PHASE I FINAL REPORT
DOE SBIR DE-FG03-98ER82708
DOE Document: DOE-OAK/SBIR

RECEIVED
JAN 24 2000
OSTI

Project Title:
Automatic Component Calibration and Error Diagnostics for Model-based Accelerator Control
Small Business:
Vista Control Systems, Inc.
Principal Investigator:
Dr. Carl Stern
Primary SLAC Research Scientist:
Dr. Martin Lee

Abstract

Phase I work studied the feasibility of developing software for automatic component calibration and error correction in beamline optics models. A prototype application was developed that corrects quadrupole field strength errors in beamline models. This application uses a decompositional approach, first analyzing the beamline to identify good and bad regions within the model, then correcting each bad region individually. This application deploys a new method for localized error correction that is described in this report. Phase I work also studied the heuristics necessary for successful use of this component calibration methodology. This study established the need for intelligent heuristics in the area of data filtering, data interpretation, and search. A full system for automated component calibration will need to provide customization options so that users can specify the particular heuristics needed to handle a given data set and problem configuration. A method for assuring maximal flexibility and customizability is to provide the necessary functionality in the form of a general toolkit and function library for building model-based applications. Initial analysis was begun on the design of such a toolkit.

1. Overview

The immediate goal of Phase I work was to explore the possibility of automating component calibration and error correction in beamline optics models. A longer term goal was to assess the feasibility of designing a general toolkit for construction of intelligent model-based diagnostic and parameter configuration applications.

A prototype was developed for analysis and correction of quadrupole field strength errors. The system finds quadrupole errors by analyzing orbit response matrix data. An analytic procedure is used which first decomposes the beamline into "good" and "bad" optics regions and then searches for the most likely cause of the orbit errors within each of the "bad" regions. This work builds on previous research at SLAC in model-based analysis [1, 2, 3] but goes beyond previous work in developing a new method for local component calibration and a new approach to the application of knowledge-based heuristics.

DISCLAIMER

The behavior of this prototype was studied on simulated and real data from the HER and SPEAR rings at SLAC. Experimentation with this prototype provided a venue for testing and evaluation of a range of techniques. It also provided an opportunity to experiment with heuristics for data pruning, region identification, and identification of miscalibrated quadrupoles within "bad" regions.

This study allowed us to verify both the feasibility and utility of automating component calibration and error correction in beamline optics models. Our experience demonstrated the remarkable time savings that are achievable using an automated tool. In addition, important heuristics for data analysis and interpretation were identified, verifying the suitability and appropriateness of a heuristics-based approach.

This experience has confirmed the usefulness of an open architecture that allows users to refine or customize general purpose analytic procedures through the specification of additional heuristics and task knowledge. By providing evidence for the feasibility of encoding methods for model-based analysis in a general and customizable form, this work has provided support and encouragement for our longer term goal, the development of a general software toolkit useful for construction of model-based applications.

2. *Importance of the Problem*

The operation of an accelerator for high-energy physics or synchrotron radiation research can be represented by a model. A model that is both accurate and properly calibrated can be used as a tool to commission a new accelerator or storage ring or to control it during its operation. Errors that arise from inaccurately calibrated performance parameters must be identified, located, and then corrected, to ensure the accelerator's performance within certain specifications.

Once an effective model is established, it is possible to apply the model to derive parameter settings for achieving performance objectives. Acceptable machine performance is achieved through error minimization, where error is defined as a discrepancy between model-based prediction and observed performance. Model-based methods of this type can be used in configuration change, orbit correction, dispersion minimization, coupling correction, and a wide variety of other applications.

Until now, error correction in models has been done manually and is, therefore, extremely time consuming because the search space for errors is large. A system automating effective algorithms and heuristics for error finding and correction, with advanced data and information management technologies and a friendly user interface, will realize order-of-magnitude improvements compared to current labor intensive methods. The benefits during commissioning and machine operation will include both savings in time and resources and also the possibility of significant gains in machine performance.

In addition to these time, resource, and performance gains, there is another more subtle reason for integrating automatic model-based tools into accelerator control systems. Automatic model-based software will provide a framework for preserving expertise and information. The various specialized techniques for error correction and model-based control will be preserved and maintained in a centralized knowledge-base rather than dwelling in the expertise of potentially transient group members. The same system will also provide a framework for preserving a history of machine states and parameter settings. Such a history will be useful both in analyzing current machine behavior and in diagnosing problems. This framework will thus support enhanced stability and continuity of machine operations.

3. Technical Approach

In Phase I we developed prototype software for calibration and error correction of quadrupole field strength errors in the beamline optics models. We used a two step decompositional approach to simplify and constrain the search for errors. These methods were tested against simulated and real data from two machines at SLAC.

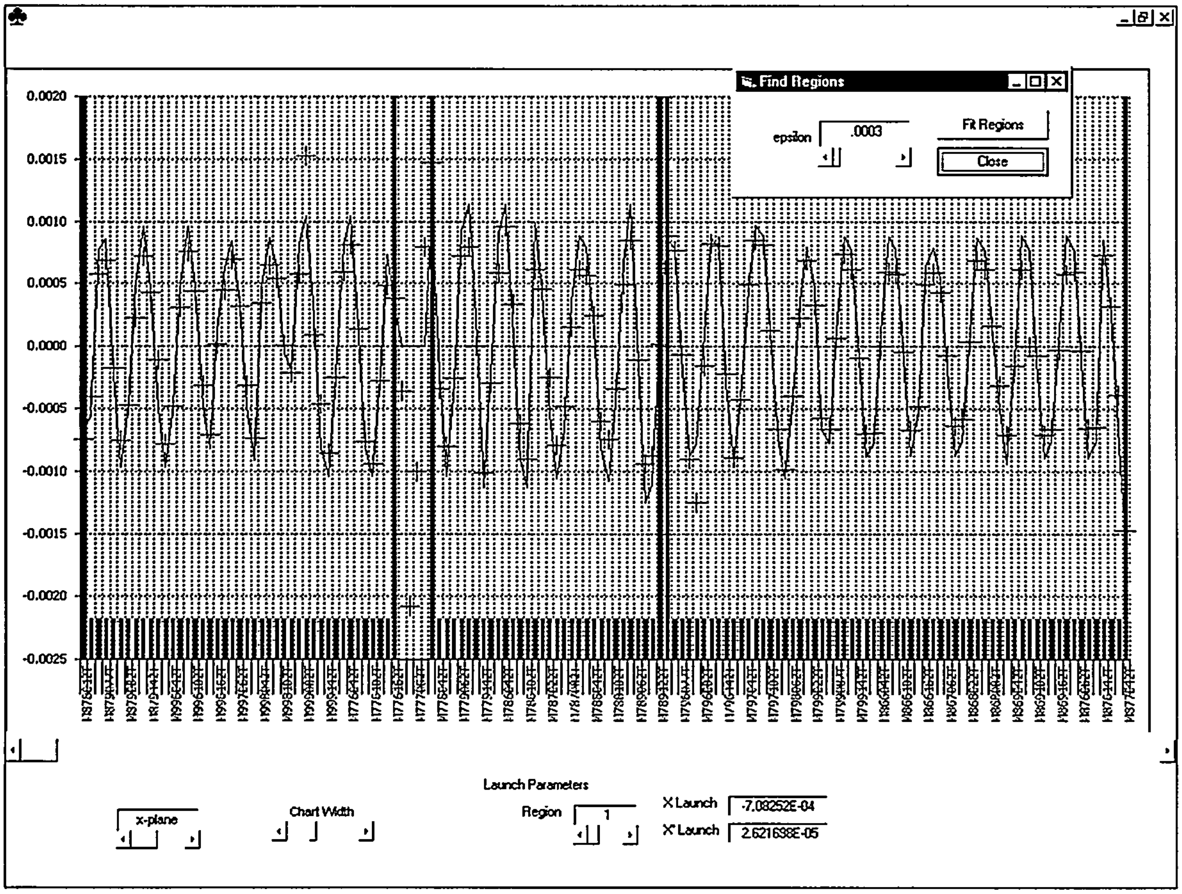
In general two approaches to model calibration are possible:

- *Global approaches* analyze and correct the entire machine model at once using some method for global error minimization.
- *Decompositional approaches* use model-based analysis to identify *good* (well calibrated) and *bad* (poorly calibrated) regions in the machine, then analyze and correct *bad* regions one-by-one.

Global approaches work well for small machines but encounter problems when applied to large machines. Difficulties arise when error minimization techniques are applied to large weakly constrained parameter spaces. In this case solutions become very sensitive to noise in the measured data, and correct solutions becomes more difficult to find. On the other hand, decompositional approaches handle scale-up far more gracefully because they are affected by the size of *bad regions* rather than the size of the *machine*.

Phase I work applied a *decompositional* approach to the analysis and correction of quadrupole field strength errors in optics models. In Phase I, a prototype system was developed that automates the identification of *good* and *bad* regions in an optics model through the analysis of orbit response data. Once the system identifies a *bad* region, an automated search procedure is applied to find and correct miscalibrated quadrupole field strengths in the *bad* region. This system was tested on both simulated and real data from the SPEAR ring and the high-energy ring (HER) of PEP-II.

Figure 1 shows a decomposition of the high-energy ring (HER) of PEP-II into good and bad regions using our Phase I prototype. Good regions have bars at the bottom. Our program discovered three good regions and two bad regions shown here. The first is about a third of the way, the second about three fifths of the way around the ring. The



epsilon .0003

Region 1

X Launch -7.08252E-04

X' Launch 2.621698E-05

Figure 1. Finding good and bad regions in the high energy ring (HER) at SLAC.

presence of the first bad region has been independently verified. The second bad region is still under study.

Analysis of multi-track orbit response data

The information necessary to evaluate the accuracy of the model with respect to quadrupole field strengths was provided by orbit response data. Corrector magnets or dipoles are used to impart a slight deflection to the beam and the offset induced by this kick is measured at every BPM around the beamline. Essentially this method involves using the beam to probe the magnetic field strengths of the elements through which it passes. Analysis compares offsets predicted by the model with those actually measured, to determine regions of the beamline where the model is accurate and regions where it is inaccurate.

When a single corrector is used, this method generates a single column of data or a single track for analysis. When multiple correctors are used, each corrector is tweaked one by one, after restoring previous correctors to their initial setting. This generates multiple tracks, one for each corrector. The information provided by each track must be consistent with all the others, since the extent of the agreement or disagreement between the model and observed system behavior is determined in each beamline section primarily by whether or not the elements in that section are represented accurately in the model. Thus using multi-track data provides a high degree of redundancy, supporting a more robust analysis than that supported by single track analysis.

We had anticipated, and have confirmed in our Phase I work, that this redundancy is necessary in order to reliably identify and correct individual quadrupole field strength errors in bad regions. While single track analysis had been previously automated in an earlier system called GOLD [4], our Phase I prototype is the first system to successfully automate multi-track analysis.

Hierarchical decomposition in model-based error analysis: a closer look

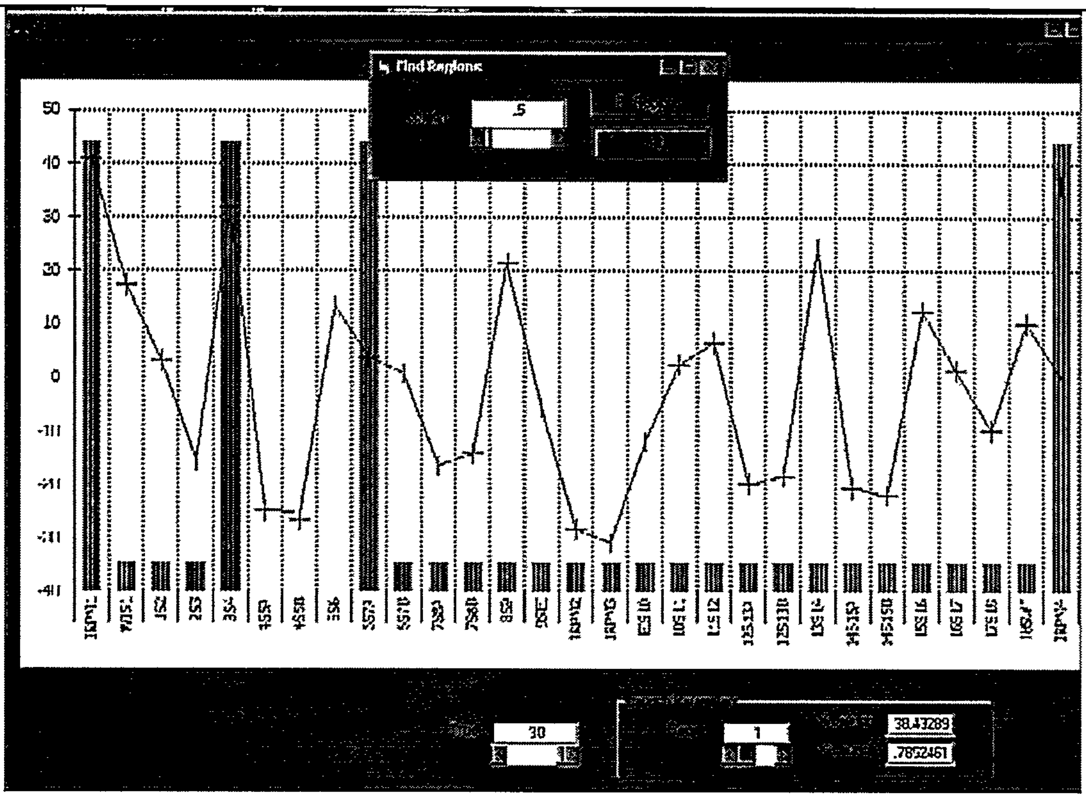
Phase I work involved a particular application of more general method. This general method for model-based error analysis uses a two step decompositional approach. The first step is to localize the error to *bad* regions in order to reduce the size of the search space. The second step is to analyze the sources of error within a *bad* region by selectively changing or perturbing model parameters in the *bad* region to find those changes that eliminate the discrepancies between prediction and observation.

In the first step, a matching procedure is used to find regions of the machine where model-based prediction of machine behavior is in acceptable agreement with observed behavior. A region of the machine where there is a good matching between model-based prediction and observation is called a *good* region.

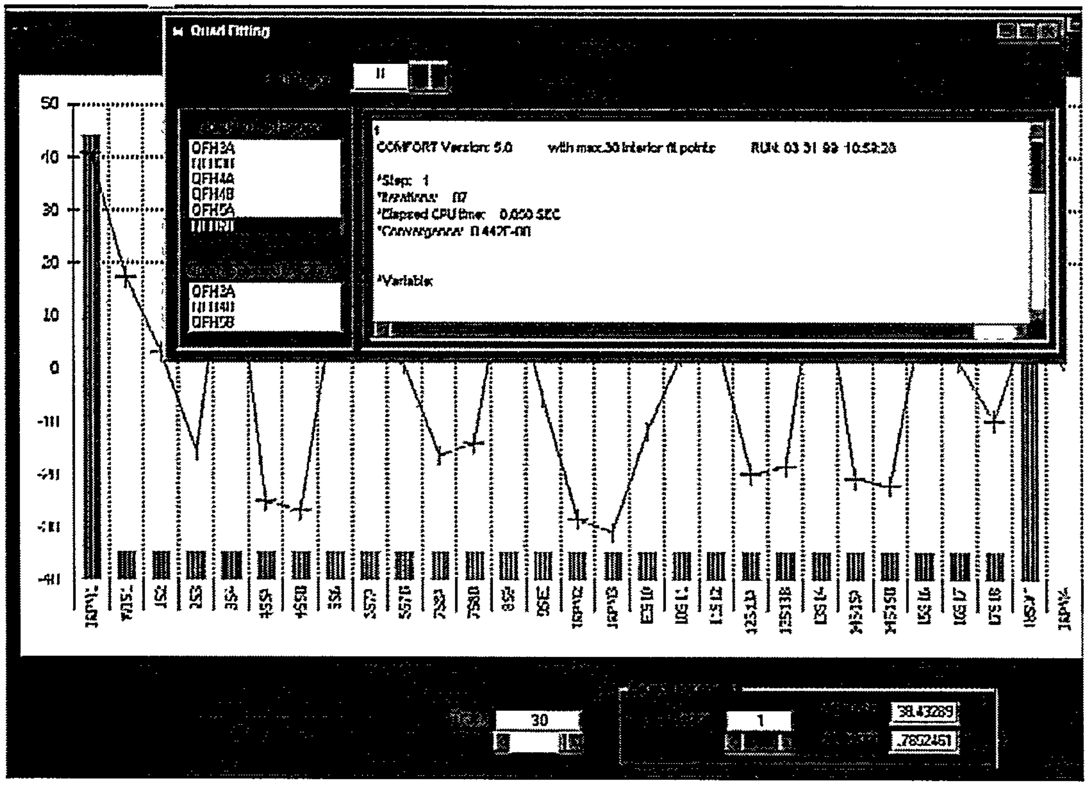
In general, error free regions can be found only when error propagation is local, i.e., when errors in one region do not affect the agreement of model-based prediction and observation in other regions. *Local error propagation* is a strong requirement. Often special measures must be taken in the analysis to enforce local error propagation.

In the Phase I analysis of orbit response data, for example, errors are localized through a special matching procedure that captures global error at the boundaries of local regions. This is achieved by employing a launch independent matching procedure, i.e., a procedure that treats the launch conditions (position and angle of the beam at the entrance to a local region) as free variables. Absorbing the global effects of errors outside the region in the region's launch conditions preserves the possibility of finding a good fit inside the region.

The second step in error correction is to search within a bad region for the source(s) of error, i.e., the source(s) of the discrepancy between prediction and observation. Search is required over a range of hypotheses where each hypothesis explains the discrepancy by attributing error to a particular set of elements. Hypotheses are tested by determining whether some change in the representation of the hypothesized elements, i.e., some change in modeled parameter values, eliminates the discrepancy between prediction and observation.



a. Analysis of SPEAR orbit data before error correction.



```
*
COMFORT Version: 5.0        with max.30 interior fit points        RUN 03 31 99 10:53:28

*Step:  1
*Iterations:  07
*Elapsed CPU time:  0.050 SEC
*Convergence:  0.442E-00


*Variable:
```

b. Analysis of SPEAR orbit data after error correction.

to find which subset of the six is incorrectly modeled. By testing over a range of hypotheses (quadrupole subsets), the prototype found the three incorrectly modeled quadrupoles. **Figure 2b** shows the analysis of the same orbit data using a corrected model. Note that the mismatched region has been disappeared, and only one good region, spanning the entire ring, is identified. The window on top displays the quadrupoles that were used for fitting and the very tight convergence that resulted.

A new procedure for local component calibration

Once a bad region is identified, the problem then becomes to recalibrate the elements in the bad region so that the model's predictions and the observed behavior of the system are again in agreement. In our prototype, the elements to be recalibrated are the quadrupoles in the bad region plus the quadrupoles between a previous BPM and the BPM at the beginning of the bad region. In searching for miscalibrated quadrupoles, it is necessary to "back up" one or two BPMs in order to account for the distance required for propagation of potential quad calibration errors. We call this larger region the "extended bad region".

Once the set of candidate quads is identified, a method is needed to identify the subset of quadrupoles that are miscalibrated and to correct the errors. The Phase I prototype implemented for the first time a unique fitting-based method for identifying and correcting quadrupole errors. This procedure first uses the result of fitting over the previous good region to find the launch conditions at the *beginning* of the extended bad region. Then, for each BPM falling within the extended bad region, the **R** matrix that induces the *observed* beam position at that BPM is calculated based on that initial launch condition. The elements of these **R** matrices serve in the next step as a set of target parameters for fitting.

Our error correction procedure selects subsets of the candidate quadrupoles as explanatory hypotheses, i.e., as elements whose miscalibration explains the divergence of the model's prediction and observed data. A fitting procedure is used to determine whether some new set of field strengths for the quads comprising the hypothesis will produce the required **R** matrix values. Hypotheses group quadrupoles one at a time, two at a time, three at a time, etc., to determine whether the **R** matrix constraints can be satisfied by a new set of field strengths for the quads comprising the hypothesis.

The success of this fitting procedure can be measured by several different parameters. One can use the convergence metric returned by the fitting procedure or the closeness of the fitting results to the target R matrix values. A final and decisive criterion of success involves inserting the newly calculated field strengths for the corrected quads back into the model to determine whether the bad region has been corrected and merges with the surrounding good regions.

This procedure was tested on simulated data from SPEAR using simulated quadrupole errors. Testing over a range of scenarios, with reasonable noise levels (< 10%), and from one to three quad errors per bad region, demonstrated the effective of this procedure within the tested ranges. We found that this procedure uniquely identifies the miscalibrated quads and successfully recalibrates their field strengths. Tests are currently underway on real data from SPEAR to determine whether this procedure will successfully correct quadrupole errors in the current model to produce a better model of the actual machine.

One issue worth noting is that this procedure is affected by a combinatorial explosion of hypotheses. Basically, any subset of the quadrupoles in a bad region constitutes a valid hypothesis. Since the number of subsets of a set grows exponentially with the size of the set, the hypothesis space grows exponentially. For this reason, heuristics are required to focus the search efficiently in order to find the correct hypotheses with a computationally tractable amount of work.

The heuristic that we tested worked effectively within the limited testing range (up to three quad errors per region). This heuristic involves testing hypotheses in order of cardinality (number of quads). In the future we plan to use *best-first search* [5] with a heuristic function consisting of the convergence metric from fitting to order the search of the hypothesis space.

4. Description of Work

This work was performed in coordination with Dr. Martin Lee from SLAC. Most programming was performed at Vista's site in New Mexico while most knowledge engineering and technical analysis was performed during the PI's visits to SLAC. The Vista PI, Dr. Stern, made four trips to SLAC during the Phase I period.

Work began in late September 1998 with our first trip to SLAC. Dr. Lee provided a theoretical description of the problem and the general technical approach that would be employed on this project. This included a description of a multi-track algorithm for decomposing the beamline into good and bad regions. The discussion also covered some rough ideas that Dr. Lee had at the time regarding methods for identifying and correcting field strength errors in bad regions. Finally, Dr. Lee provided necessary information about the modeling code COMFORT and the SPEAR beamline model as represented in COMFORT.

These first technical discussions defined a set of initial programming tasks, including construction of an initial prototype implementing the multi-track decompositional analysis procedure. Programming tasks also included integrating COMFORT with the prototype, and using COMFORT in conjunction with the SPEAR model to build a simple simulation useful for testing.

The second trip to SLAC followed in late October. During this visit we experimented with the first version of the prototype, using the prototype to analyze scenarios generated by the simulation. The multi-track method for analyzing orbit response data to decompose the model into good and bad regions was tested and found effective and robust. At this time we also experimented with two methods for analyzing errors in bad region, applying these methods manually to data generated by the simulation. Finally, Dr. Lee proposed the local component calibration method described above, and initial testing of this method gave very positive results.

Between the second and third trips to SLAC an unanticipated opportunity to exercise and test the prototype was presented. The high energy ring (HER) of PEP-II at SLAC, in commissioning stage at the time, was affected by orbit errors that required analysis and correction. After approximately a week of effort to modify the Phase I prototype and integrate the HER model, the prototype was applied to data from HER. Because of the nature of the available data, only single track analysis (see the section above, *Analysis of multi-track orbit response data*) was possible. However, the prototype clearly identified two bad regions in the HER beamline. The first has been confirmed by independent analysis and the second is still under study.

The focus of the third trip to SLAC in December was 1) experimentation to evaluate the effectiveness of the prototype, and 2) identification and evaluation of knowledge-based heuristics to further improve the reliability and robustness of analytic procedures. We evaluated heuristics for addressing a variety of problems:

1. Data preprocessing to flag and filter our bad data points and data sets;
2. Selection of appropriate epsilon values for fitting based on the stability of results in decompositional analysis;
3. Selection of robust region boundaries;
4. Search strategies and evaluation functions for finding and identifying optimal error hypotheses.

We observed that human experts in model analysis rely on heuristics, particularly certain kinds of pattern-based triggers, for recognition of anomalous data as well as recognition of "bad regions". Bad BPM data, for example, is identified by a certain "signature", which includes an acceptably tight fitting curve that breaks at a single point while continuing smoothly past that point to subsequent BPMs. Bad or noisy tracks have a similar characteristic signature, i.e., region boundaries systematically inconsistent with the region boundaries produced analysis of the remaining tracks.

We experimented by varying the size of epsilon, the fitting tolerance for a least squares fit, in the analysis of good and bad regions. We monitored the variation in the number of regions and the position of region boundaries as a function of the size of epsilon. Finding the transitional values of epsilon with respect to the appearance, movement, and break-up of stable regions, provided us a heuristic for selecting optimal values for epsilon. These optimal epsilon values, in turn, provide us with parameters useful for determining robust region boundaries.

We also did multiple runs of our new method for identifying and correcting miscalibrated quadrupoles in a bad region. We experimented with a variety of search strategies for addressing the combinatorial explosion of potential error hypotheses. We found the strategy of searching over single quadrupole-hypotheses first, then two-quadrupole hypotheses, then three-quadrupole hypotheses, etc., worked quickly and efficiently over small search spaces. Combining this with a simple heuristic yielded further efficiency. This heuristic is to order the search over N-quadrupole hypotheses by the quality of fitting found for the N-1 quadrupole hypotheses. Intuitively, this means N-quadrupole hypotheses that are supersets of the good fitting N -1 quadrupole hypotheses are searched first. This heuristic is, of course, rather primitive. We hand tested a more sophisticated heuristic, *best-first search* [5]. It provided an even more focused and efficient search strategy. Further work on evaluation functions *for best-first search* in this context is expected to yield additional improvements in efficiency and robustness.

5. *Feasibility Issues Related to Customization and Generalization*

Two major questions or issues related to the feasibility of our long term design goals are:

- How can the procedures in this application be implemented in a customizable way so that the user can selectively specify additional heuristic and task knowledge that will effectively control program behavior?

- How can the procedures comprising our prototype be reimplemented in a more general way so that their basic functionality can be incorporated into a wider range of applications?

Our first concern in this study was to determine where, i.e., at what points in the analytic algorithms described above, heuristics could be profitably employed. Then, for each application of heuristics, we examined how such the selection of heuristics might be reasonably incorporated into the user interface. Analysis of the experiments described above, i.e. applying a variety of heuristics to specific procedures in the prototype, has led to an initial design supporting user customization. This design currently addresses customization of epsilon selection, data analysis and filtering, and user control over various parameters of the search engine.

We view this as only the first step in decomposing the program into a set of more flexible functions and procedures that can be readily assembled and customized as appropriate for new hardware contexts and new analytic problems. The shape that such a function and procedure library will take is prefigured to some extent in the set of functions and procedures implemented in the Phase I prototype. In our current view, the following procedures are likely candidates for inclusion in a toolkit library:

- A variety of data analysis procedures for detecting and flagging bad data.

- A variety of procedures for generating local model-based predictions.
  This will include predictions that take arbitrary entrance conditions (within a realistic

range) and calculate the effects of beam propagation over a local region.

- A local matching procedure.
  This procedures interfaces to the procedures above to generate set of local model-based predictions. It then attempts to match these model-based predictions with observational data. It will include the capability of capturing global error at the entrance to the local region, as described in the section on *Hierarchical Decomposition* above.

- A global model decomposition procedure.
  This procedure decomposes the model into good and bad regions based on an interpretation of the results of local matching returned by the procedure above.

- A variety of local component recalibration procedures.
  These are procedures that search through components in a bad region to identify components that are miscalibrated and correct the errors. The current prototype applies the fitting procedures in the modeling code COMFORT, attempting to fit the quadrupole field strengths to reproduce the observed orbit displacements in the bad region. This is only one possible method for error identification and recalibration. There must be a variety of such methods incorporated into the toolkit because different types of problems will necessitate different approaches. This will include gradient descent error minimization techniques as well as other search methods, such as genetic algorithms and simulated annealing, that are less prone to becoming trapped in local minima. What all such methods will have in common is that they are methods for searching over perturbations to the model in a local region, looking for perturbations that best bring the model's predictions back into agreement with observed behavior.

Clearly achieving our ultimate goal, a general set of customizable tools that can be used to expedite the construction of analytic model-based applications, will require a great deal of further experimentation and study. It will be necessary, in particular, to develop several additional model calibration and parameter configuration applications using prototyped components from this toolkit in order to assess the full range of requirements for such a toolkit. This is essentially the work that we have proposed for Phase II.

6. Summary and conclusions.

We have constructed a prototype that automates error correction of quadrupole field strength errors. This system has been tested on simulated and real data from the SPEAR and HER rings at SLAC. It has performed successfully in identifying bad regions in these rings. In test simulations on SPEAR data it has successfully identified up to three miscalibrated quadrupoles per bad region.

The Phase I implements a new multi-track method for local error correction. This method searches over error hypotheses, testing each hypothesis by applying a newly designed matching procedure. This matching procedure attempts to optimize quadrupole field strengths for the quadrupoles specified by the error hypothesis, i.e., it attempts to find a set of field strengths that minimizes the discrepancy between model-based prediction and observation in the local region under analysis. The best error hypothesis, as that which support the best error minimization, is then selected and used for field strength recalibration in the local region.

An important issue addressed by our Phase I research is the automation of heuristic methods. Clearly, when human experts perform manual data analysis and data interpretation, they employ a variety of heuristics. They need to apply heuristic strategies for data analysis and filtering and specific pattern recognition capabilities in data interpretation, especially in the analysis of challenging situations and noisy data sets.

Similarly, we believe software tools for automatic data analysis and interpretation will also require the application of heuristics if they are to work effectively over a wide range of cases. Even with progress in automating certain heuristics, we expect that automated systems for model-based analysis will require the assistance of a skilled human user, at least in the short term, when confronted with difficult or anomalous cases.

Our goal however, is to automate a set of heuristics sufficient to handle the large majority of cases, specifically the normal and near normal cases. We have begun this process, showing that it is possible to incorporate at least some automated heuristics for data analysis and pattern recognition capabilities. Our current design includes some of these as user selectable options for customizing program behavior to meet the requirements of the current problem. We see the process of automating heuristics as an ongoing incremental process, progressively expanding the range of cases that the system can handle automatically and reducing both the time requirements and cognitive burden on the human user.

Phase I work has also made progress towards our long term goal of developing a set of software tools for constructing model-based applications. We are working towards a more general and flexible design that decomposes the integrated functionality in our Phase I prototype into a library of discrete functions and procedures that can be assembled to construct model-based applications. We have discussed in the preceding section how the Phase I prototype has contributed to that development process by contributing insight into the necessary functionality and the outline of a potential design.

REFERENCES

[1] Lee, M. J. (SLAC). GLAD: A Generic Lattice Debugger. SLAC-PUB-5700, Nov 1991. Invited talk given at International Conference on Accelerator and Large Experimental Physics Control Systems, Tsukuba, Japan, Nov. 11-15, 1991.

[2] Lee, M. J. and Clearwater, S. H. 1987. Generic orbit and lattice debugger. *Proceedings of the Model Based Accelerator Controls Workshop*, Brookhaven National Laboratory.

[3] Lee, M. J. (SLAC), Clearwater, S. H. (Los Alamos), Kleban, S. (SLAC). Finding Beam Errors Automatically. SLAC_CN-351, (Received Jun 1987).

[4] Lee, M. J. (SLAC), Clearwater, S. H. (Los Alamos). GOLD: Integration of Model Based Control Systems With Artificial Intelligence and Workstations. SLAC-PUB-4396, Aug 1987. Invited talk at Workshop on Model Based Accelerator Controls, Upton, N.Y., Aug 17, 1987.

[5] Luger, G.F. & Stubblefield, W.A. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving,* Third Edition, London: Addison-Wesley, 850 pages, 1998.