REAL-TIME REMOTE SCIENTIFIC MODEL VALIDATION

34075

P. 8

Richard Frainier and Nicolas Groleau
Recom Technologies, Inc.

NASA Ames Research Center, m/s 269-2
Moffett Field, CA 94035-1000, USA

ABSTRACT

This paper describes flight results from the use of a CLIPS-based validation facility to compare analyzed data from a space life sciences (SLS) experiment to an investigator's pre-flight model. The comparison, performed in real-time, either confirms or refutes the model and its predictions. This result then becomes the basis for continuing or modifying the investigator's experiment protocol. Typically, neither the astronaut crew in Spacelab nor the ground-based investigator team are able to react to their experiment data in real time. This facility, part of a larger science advisor system called Principal-Investigator-in-a-Box, was flown on the Space Shuttle in October, 1993. The software system aided the conduct of a human vestibular physiology experiment and was able to outperform humans in the tasks of data integrity assurance, data analysis, and scientific model validation. Of twelve pre-flight hypotheses associated with investigator's model, seven were confirmed and five were rejected or compromised.

INTRODUCTION

This paper examines results from using a CLIPS-based scientific model validation facility to confirm, or refute, a set of hypotheses associated with a Shuttle-based life-science experiment. The model validation facility was part of a larger software system called "PI-in-a-Box" (Frainier et al., 1993) that was used by astronauts during the October, 1993 Spacelab Life Sciences 2 (SLS-2) mission. The model validation facility (called Interesting Data Filter in the PI-in-a-Box system) compares the output of the scientific data analysis routines with the investigator's pre-flight expectations in real-time.

The model validation facility compares analyzed data from the experiment with the investigator's model to determine its fit with pre-flight hypotheses and predictions. The fit can be either statistical or heuristic. Deviations are reported as "interesting". These deviations are defined as "needing confirmation", even if not part of the original fixed protocol. If confirmed, then at least a portion of the theoretic model requires revision. Further experiments are needed to pinpoint the deviation. This idea is at the heart of the iterative process of "theory suggesting experiment suggesting theory".

THE ROTATING DOME EXPERIMENT

The PI-in-a-Box software system was associated with a flight investigation called the Rotating Dome Experiment (RDE). This was an investigation into the effects of human adaptation to the micro-gravity condition that exists in Earth-orbiting spacecraft. A sensation, called "angular vection", was induced in a set of human subjects by having them view a rotating field of small, brightly-colored dots. After a few seconds, the subject perceives that s/he is rotating instead of the constellation of dots. This perception of self-rotation generally persists throughout the time that the dots are rotating, though occasionally the subject realizes that it is in fact the dots that are

rotating. This sudden cessation of the sensation of vection is termed a "dropout". With the RDE, the field of dots rotates in a set direction (clockwise/counter-clockwise) with a set speed for 20 seconds. There is a 10-second pause, and then the rotation resumes (though with a new direction and/or angular speed). There are six such 20-second trials for each experiment run.

There are three experiment conditions for RDE subjects. In the first, called "free-float", the subject grips a biteboard with his/her teeth in front of the rotating dome (and is otherwise floating freely). In the second, called "tether", the subject is loosely tethered to the front of the rotating dome without the biteboard. In the third, called "bungee", the subject is attached to the "floor" of the laboratory by a set of bungee cords, again teeth gripping a biteboard.

There are eight main parameters measured with respect to angular vection during the RDE. Four of these parameters are "subjective", meaning that the subject consciously reports them by manipulating a joystick/potentiometer: These are the time interval from the start of dome rotation to the onset of the sensation of vection (measured in seconds), the average rate of perceived vection (expressed as a percent of the maximum), the maximum rate of perceived vection (also expressed as a percent of the maximum), and the number of times during a 20-second trial that the sensation suddenly ceases (the dropout count, an integer). The remaining four parameters are "objective", meaning that the subject's involuntary movements are recorded. These are the first and second head movements associated with the torque strain gage mounted on the biteboard and the same head movements associated with subject neck muscle activity detectors (electromyograms). These eight parameters were measured for each 20-second trial of a run.

In the flight system, twelve distinct hypotheses were identified. These were all associated with the joystick-generated subjective parameters. They are:
1. There should be some sensation of angular vection.
2. The average time for the onset of the sensation of vection for the six trials of a run should be greater than 2 seconds* .
3. The average time for the onset of the sensation of vection for the six trials of a run should be less than 10 seconds.
4. Early in a mission, before adaptation to micro-gravity is significantly underway, the average of the six trials' maximum sensation of vection should be less than 90%.
5. Late in a mission, after adaptation to micro-gravity is complete, the average of the six trials' maximum sensation of vection should be more than 80%.
6. Tactile cues decrease the sensation of vection, therefore, the average of the six trials' maximum sensation of vection for a free-float run should be more than that of a bungee run.
7. The average of the six trials' average sensation of vection should be more than 30%.
8. The average of the six trials' average sensation of vection should be less than 80%.
9. Tactile cues decrease the sensation of vection, therefore, the average of the six trials' average sensation of vection for a free-float run should be more than that of a bungee run.
10. There should be at least one dropout during a bungee run.
11. There should not be an average of more than two dropouts per trial during a free-float run.
12. The average of the six trials' dropout count for a free-float run should be less than that of a bungee run.

MODEL VALIDATION

Real-time, quick-look data acquisition and analysis routines extract significant parameters from the experiment that are used by the model validation facility (see Figure 1). With the RDE,

---

*Strictly speaking, the numeric value of this hypothesis (and most other hypotheses) was subject to adjustment on a subject-by-subject basis as a result of pre-flight "baseline data" measurements. This was due to the significant variability between individual human subjects.

predictions were formed on the basis of data from two sources. The first source was previously-collected flight data. (The RDE was flown on three earlier missions: SL-1, D-1, and SLS-1.) The second source was from SLS-2 crew responses recorded on earth before the flight during baseline data collection sessions. These predictions were used to define thresholds that, if violated, indicated significant deviations from the investigator's model. Many space life-sciences investigations (including the RDE) are exploratory in nature, and the investigator team expected significant deviations for perhaps 20% of the experiment runs. When detected, these deviations were made available for display to the astronaut-operator. It is then that the reactive scientist briefly reflects on the situation and try to exploit the information to increase the overall science return of the experiment. This would most likely result in a change to the experiment protocol.
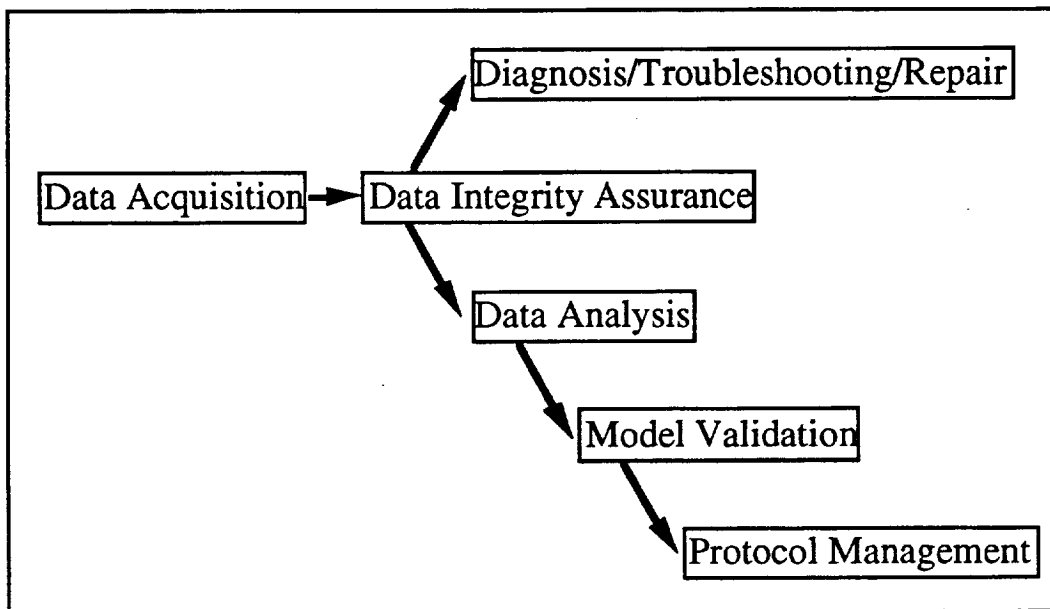

Figure 1: Flow of control.

For the PI-in-a-Box system, the model validation facility was named the Interesting Data Filter (IDF). The IDF was a set of CLIPS rules and facts that compared current experiment results with the investigator's preflight expectations. There were approximately two dozen rules[†] and 40 facts that comprised the pre-flight hypotheses.

FLIGHT RESULTS

Results of the flight use with respect to the 12 hypotheses are listed in Table I:
* the first column identifies the hypothesis number from the list of hypotheses presented earlier.
* the second column is the binomial probability of observing the given outcome assuming 95% of the run results agree with the model.
* the third column is our conclusion with respect to the hypothesis given the overall SLS-2 flight evidence. The hypothesis is rejected when the probability of observing the flight results given the hypothesis is < 0.001; it is compromised when the probability of observing the flight results given the hypothesis is < 0.01; it is suspect[‡] when the probability of observing the flight results given the hypothesis is < 0.05; and it is accepted otherwise.

[†]See Appendix for listing of CLIPS rules.
[‡]This case does not occur for this data set.

- the fourth column summarizes the mission results. This is expressed as a ratio where the denominator represents the number of experiment runs producing data that bears on the hypothesis and the numerator represents the subset of those experiment runs whose data supports the hypothesis. The entry "n/a" denotes that the hypothesis was not applicable to that flight day.
- the last three columns present a more detailed view of the results from each of the three flight days (fd) when the system was in use.

These results indicate that seven of 12 pre-flight hypotheses were accepted. Five hypotheses were either rejected or compromised, indicating a need to modify the existing model with respect to the pattern of human adaptation to weightlessness over time, with respect to the importance of dropouts as an indication of adaptation, and with respect to the influence of tactile cues.

Table I: Flight data confirmation of vection-related hypotheses.

hyp #	probability	result	mission total	FD2	FD8	FD11
1	1.0000	accepted	38/38	12/12	14/14	12/12
2	0.2642	accepted	18/20	6/6	8/8	4/6
3	1.0000	accepted	38/38	12/12	14/14	12/12
4	1.0000	accepted	12/12	12/12	n/a	n/a
5	0.0000	rejected	0/12	n/a	n/a	0/12
6	0.0040	compromised	17/22	3/4	9/10	5/8
7	0.8576	accepted	37/38	11/12	14/14	12/12
8	1.0000	accepted	38/38	12/12	14/14	12/12
9	0.0040	compromised	17/22	4/4	8/10	5/8
10	0.0000	rejected	5/13	1/4	3/5	1/4
11	1.0000	accepted	13/13	4/4	5/5	4/4
12	0.0009	rejected	14/22	4/4	5/10	5/8

CONCLUSION

A scientific model validation facility has been devised for space science advisor systems that appears to be a useful framework for confirming or refuting pre-flight hypotheses. This facility is a key step to achieving truly reactive space-based laboratory science.

ACKNOWLEDGMENTS

REFERENCES

Frainier Richard, Groleau Nicolas, Hazelton Lyman, Colombano Silvano, Compton Michael, Statler Irv, Szolovits Peter, and Young Larry, "PI-in-a-Box: a knowledge-based system for space science experimentation", AI Magazine, Menlo Park, CA, vol. 15, no. 1, Spring, 1994, pp. 39-51.

APPENDIX: MODEL VALIDATION FLIGHT RULE-SET

```
;; ;these rules follow CLIPS v4.3 syntax

;;start up IDF
(defrule idf-startup
    (idf)
    (disk-drive ?disk)
    (interface-directory ?dir)
    =>
    (bind ?predictions-file (str_cat ?disk ?dir "BDC-predictions"))
    (bind ?input-file (str_cat ?disk ?dir "idf-input"))
    (load-facts ?predictions-file)
    (load-facts ?input-file)
    (assert (idf-result not-interesting)) ; the default result
    (open (str-cat ?disk ?dir "idf-stats") idf-stats "w")) ; no append w/o file size limit checking!


;; formula: SD_square = (1/n)(sum_of squares - (square of sum)/n)
(defrule compute_statistics_for_six_trials
    (declare (salience 10))
    (?parameter trial_data ?t1 ?t2 ?t3 ?t4 ?t5 ?t6)
    =>
    (bind ?sum (+ (+ (+ (+ (+ ?t1 ?t2) ?t3) ?t4) ?t5) ?t6))
    (bind ?sum_of_squares (+ (** ?t1 2) (** ?t2 2) (** ?t3 2) (** ?t4 2) (** ?t5 2) (** ?t6 2)))
    (bind ?mean (/ ?sum 6))
    (bind ?SD_square (/ (- ?sum_of_squares (/ (** ?sum 2) 6)) 6))
    (bind ?SD (sqrt ?SD_square))
    (assert (?parameter sum ?sum))
    (assert (?parameter sum of squares ?sum_of_squares))
    (assert (?parameter experiment_result ?mean))
    (assert (?parameter standard deviation ?SD)))


;;;; Parameter-specific rules to detect interestingness


;; ONSET_OF_VECTION

;; Onset of vection is interesting if it's non-existent (that is, less than 0.03 seconds)
(defrule no-vection-detected
    (declare (salience 5))
    (Onset_Of_Vection experiment_result ?x&:(< ?x 0.03))
    =>
    (assert (no-vection-detected)))

(defrule no-vection-detected--interesting
    (declare (salience 5))
    (no-vection-detected)
    (Maximum_Vection_Intensity experiment_result ?x&:(< ?x 10))
    =>
    (assert (Onset_Of_Vection conclusion potentially_interesting "No vection was detected.")))

;; Onset of vection is interesting if it's consistently < threshold (but >= 0.03)
(defrule onset_of_vection_less_than_2
    (Onset_Of_Vection experiment_result ?mean_found)
    (subject ?subj)
    (BDC-datum ?subj quick_onset ?threshold)
    (test (and (> ?mean_found 0.03) (< ?mean_found ?threshold)))
    =>
    (bind ?msg (str_cat "Mean onset of vection is less than " ?threshold " seconds"))
    (assert (Onset_Of_Vection conclusion potentially_interesting ?msg)))
```

```
;; Onset of vection in flight is interesting if it's consistently > threshold
(defrule onset_of_vection_greater_than_10
    (environment flight)
    (subject ?subj)
    (BDC-datum ?subj slow_onset ?threshold)
    (Onset_Of_Vection experiment_result ?mean_found&:(> ?mean_found ?threshold))
    =>
    (bind ?msg (str_cat "Mean onset of vection is greater than " ?threshold " seconds"))
    (assert (Onset_Of_Vection conclusion potentially_interesting ?msg)))


;; MAXIMUM_VECTION_INTENSITY

(defrule early_interesting_maximum_vection
; Early in flight (Day 0 / Day 1), maximum vection is interesting if it's consistently > threshold.
    (environment flight)
    (day 0|1)
    (subject ?subj)
    (BDC-datum ?subj early_hi_max ?threshold)
    (Maximum_Vection_Intensity experiment_result ?mean_found&:(> ?mean_found ?threshold))
    (not (no-vection-detected))
    =>
    (bind ?msg (str_cat "Max vection intensity mean is greater than " ?threshold "%"))
    (assert (Maximum_Vection_Intensity conclusion potentially_interesting ?msg)))


(defrule late_interesting_maximum_vection
; Late in the flight, maximum vection is interesting if it's consistently < threshold
    (environment flight)
    (day ?day&:(> ?day 7))                  ; "Late" is Day 8 or later
    (subject ?subj)
    (BDC-datum ?subj late_lo_max ?threshold)
    (Maximum_Vection_Intensity experiment_result ?mean_found&:(< ?mean_found ?threshold))
    (not (no-vection-detected))
    =>
    (bind ?msg (str_cat "Max vection intensity mean is less than " ?threshold "%"))
    (assert (Maximum_Vection_Intensity conclusion potentially_interesting ?msg)))

;; Max vection is interesting if tactile > free-float
(defrule free-below-bungee--interesting--maximum-vection
    (body_position free-flt)
    (Maximum_Vection_Intensity experiment_result ?ff)
    (Maximum_Vection_Intensity running_mean ?val&:(> ?val ?ff)) ; bungee mean
    =>
    (assert (Maximum_Vection_Intensity conclusion potentially_interesting
            "Subj's max. vection < bungee cond. max. vection")))

(defrule bungee-above-free--interesting--maximum-vection
    (body_position bungee)
    (Maximum_Vection_Intensity experiment_result ?b)
    (Maximum_Vection_Intensity running_mean ?val&:(> ?b ?val)) ; free-float mean
    =>
    (assert (Maximum_Vection_Intensity conclusion potentially_interesting
            "Subj's max. vection > free-float cond. max. vection")))


;; AVERAGE_VECTION_INTENSITY

(defrule low_average_vection_intensity
; Average vection intensity is interesting if it's consistently < threshold
    (environment flight)
    (subject ?subj)
    (BDC-datum ?subj lo_average ?threshold)
    (Average_Vection_Intensity experiment_result ?mean_found&:(< ?mean_found ?threshold))
```

128

```
    (not (no-vection-detected))
    =>
    (bind ?msg (str_cat "Avg. vection intensity mean is less than " ?threshold "%"))
    (assert (Average_Vection_Intensity conclusion potentially_interesting ?msg)))


(defrule high_average_vection_intensity
; Average vection intensity is interesting if it's consistently > threshold
    (environment flight)
    (subject ?subj)
    (BDC-datum ?subj hi_average ?threshold)
    (Average_Vection_Intensity experiment_result ?mean_found&:(> ?mean_found ?threshold))
    (not (no-vection-detected))
=>
    (bind ?msg (str_cat "Avg. vection intensity mean is greater than " ?threshold "%"))
    (assert (Average_Vection_Intensity conclusion potentially_interesting ?msg)))


;; Average vection is interesting if tactile > free-float
(defrule free-below-bungee--interesting--average-vection
    (body_position free-flt)
    (Average_Vection_Intensity experiment_result ?ff)
    (Average_Vection_Intensity running_mean ?val&:(> ?val ?ff)) ; bungee mean
    =>
    (assert (Average_Vection_Intensity conclusion potentially_interesting
            "Subj's ave. vection < bungee cond. ave. vection")))


(defrule bungee-above-free--interesting--average-vection
    (body_position bungee)
    (Average_Vection_Intensity experiment_result ?b)
    (Average_Vection_Intensity running_mean ?val&:(> ?b ?val)) ; free-float mean
    =>
    (assert (Average_Vection_Intensity conclusion potentially_interesting
            "Subj's ave. vection > free-float cond. ave. vection")))



;; DROPOUTS

;; Number of dropouts is interesting if it's consistently 0 under tactile conditions
(defrule interesting_dropouts_tactile
    (environment flight)
    (body_position bungee)
    (Dropouts experiment_result 0)
    (not (no-vection-detected))
    =>
    (assert (Dropouts conclusion potentially_interesting
            "There were no dropouts with bungees attached")))


;; Number of dropouts is interesting if it's consistently >2 under free-float conditions
(defrule interesting_dropouts_free
    (environment flight)
    (body_position free-flt)
    (Dropouts experiment_result ?mean_found&:(> ?mean_found 2))
    =>
    (assert (Dropouts conclusion potentially_interesting
            "Mean number of free-float dropouts is greater than 2")))


;; Number of dropouts is interesting if tactile consistently < free-float
(defrule free-above-bungee--interesting--dropouts
    (body_position free-flt)
    (Dropouts experiment_result ?ff)
    (Dropouts running_mean ?val&:(> ?ff ?val)) ; bungee mean
    =>
    (assert (Dropouts conclusion potentially_interesting
            "Subj's dropout count > bungee cond. dropout count")))
```

129

```
(defrule bungee-below-free--interesting--dropouts
   (body_position bungee)
   (Dropouts experiment_result ?b)
   (Dropouts running_mean ?val&:(< ?b ?val)) ; free-float mean
   =>
   (assert (Dropouts conclusion potentially_interesting
         "Subj's dropout count < free-float cond. dropout count")))


;;; OUTPUT STATS

(defrule output_idf_stats
   (?parameter experiment_result ?mean)
   (?parameter standard deviation ?SD)
   (body_position ?cond)
   (subject ?subj)
   =>
   (fprintout idf-stats "Subject: " ?subj " Cond: " ?cond " " ?parameter " mean: " ?mean " SD: " ?SD crlf))

;; Output Interestingness info to "Session History" file for Session Manager
(defrule record-interestingness--start
   (declare (salience -100))
   (?parameter conclusion ?interesting ?source)
   ?f <- (idf-result not-interesting)
   (disk-drive ?disk-drive)
   (interface-directory ?interface-dir)
=>
   (retract ?f)
   (assert (idf-result interesting))
   (open (str-cat ?disk-drive ?interface-dir "history-session") history-session "a")
   (assert (record-interestingness)))

; potentially_interesting     => medium
; certainly_interesting       => high
(defrule record-interestingness
   (record-interestingness)
   ?int <- (?parameter conclusion ?interesting ?source)
   (subject ?subj)
   (body_position ?cond)
   (current-step ?step)
   (this-session ?session)
=>
   (retract ?int)
   (if      (eq ?interesting potentially_interesting)
   then    (bind ?level medium)
   else    (bind ?level high))
   (fprintout history-session "(int-hist class interesting session " ?session " step " ?step " subj "
         ?subj " cond " ?cond " source " ?source " level " ?level ")" crlf))

(defrule record-interestingness--end
   (record-interestingness)
   (not (?parameter conclusion ?interesting ?source))
=>
   (close)
   (assert (ctrl--stop idf)) ; inhibit rules-control "abnormal" message
   (printout "hyperclips" "interesting")) ; return to HyperCard

(defrule no-interesting-results
   (declare (salience -200))
   (idf-result not-interesting)
=>
   (close)
   (assert (ctrl--stop idf))
   (printout "hyperclips" "as-expected")) ; return to HyperCard
```

130