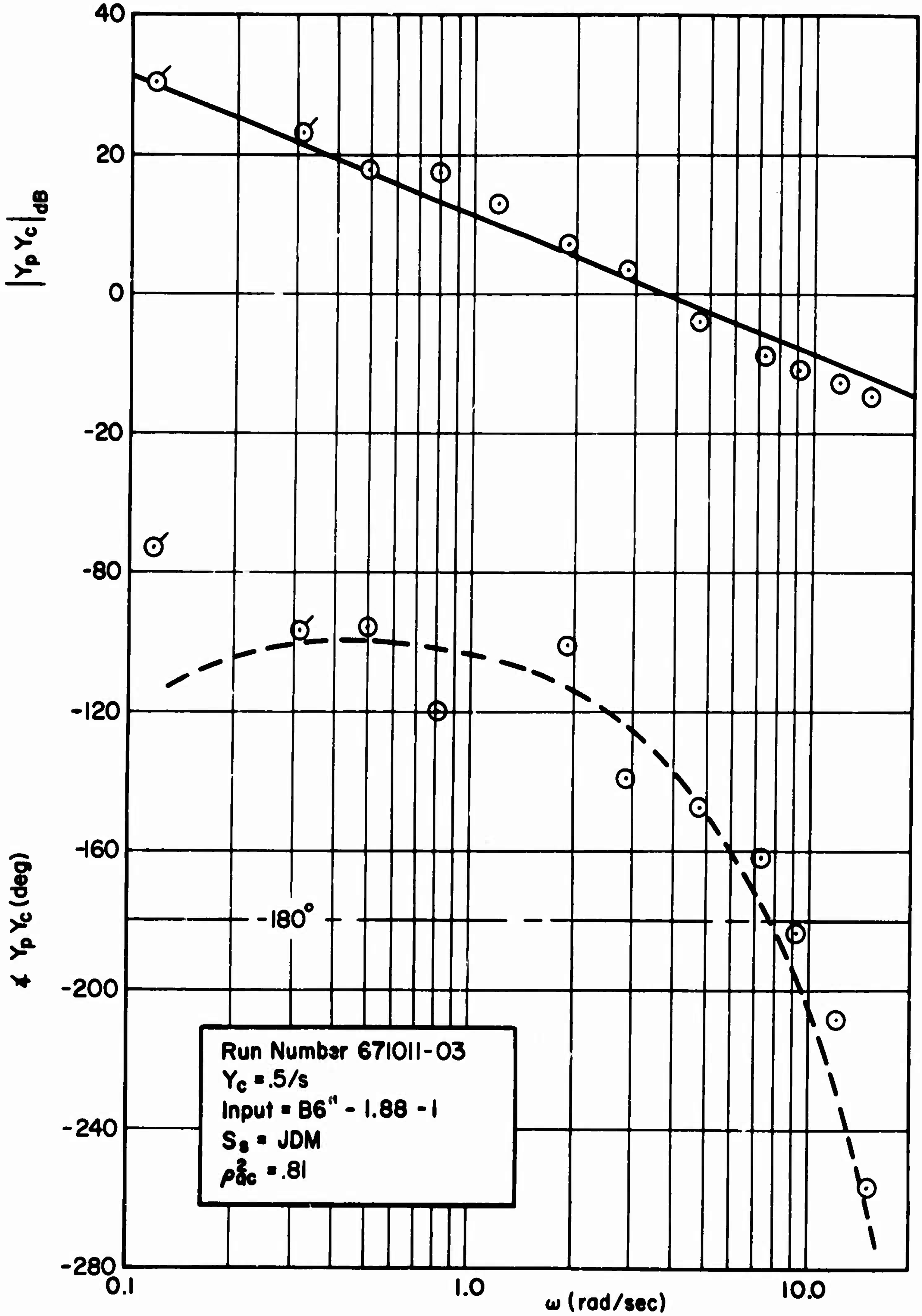

40
20
0
-20
-120
-160
-180°
-200
-240
-280
-320
$|Y_p Y_c|_{dB}$
∡ $Y_p Y_c$ (deg)
0.1
1.0
10.0
ω (rad/sec)
-326
Run Number 671009-07
Y_c = 1/s(s+1)
Input = B6"-1.88-1
S_s = JDM
ρ^2_{ac} = .604

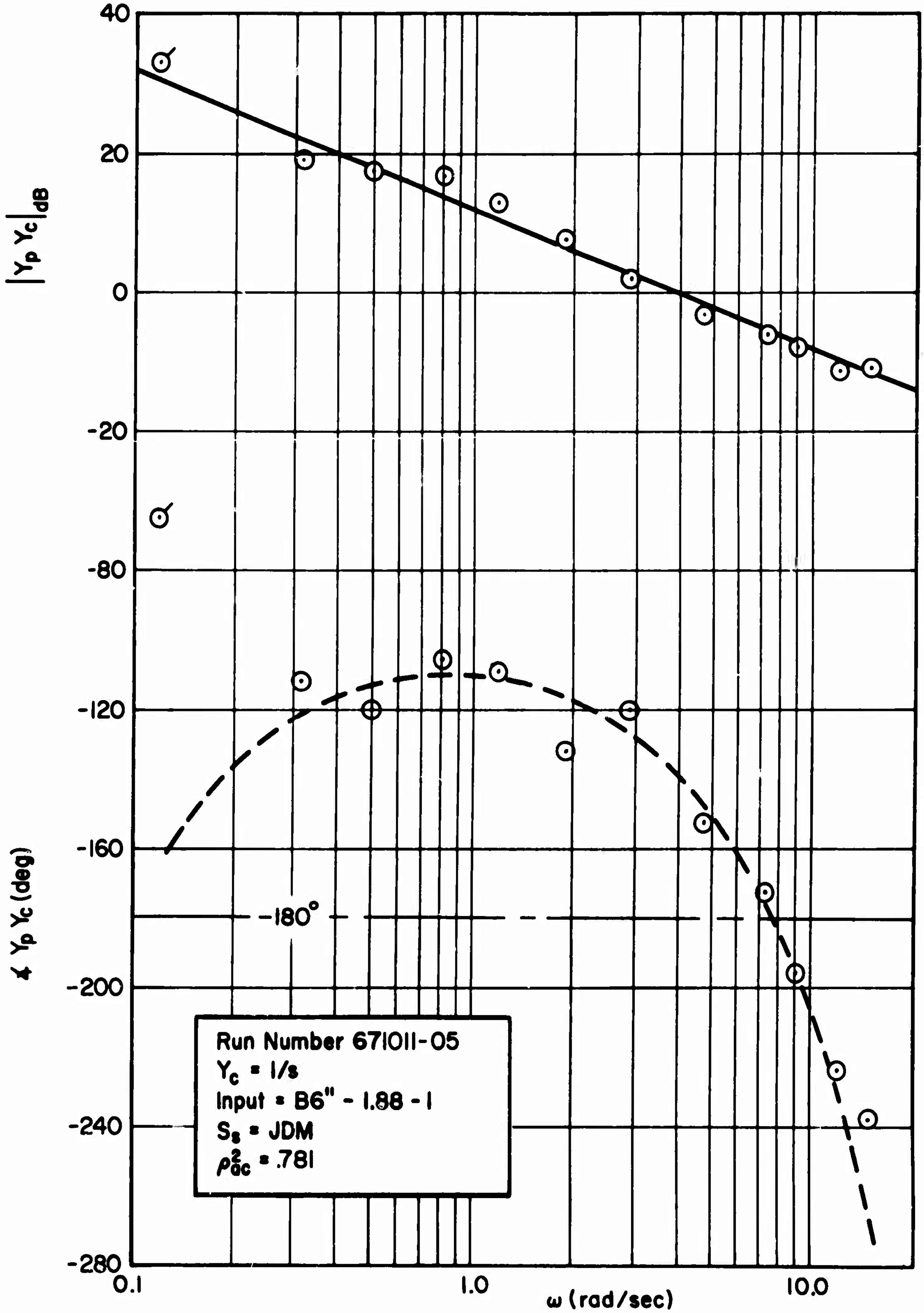


20
0
-40
-80
-120
-160
-200
-240
-280
0.1
1.0
10.0
ω (rad/sec)
|Yp Yc|dB
∢ Yp Yc (deg)
-180°
Run Number 671009-11
Yc = .5/.7,16
Input = B6" - 1.88 -1
Ss = JDM
ρ²ac =.96

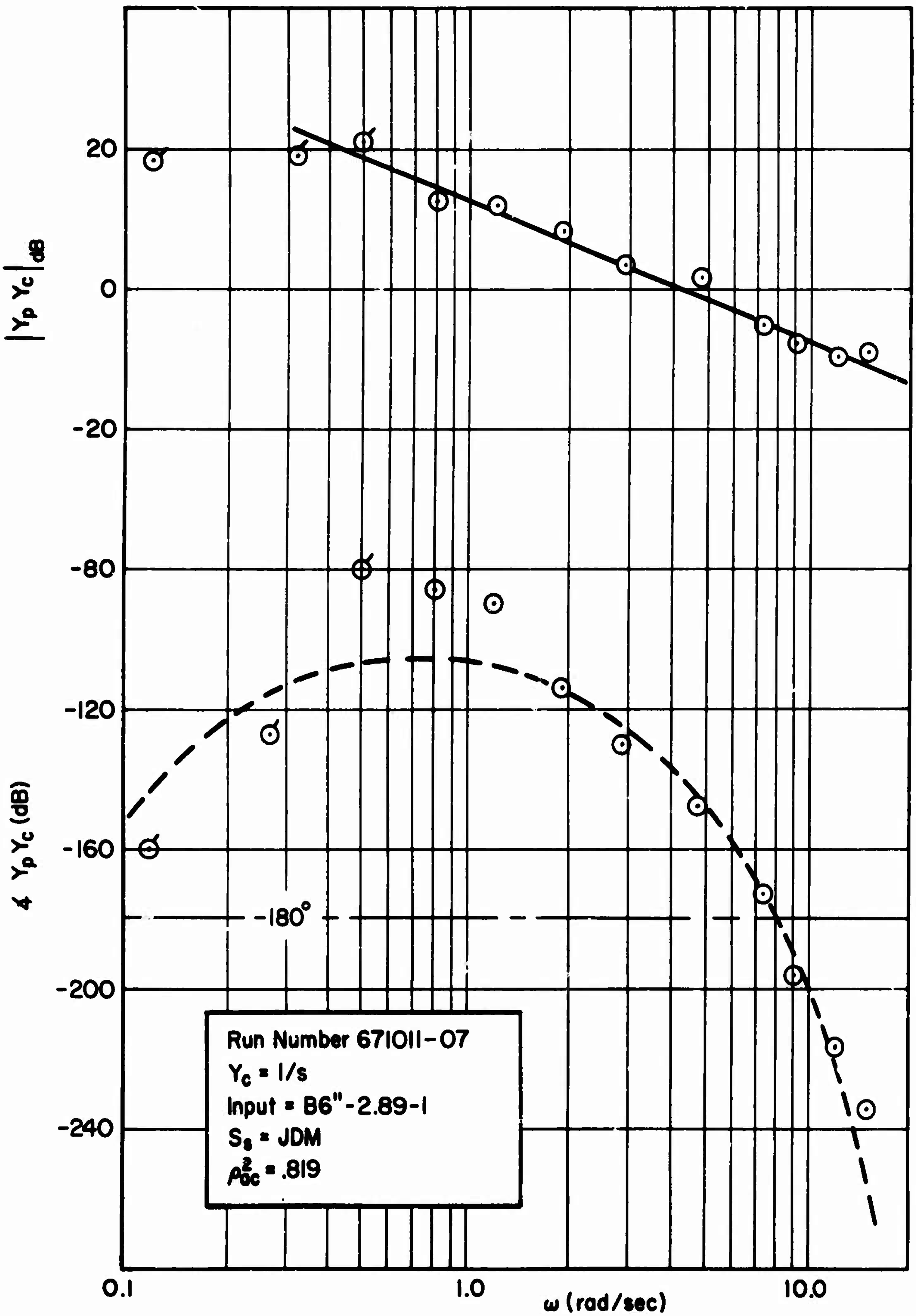


20
0
-20
280
240
200
160
120
80
40
0.1
1.0
10.0
$|Y_p Y_c|_{dB}$
$\measuredangle$ $Y_p Y_c$ (deg)
ω (rad/sec)
180°
32°
28°
Run Number 671011-01
Y_c = 10/s(s+1)
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} =.0973

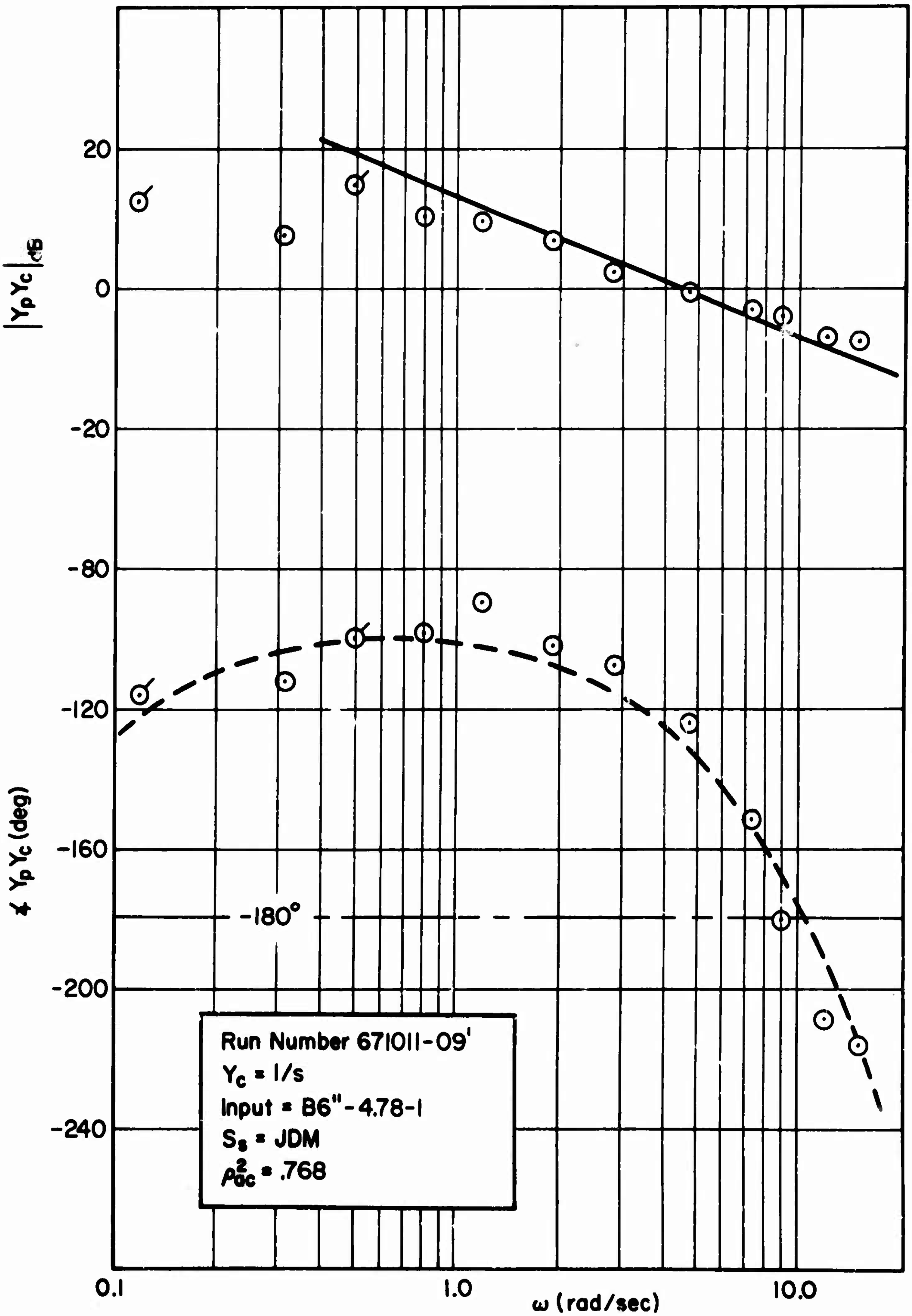

40
20
0
-20
-80
-120
-160
-200
-240
-280
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
-180°
Run Number 671011-03
Y_c = .5/s
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} = .81
0.1
1.0
10.0
ω (rad/sec)



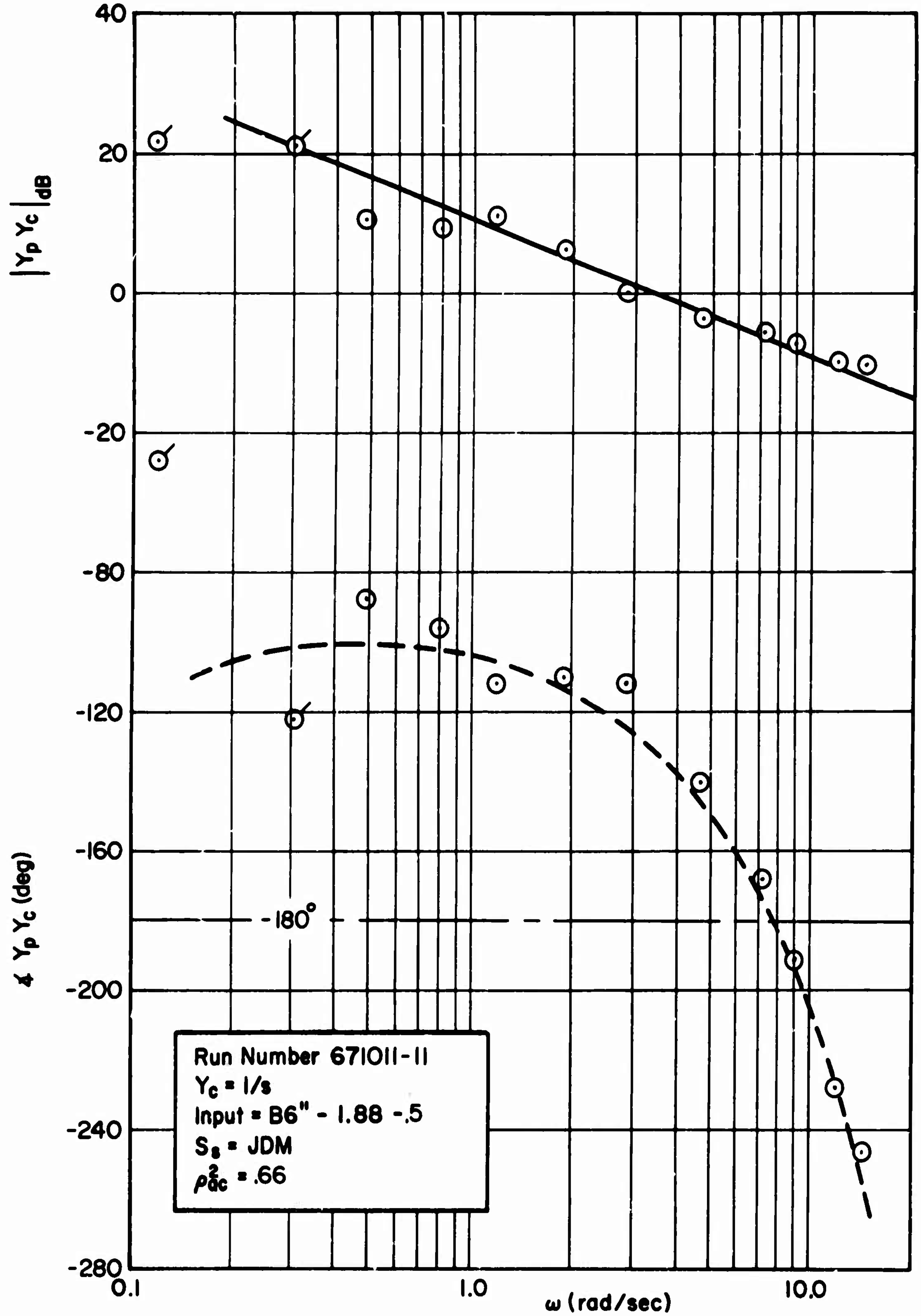
40
20
0
-20
-80
-120
-160
-200
-240
-280
$|Y_p Y_c|_{dB}$
∢ $Y_p Y_c$ (deg)
-180°
Run Number 671011-05
Y_c = 1/s
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} = .781
0.1
1.0
10.0
ω (rad/sec)



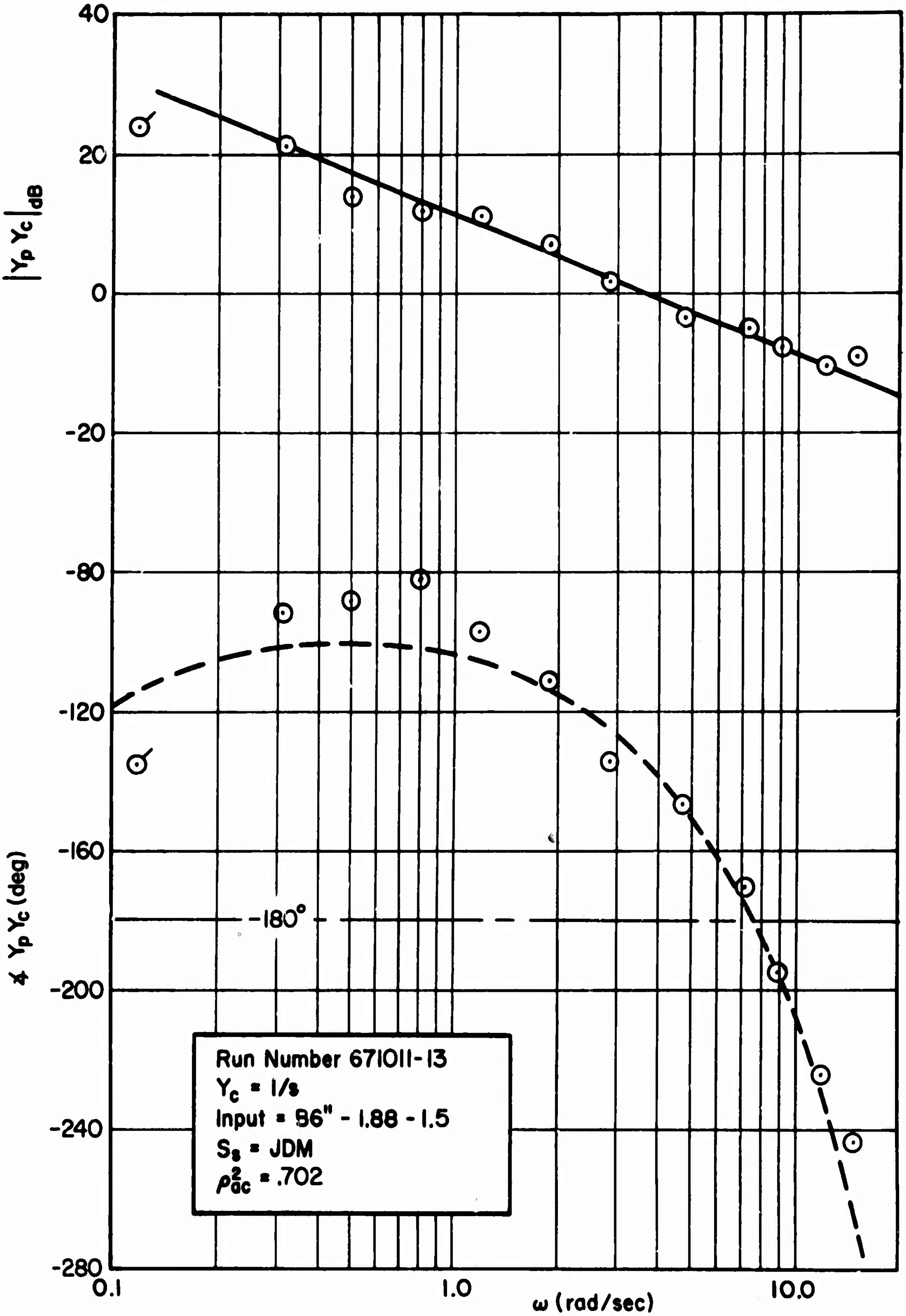
20
0
-20
-80
-120
-160
-200
-240
-180°
$|Y_p Y_c|_{dB}$
∡ $Y_p Y_c$ (dB)
Run Number 671011-07
Y_c = 1/s
Input = B6"-2.89-1
S_s = JDM
ρ^2_{ac} = .819
0.1
1.0
10.0
ω (rad/sec)

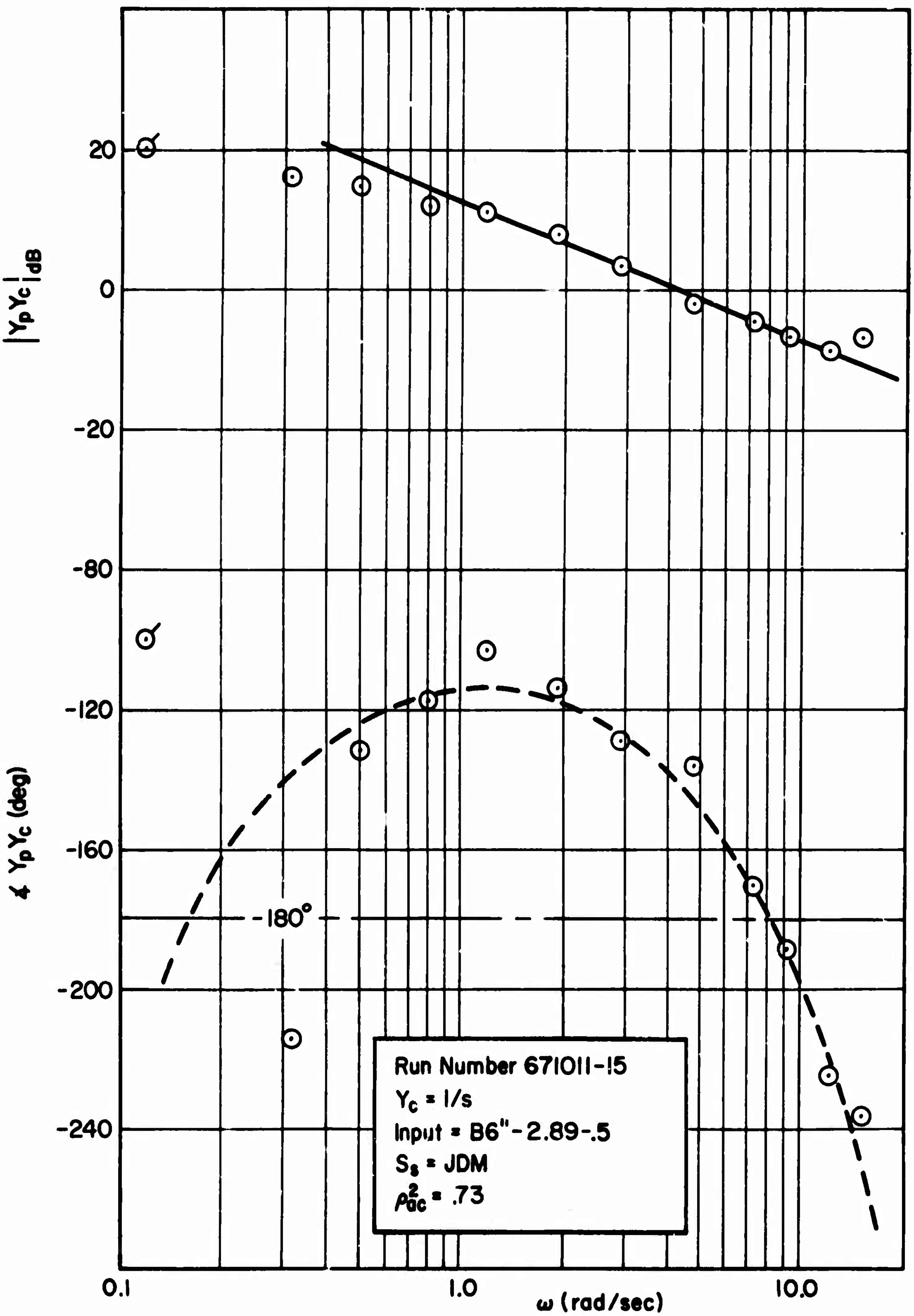


20
0
-20
-80
-120
-160
-200
-240
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
-180°
Run Number 671011-09'
Y_c = 1/s
Input = B6"-4.78-1
S_s = JDM
ρ^2_{ac} = .768
0.1
1.0
10.0
ω (rad/sec)

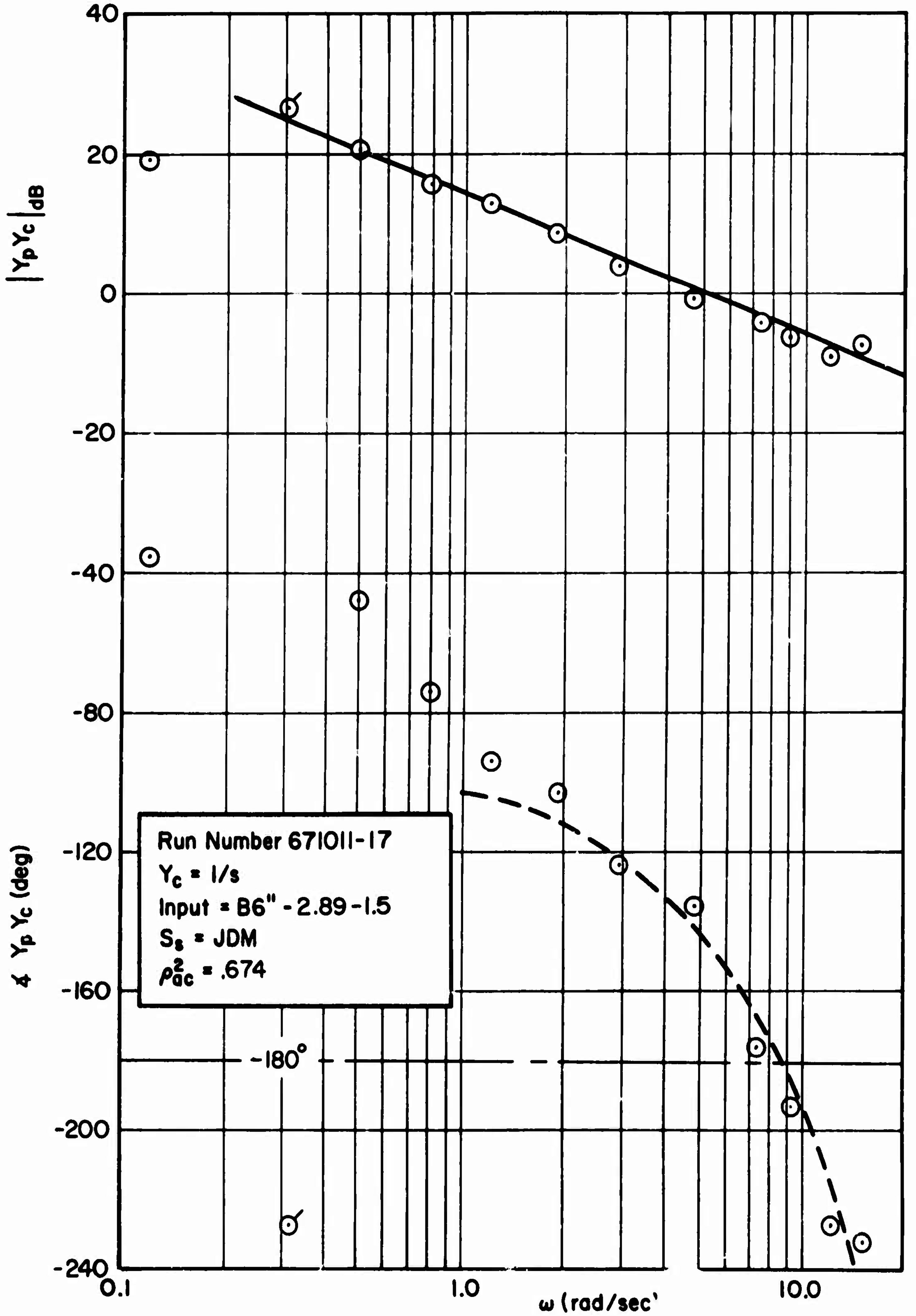


40
20
0
-20
-80
-120
-160
-200
-240
-280
0.1
1.0
10.0
ω (rad/sec)
|Yp Yc|dB
∢ Yp Yc (deg)
-180°
Run Number 671011-11
Yc = 1/s
Input = B6" - 1.88 -.5
Ss = JDM
ρ²ac = .66

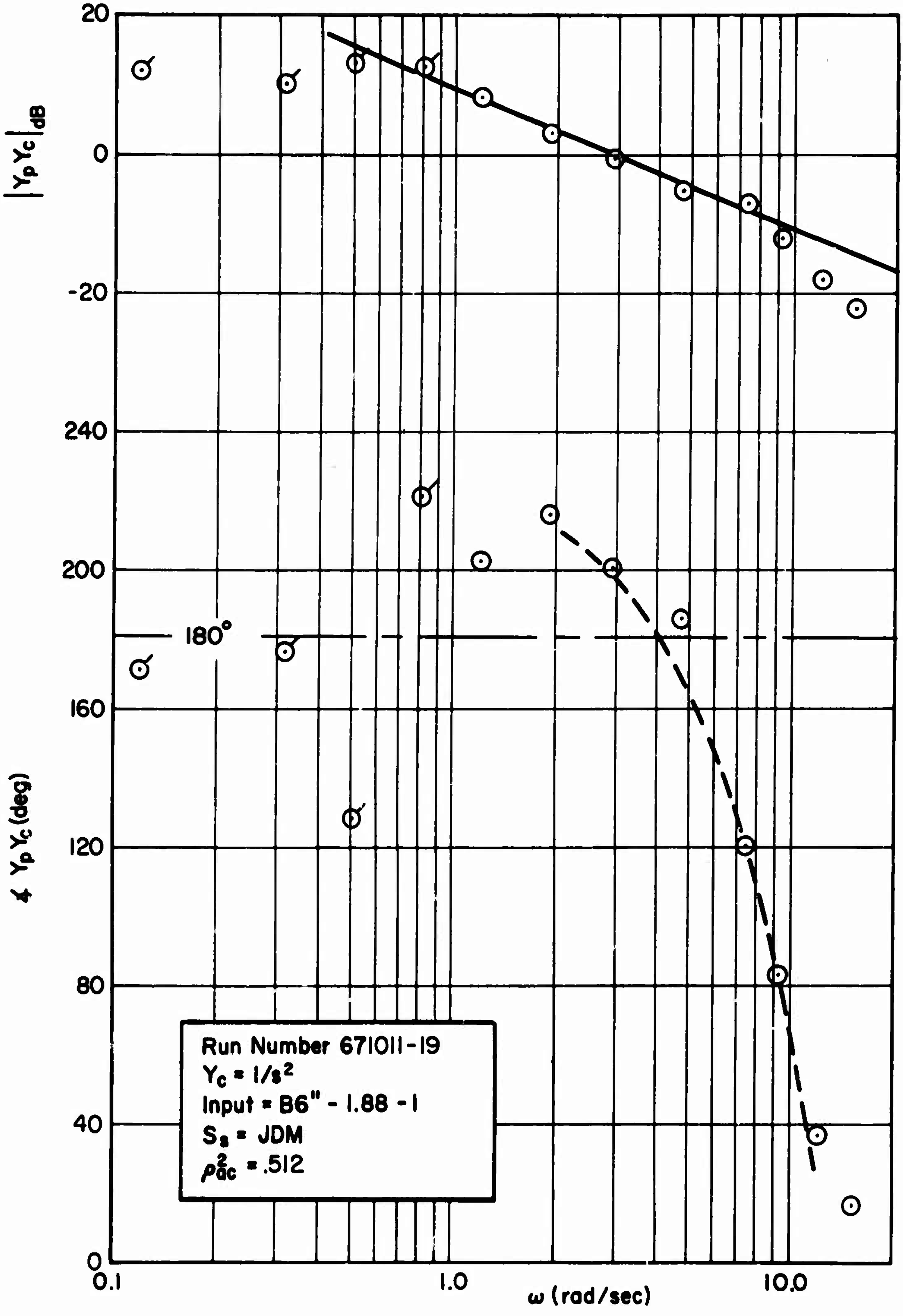

40
20
0
-20
-80
-120
-160
-200
-240
-280
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
-180°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671011-13
Y_c = 1/s
Input = 86" - 1.88 - 1.5
S_s = JDM
ρ^2_{ac} = .702

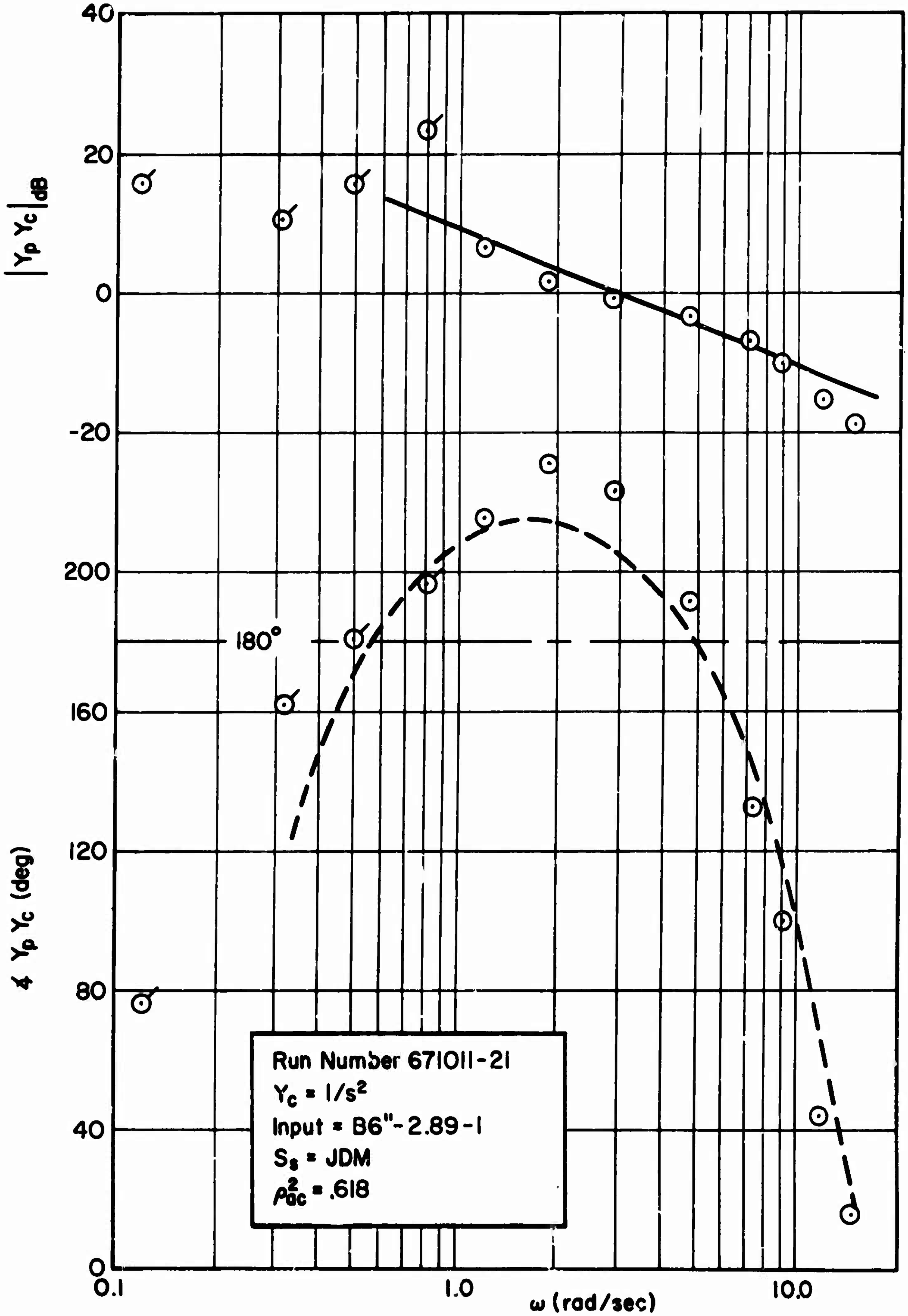


20
0
-20
-80
-120
-160
-200
-240
$|Y_p Y_c|_{dB}$
∡ $Y_p Y_c$ (deg)
-180°
Run Number 671011-15
Y_c = 1/s
Input = B6"-2.89-.5
S_s = JDM
ρ_{ac}^2 = .73
0.1
1.0
10.0
ω (rad/sec)

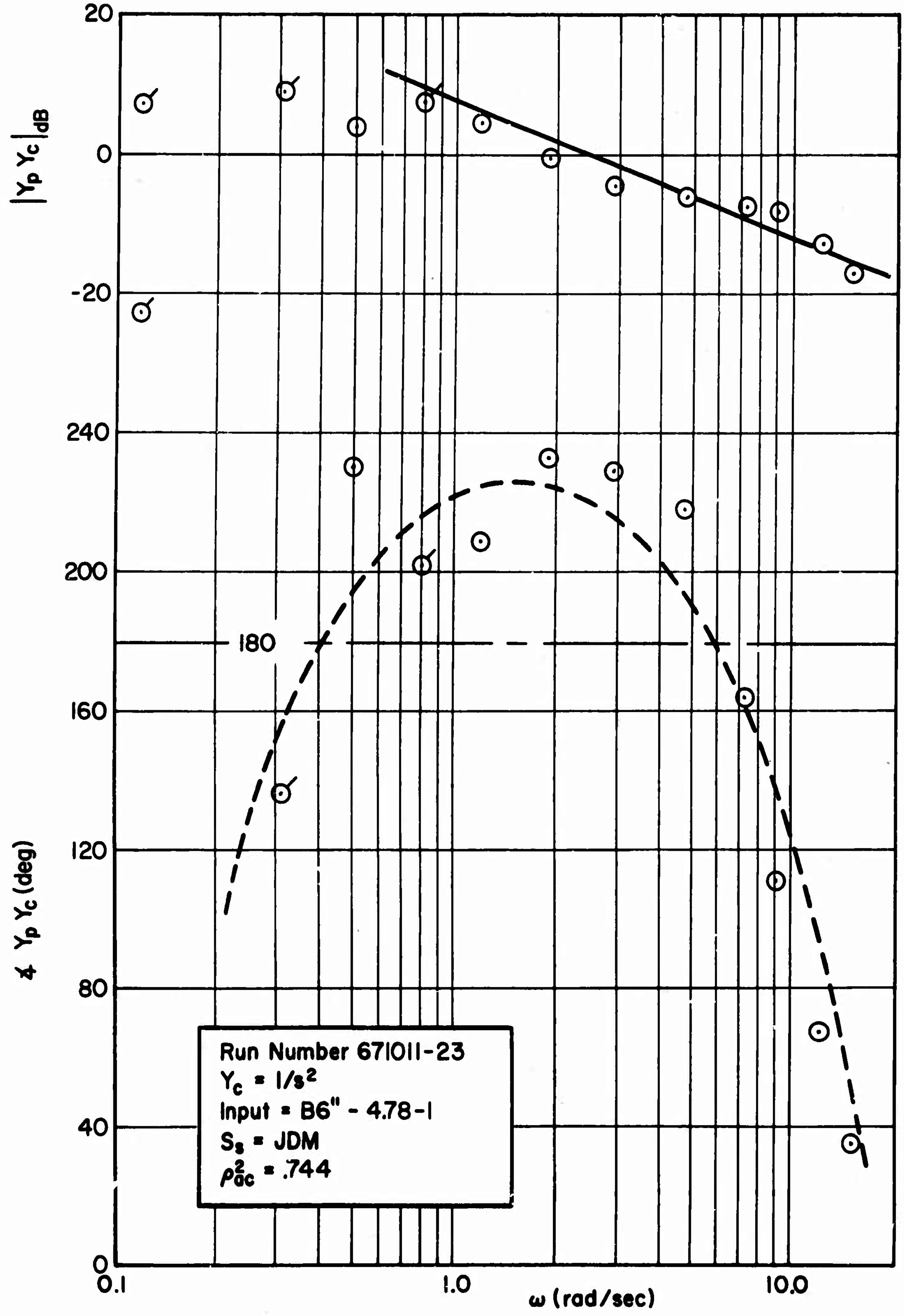


40
20
0
-20
-40
-80
-120
-160
-200
-240
$|Y_p Y_c|_{dB}$
∢ $Y_p Y_c$ (deg)
Run Number 671011-17
Y_c = 1/s
Input = B6" - 2.89 - 1.5
S_s = JDM
ρ^2_{ac} = .674
-180°
0.1
1.0
10.0
ω (rad/sec)

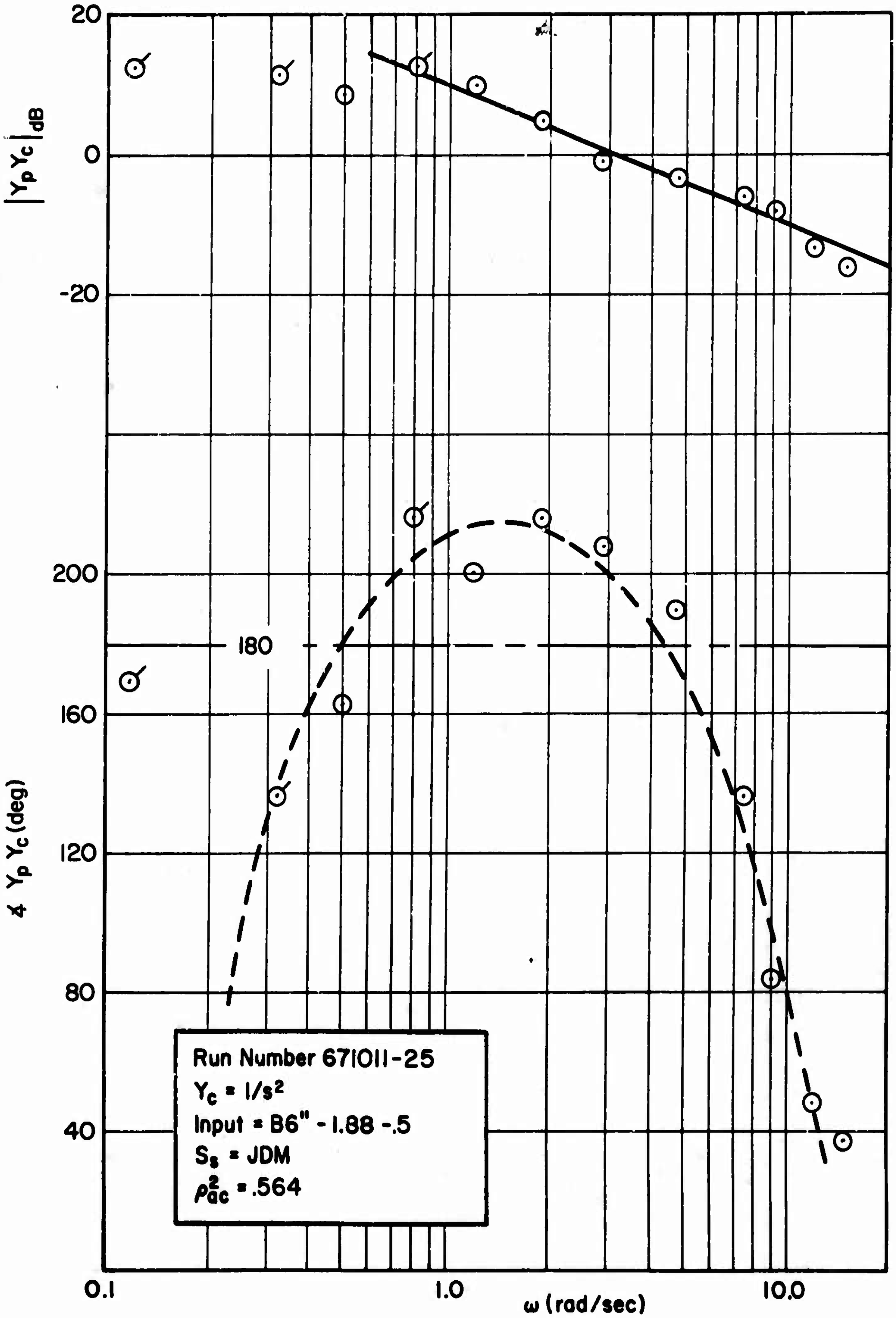

20
0
-20
240
200
180°
160
120
80
40
0
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
Run Number 671011-19
$Y_c = 1/s^2$
Input = B6" - 1.88 - 1
S_s = JDM
ρ_{ac}^2 = .512
0.1
1.0
10.0
ω (rad/sec)



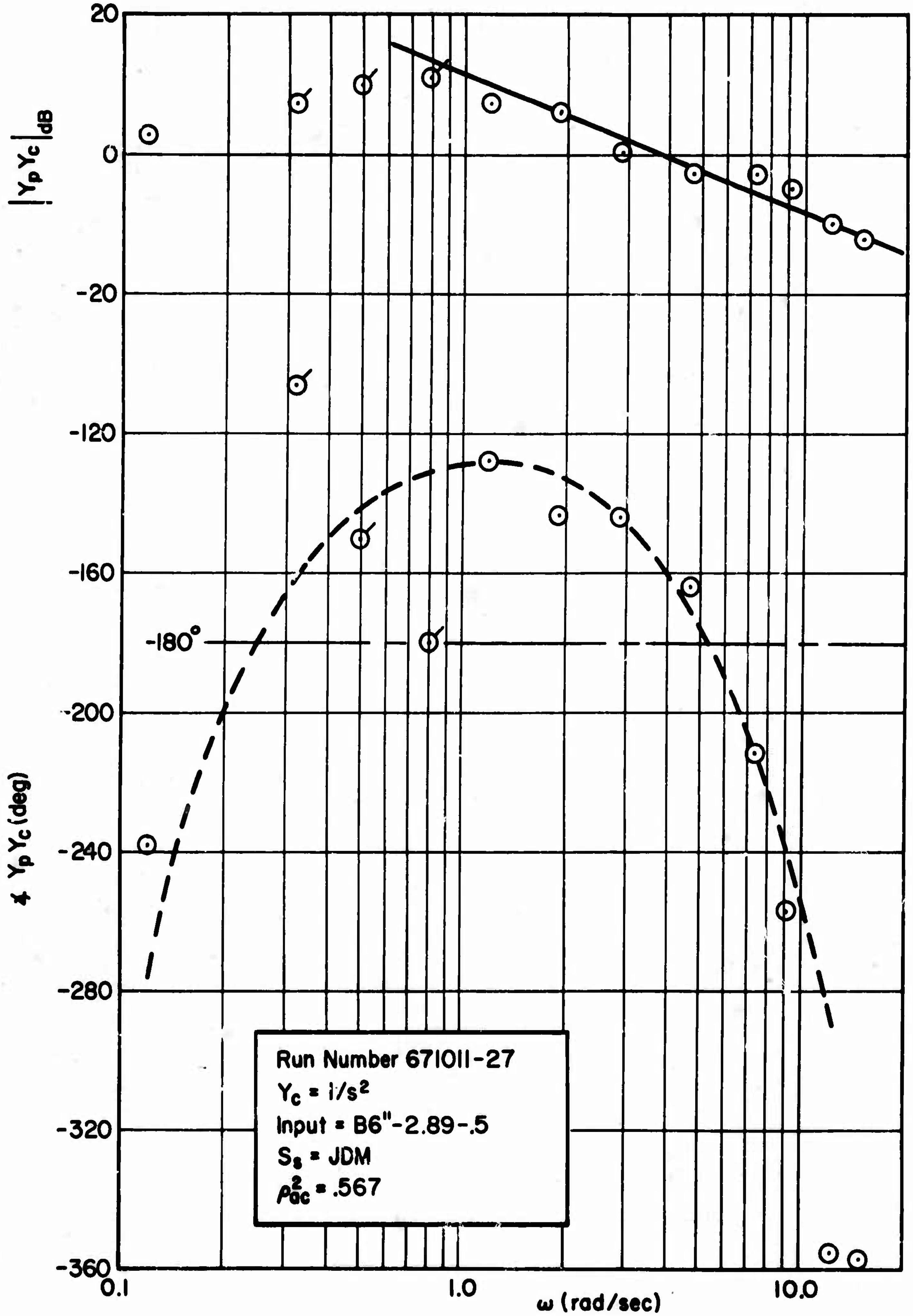
40
20
0
-20
200
160
120
80
40
0
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
180°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671011-21
$Y_c = 1/s^2$
Input = B6"-2.89-1
S_s = JDM
ρ_{ac}^2 = .618



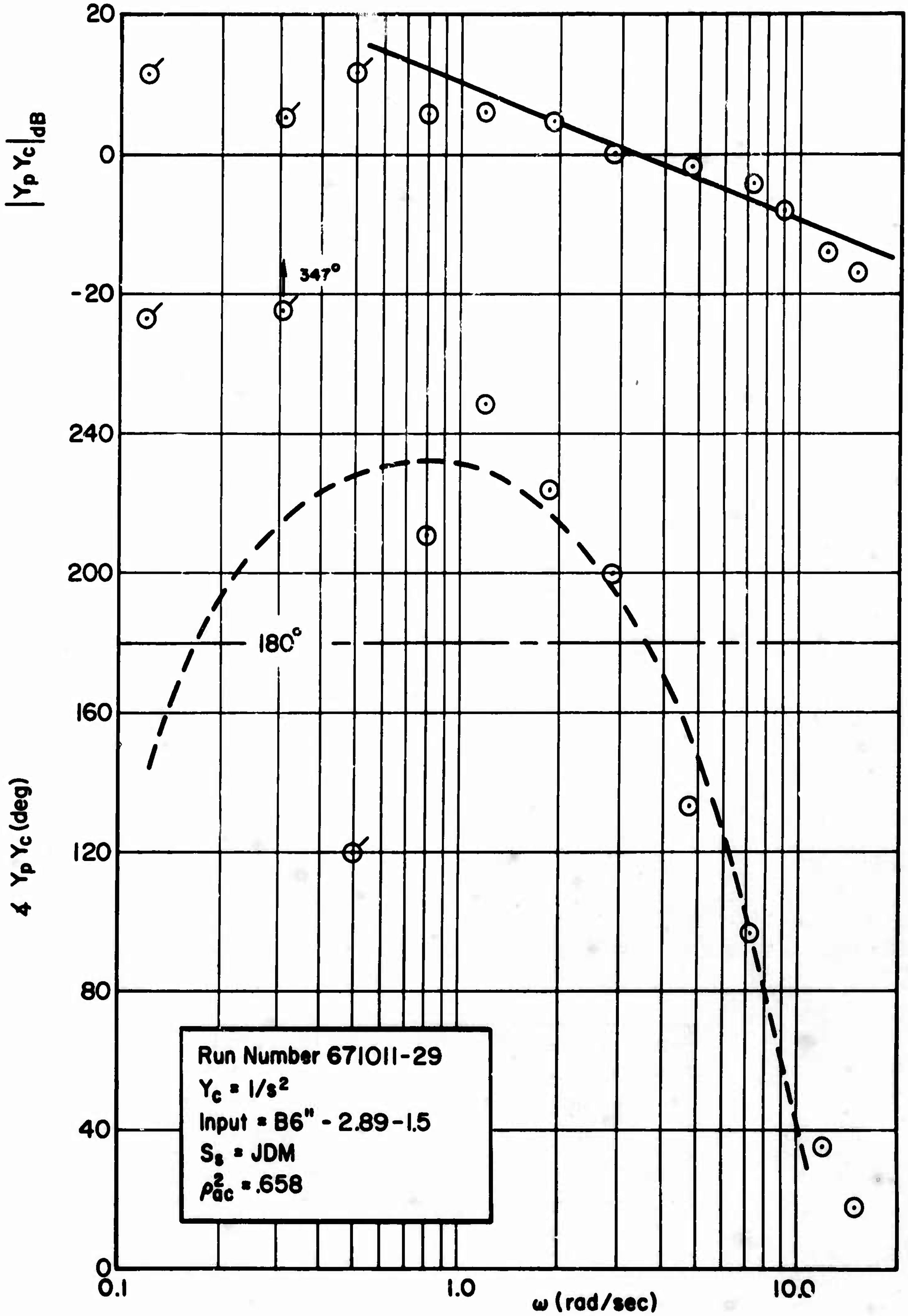
20
0
-20
240
200
180
160
120
80
40
0
0.1
1.0
10.0
ω (rad/sec)
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
Run Number 671011-23
$Y_c = 1/s^2$
Input = B6" - 4.78-1
S_s = JDM
ρ^2_{ac} = .744



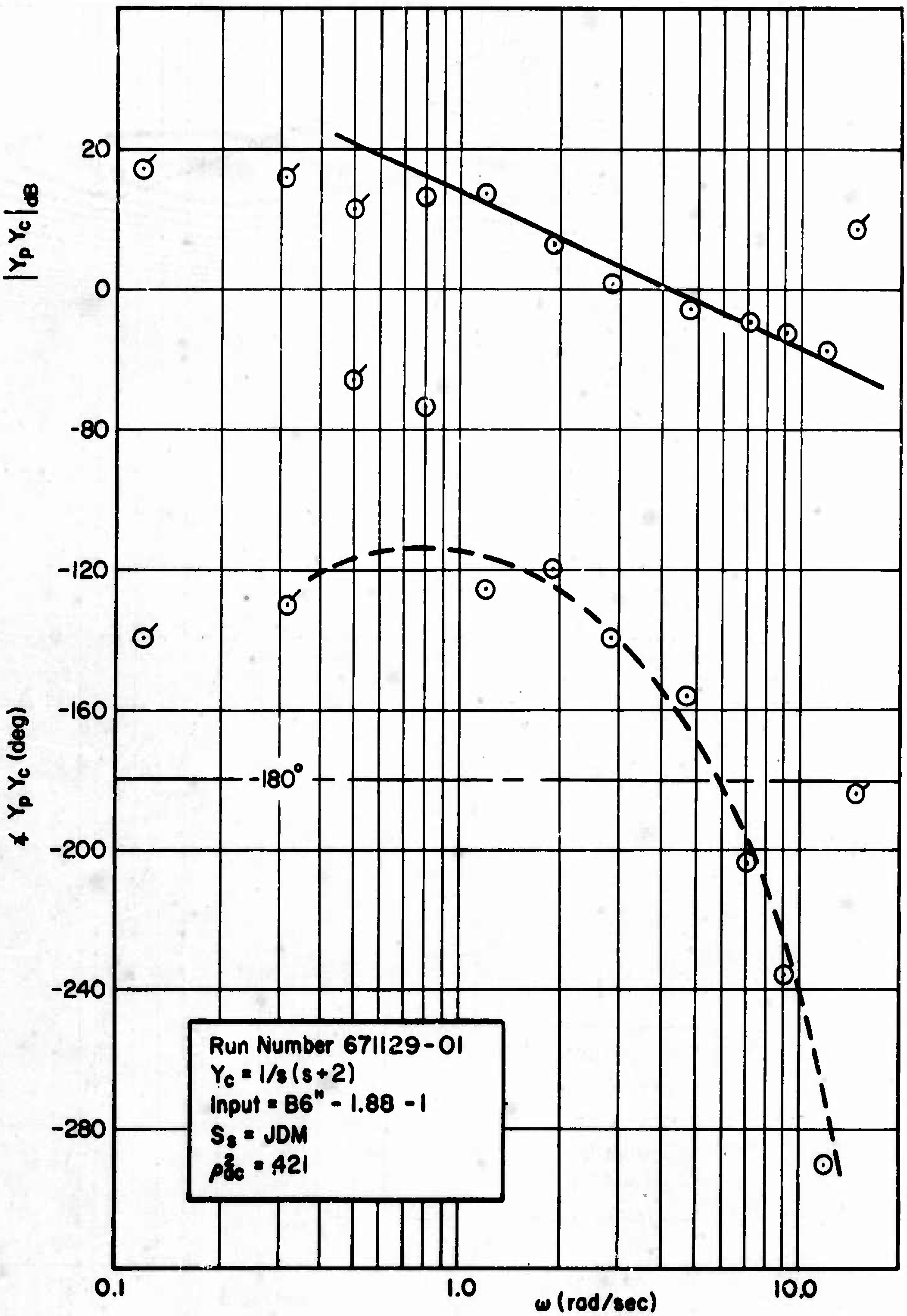
20
0
-20
$|Y_p Y_c|_{dB}$
200
180
160
120
80
40
$\measuredangle Y_p Y_c$ (deg)
Run Number 671011-25
$Y_c = 1/s^2$
Input = B6" - 1.88 -.5
S_s = JDM
ρ_{ac}^2 = .564
0.1
1.0
10.0
ω (rad/sec)



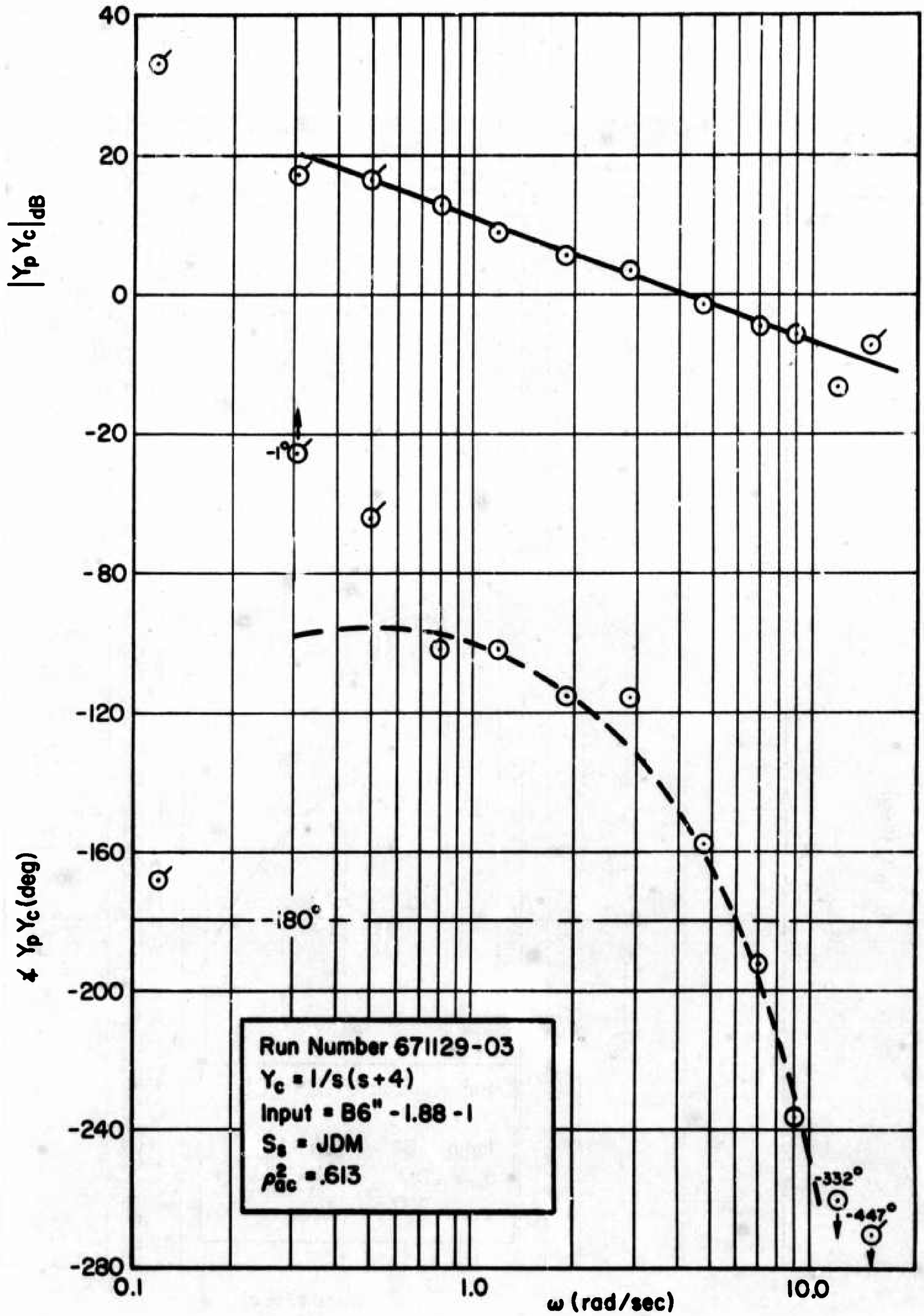
20
0
-20
-120
-160
-180°
-200
-240
-280
-320
-360
0.1
1.0
10.0
ω (rad/sec)
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
Run Number 671011-27
$Y_c = 1/s^2$
Input = B6"-2.89-.5
S_s = JDM
$\rho_{ac}^2 = .567$



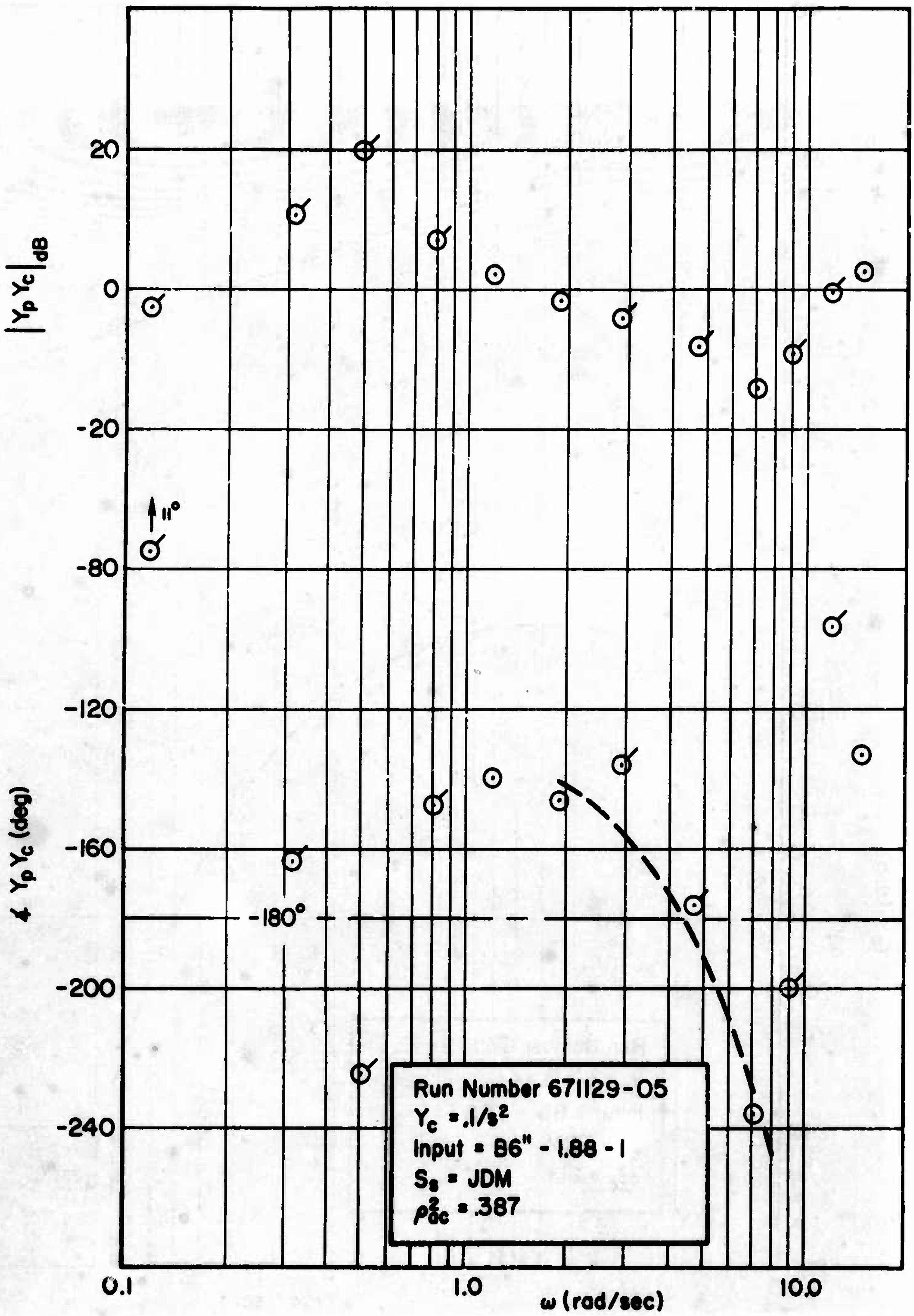
20
0
-20
240
200
160
120
80
40
0
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
347°
180°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671011-29
$Y_c = 1/s^2$
Input = B6" - 2.89-1.5
S_s = JDM
$\rho_{ac}^2 = .658$

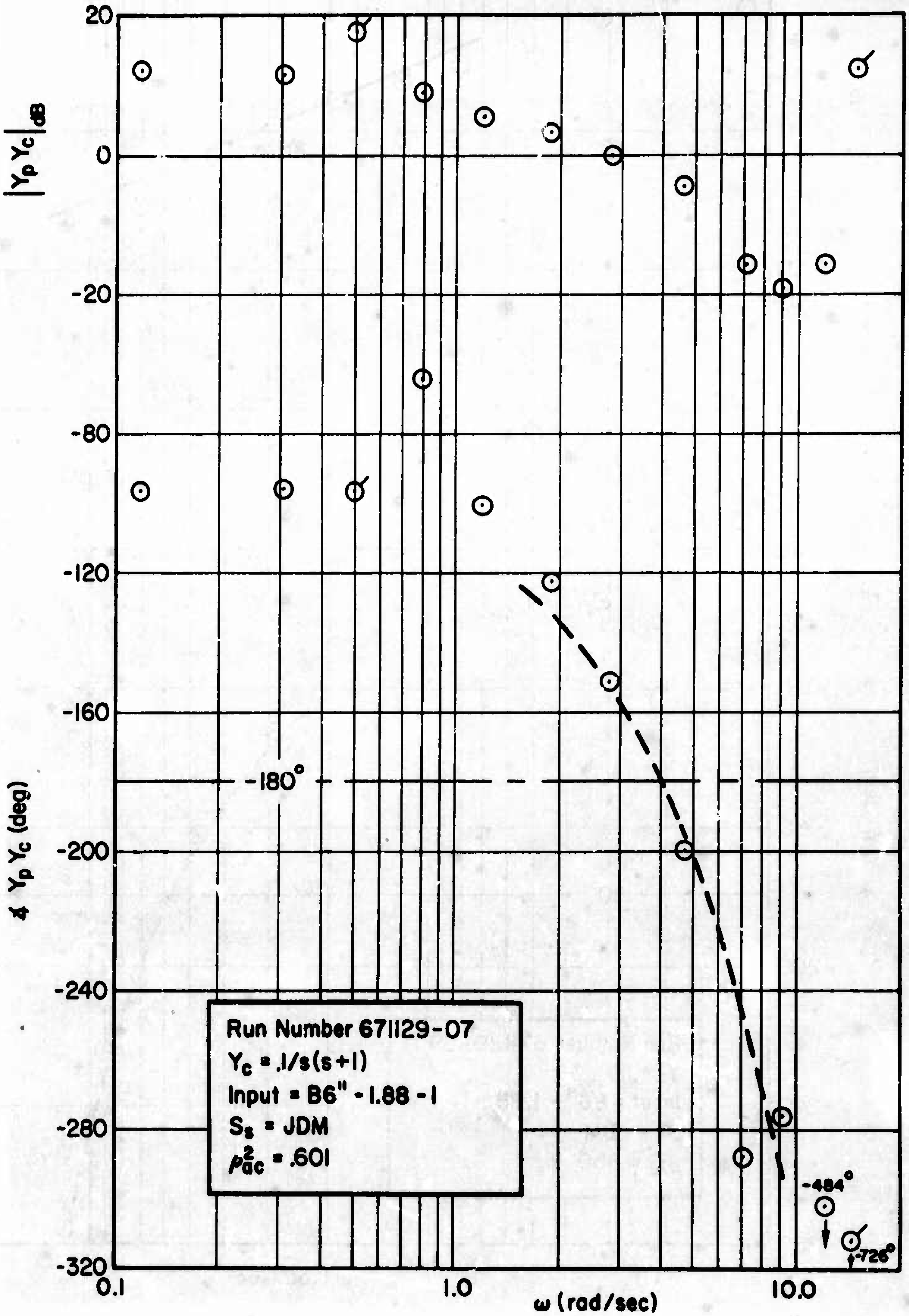


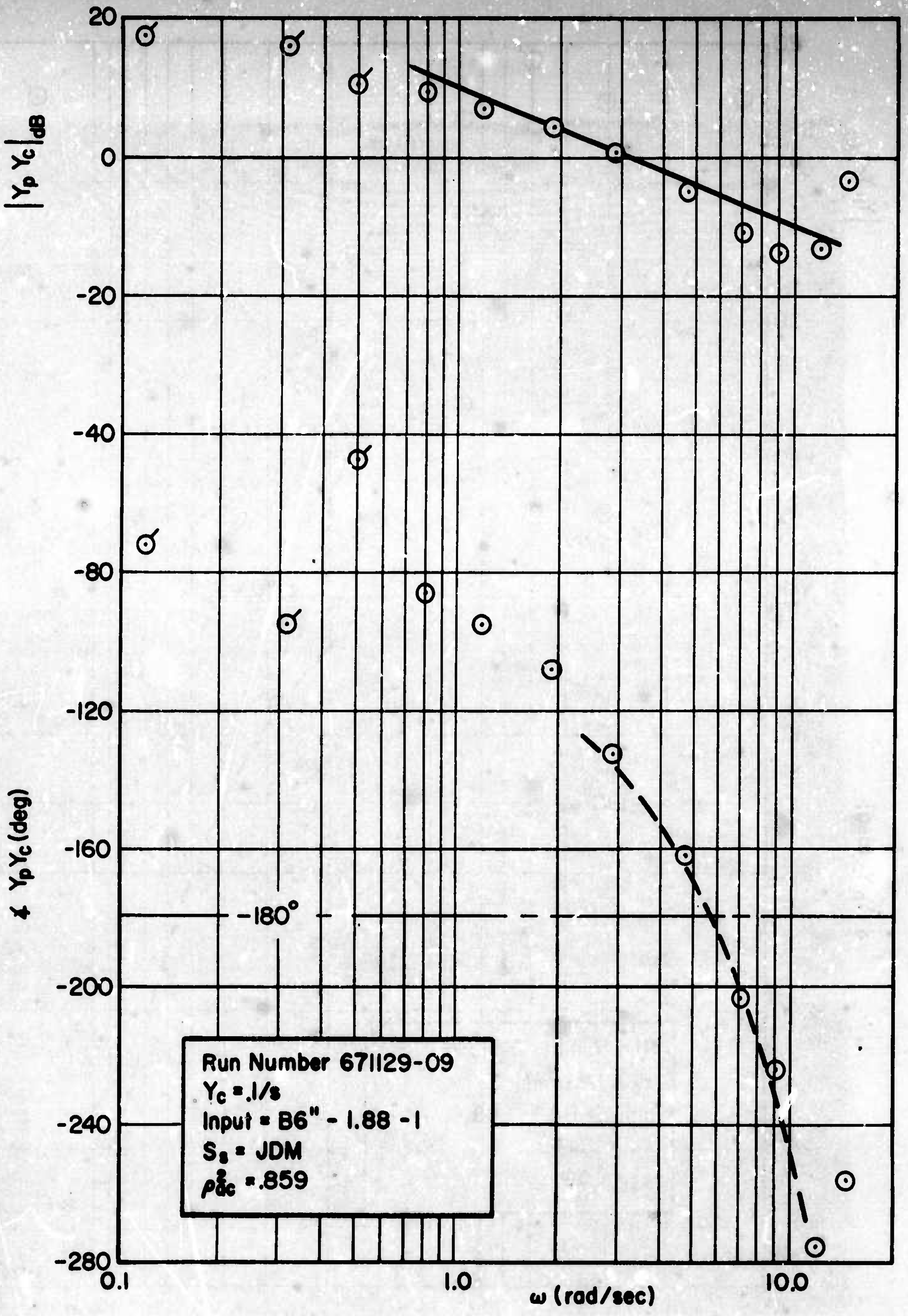
20
0
-80
-120
-160
-200
-240
-280
-180°
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
0.1
1.0
10.0
ω (rad/sec)
Run Number 671129-01
$Y_c = 1/s(s+2)$
Input = B6" - 1.88 - 1
S_8 = JDM
ρ^2_{ac} = .421



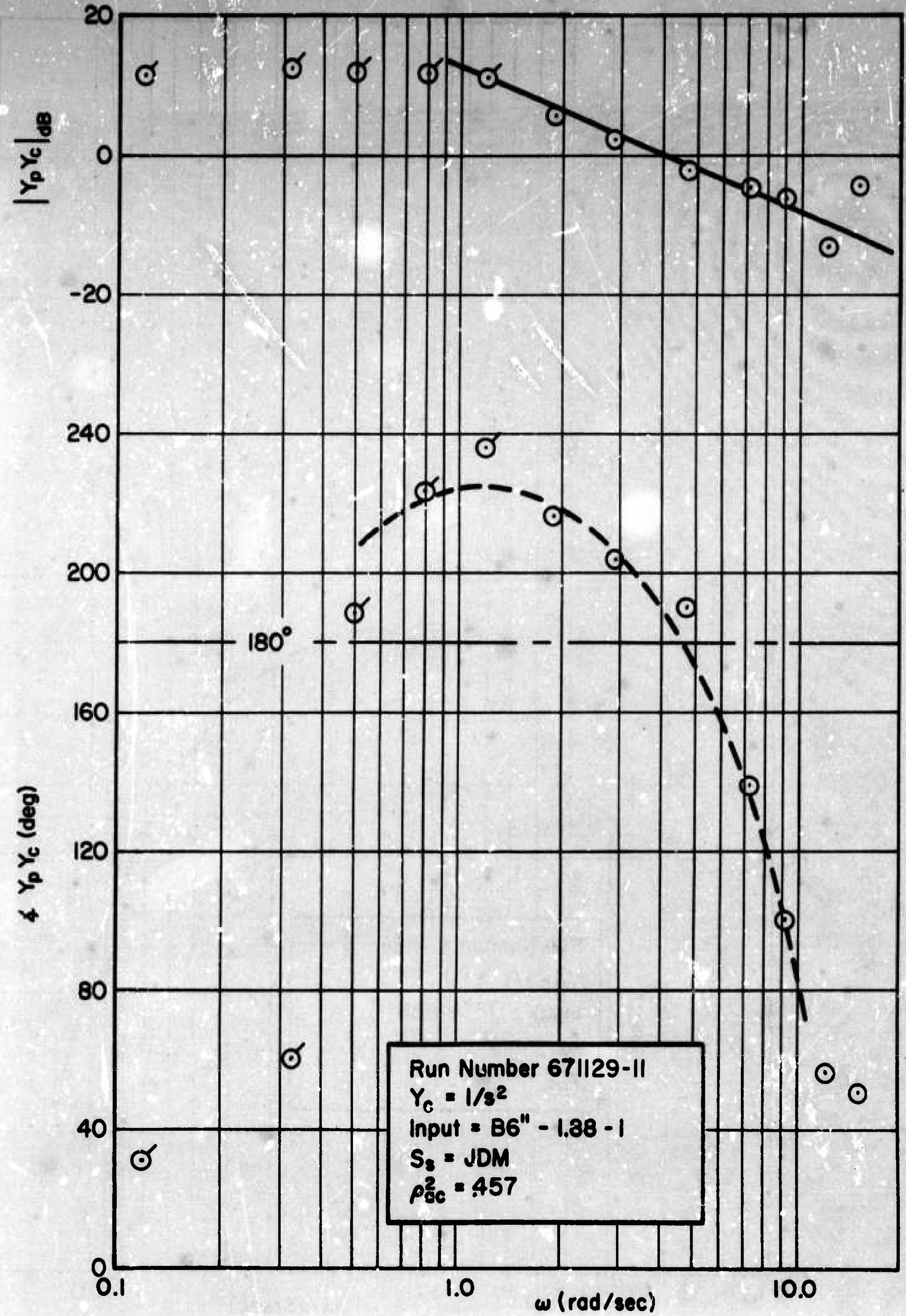
40
20
0
-20
-80
-120
-160
-200
-240
-280
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
-1°
-180°
-332°
-447°
Run Number 671129-03
Y_c = 1/s(s+4)
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} = .613
0.1
1.0
10.0
ω (rad/sec)


20
0
-20
-80
-120
-160
-200
-240
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
11°
-180°
Run Number 671129-05
$Y_c = .1/s^2$
Input = B6" - 1.88 - 1
S_s = JDM
$\rho^2_{ac} = .387$
0.1
1.0
10.0
ω (rad/sec)

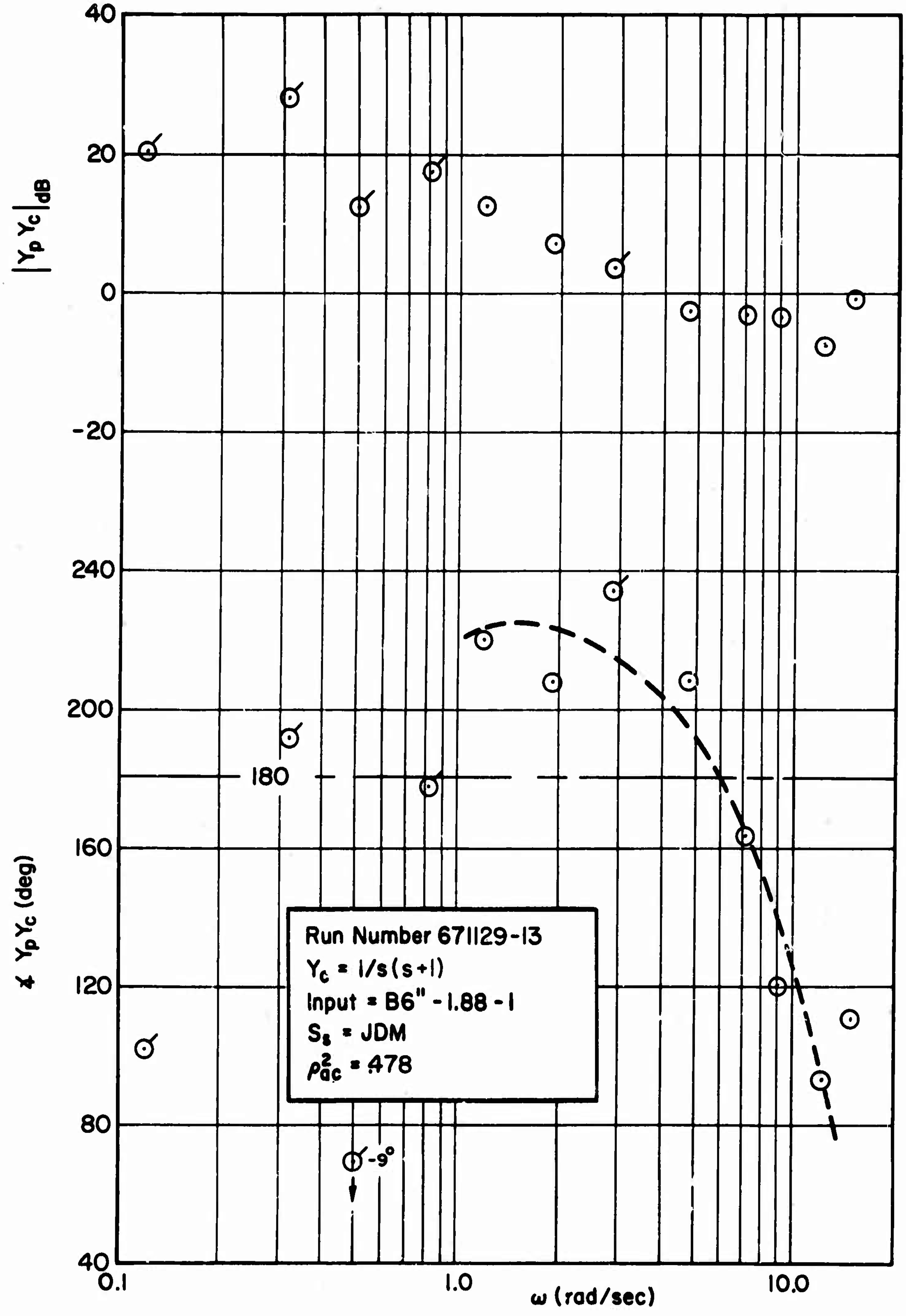

20
0
-20
-80
-120
-160
-200
-240
-280
-320
$|Y_p Y_c|_{dB}$
∡ $Y_p Y_c$ (deg)
-180°
-484°
-726°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671129-07
Y_c = .1/s(s+1)
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} = .601



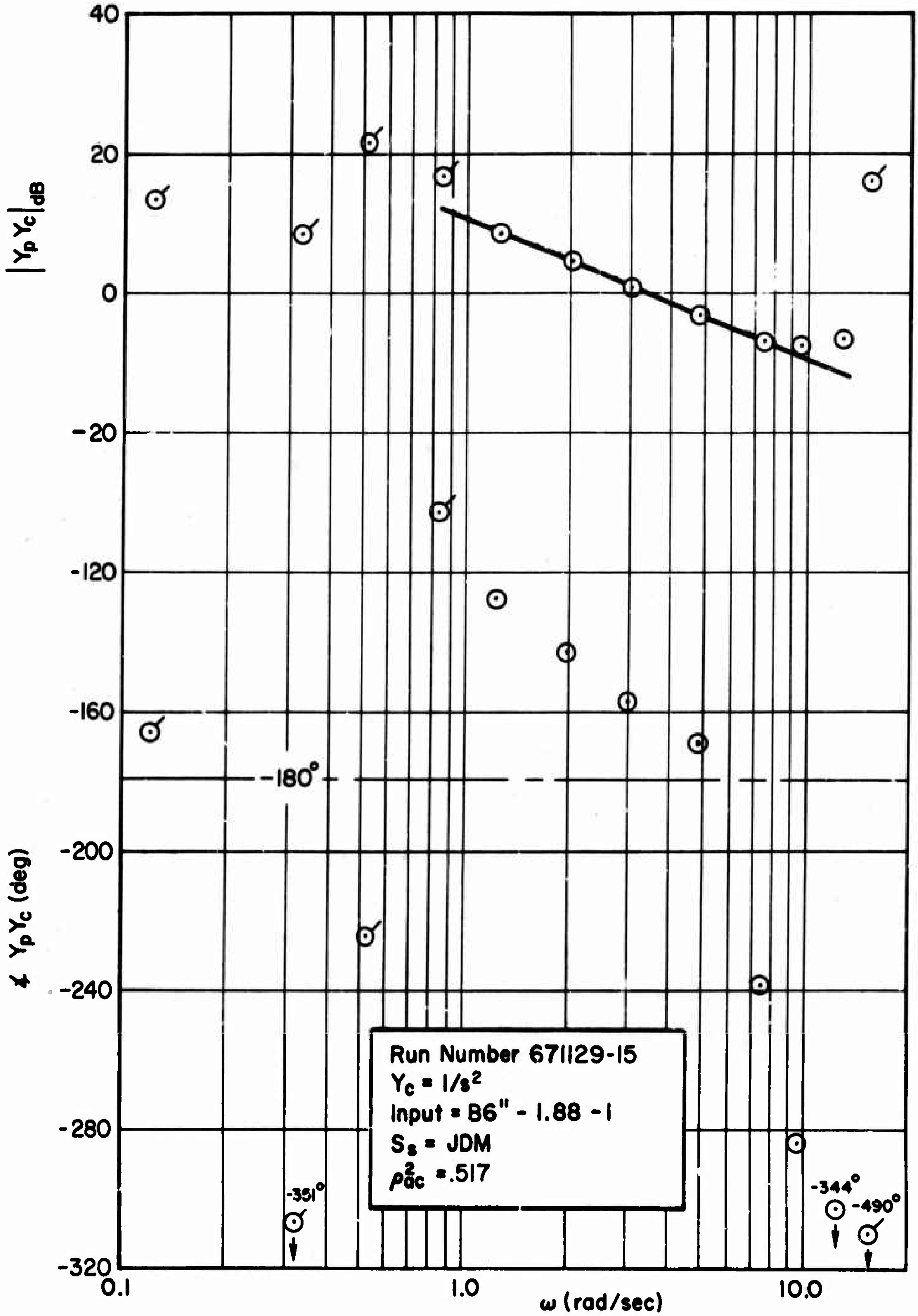
20
0
-20
-40
-80
-120
-160
-200
-240
-280
$|Y_p Y_c|_{dB}$
∡ $Y_p Y_c$ (deg)
-180°
Run Number 671129-09
Y_c = .1/s
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} = .859
0.1
1.0
10.0
ω (rad/sec)



20
0
-20
240
200
160
120
80
40
0
0.1
1.0
10.0
ω (rad/sec)
$|Y_p Y_c|_{dB}$
∢ $Y_p Y_c$ (deg)
180°
Run Number 671129-11
$Y_c = 1/s^2$
Input = B6" - 1.38 - 1
S_s = JDM
ρ^2_{ac} = .457



40
20
0
-20
240
200
180
160
120
80
40
0.1
1.0
10.0
ω (rad/sec)
|Yp Yc|dB
∡ Yp Yc (deg)
-9°
Run Number 671129-13
Yc = 1/s(s+1)
Input = B6" - 1.88 - 1
Ss = JDM
ρ²ac = .478



40
20
0
-20
-120
-160
-180°
-200
-240
-280
-320
$|Y_p Y_c|_{dB}$
$\measuredangle Y_p Y_c$ (deg)
-351°
-344°
-490°
0.1
1.0
10.0
ω (rad/sec)
Run Number 671129-15
$Y_c = 1/s^2$
Input = B6" - 1.88 - 1
S_s = JDM
ρ^2_{ac} = .517

APPENDIX D

COOPER RATINGS FROM STUDY OF McRUER (5)

The data contained in Table D-I are Cooper ratings which were obtained during an experimental series in 1963. These data, together with the Cornell ratings shown in Figs. 40 and 41, provide a valuable comparison between the Cooper and Cornell scales which has heretofore not been available.

TABLE D-I

PREVIOUSLY UNPUBLISHED COOPER RATINGS
OBTAINED DURING McRUER'S (5) STUDY

Pilot: RH

$Y_c = K/s$		$Y_c = 2.5/s(s-\omega_n)$	
Input: B6-0.40(10)-1/2		Input: B6-0.24(10)-1/4	
K	PR	ω_n	PR
10	2	0	6.5
10	2	1	8.5
50	3.5	2	9
5	1.5	0.5	6
2	1.5	1.5	8.5
10	2	1	8
2	2	0.5	6.5
20	4	2	9
100	8	0	6
1	4	1.5	8.5
5	1.5	0.5	7
100	8.5	0	6.5
50	6	2	9
1	4	1	8.5
20	5	1.5	9
50	6.5		
20	3.5		
5	1.5		
2	2		
100	8		
1	4		

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Systems Technology, Inc. 13766 Hawthorne Boulevard Hawthorne, California 90250	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP N/A

3. REPORT TITLE

PILOT RATING TECHNIQUES FOR THE ESTIMATION AND EVALUATION OF HANDLING QUALITIES

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

John D. McDonnell

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 1968	205	58

8a. CONTRACT OR GRANT NO. b. PROJECT NO. 8219 c. Task 821910 d. Work Unit Number 001	9a. ORIGINATOR'S REPORT NUMBER(S) STI-TR-166-1 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFFDL-TR-68-76

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	AFFDL (FDCC) Wright-Patterson AFB, Ohio 45433

13. ABSTRACT

Although rating scales of varied forms have been widely used to estimate and evaluate handling qualities over the past decade, a number of deficiencies in both method and data base have been apparent. This investigation was aimed at overcoming many of these deficiencies by attempting to resolve the difficulties experienced with rating scales themselves, and by extending and adding to already existing relationships between ratings and pilot/vehicle system parameters.

Rating scales have come under increasing criticism for problems such as wording ambiguity, the dual mission character of some scales, the nonuniformity in the distribution of descriptors across the scale, and the misuse of scales which has occurred when ratings have been averaged. Psychometric methods provide an approach to these problems, and in this study were used to scale several phrases descriptive of vehicle handling qualities. Thus, quantitative characteristics were derived for contemporary scales through the use of the Method of Successive Intervals.

An experiment was conducted which added to available data relating Cooper ratings and pilot/vehicle parameters, and which also tested some potential alternate scale candidates. The correlation results indicate that ratings are probably based on performance and the degree of difficulty experienced in maintaining the performance. The difficulty is most easily represented by the pilot equalization required and the vehicle stick characteristics.

DD FORM 1 NOV 65 1473

Unclassified

Security Classification

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Pilot Opinion						
Pilot Rating						
Human Response						
Handling Qualities						

Unclassified
Security Classification