Modeling and Simulation of System of Systems via Data Analytics – The Case for “Big Data” in SoS

Mo Jamshidi, Ph.D., DEgr. (h.c.), Fellow IEEE

Lutcher Brown Endowed Chaired Professor
Member, Univ. Texas System’s Chancellor’s Council
Department of Electrical and Computer Engineering
and Director Autonomous Control Engineering - ACE Laboratory
The University of Texas, San Antonio, Texas
www.wacong.org/ace
moj@wacong.org, mo.jamshidi@utsa.edu

2013 LABEX Workshop
Compiegne, France
September 4, 2013
1. Introduction to System of systems – SoS (SoS vs. Cyber-Physical Systems - CPS)
2. The need to go from SoS to SoS Engineering
3. Modeling of SoS via BIG DATA Analytics
4. Micro Grid and private cloud at the University of Texas, San Antonio - Open Cloud Paradigm
5. Conclusions and videos and movies
Preliminary Comments

- Internet has connected people of the world since ~ 1995

- IT has given us crossing of cyber and physical spaces or Cyber-physical systems - CPS

- **System of Systems** is a generalization of connectivity of systems or systems and people – or a more comprehensive crossing of cyberspace and physical space! Connectivity of CPS is the Internet, while that of SoS is more general infrastructure.

- SoS and/or CPS is getting into new application cases in a very persisting pace – from IT to defense, energy, space, environment, healthcare, services, earth studies,
A SoS is an integration of a finite number of constituent systems which are independent and operatable, and which are networked together for a period of time to achieve a certain higher goal. (Jamshidi, 2005, 2009)

- Operational independence of component systems
- Managerial independence of component systems
- Geographical distribution
- Emergent behaviour
- Evolutionary development processes

(Maier, 1998)
SoS relationship to CPS

• Cyber-Physical Systems (CPS)
  – concern increasing use of embedded software
  – direct and real-time interface between the virtual and physical worlds.

• CPS is a special type of SoS
  – CPS interconnectivity restricted to cyberspace
  – SoS interconnectivity is ubiquitous (e.g. NCOIC Interoperability Framework) … CPS ~ Virtual SoS

• Similarities with SoSE, but different emphasis.
  – CPS is mainly concerned with control
  – SoSE is concerned with configurations, authorities and responsibilities, interoperability, and emergent behaviour
SoS vs. CPS

Systems Engineering

SoS

CPS

Computer Science
What is a system of systems

Unanticipated benefits of SoS extension beyond MP3 player (Blogs, PODCAST) or Internet purchases

Others: **iPAD** **iPHONE**

Retail businesses

Freeways

Transportation SoS: Roads +GPS+ ONSTAR

Aircraft
FROM SoS TO SoS
ENGINEERING?
System Engineering  
**vs.**  
SoS (CPS) Engineering  

System Engineering is a discipline  
(at least 5 decades old)  

**BUT**  

SoS (or CPS) Engineering  
(at the present time)  
is only an opportunity
Challenges … SOS ➔ SoSE

- Control Theory
- Cyber-Security
- Simulation
- Artificial Intelligence
- Emergence
- Net Centricity
- Ethics
- Architecture
- Modeling (Big Data)
MODELING of SoS via Big Data Analytics
SoS BIG DATA ANALYTIC

Inputs

SoS

Model

Outputs
SoS BIG DATA ANALYTIC

Modeling of System of Systems via Data Analytics – Case for “Big Data” in SoS
Objective

– Advances in sensor technology, the Internet, wireless communication, and inexpensive memory have all contributed to an explosion of “Big Data”.

– Recall: SoS are integrated independently operating systems to achieve a higher goal than the sum of the parts.
Objectives, Cont’d

Today’s SoS are also contributing to the existence of unmanageable “Big Data”. Recent efforts have developed a promising approach, called “Data Analytics”, including:

• K-means Clustering, Fuzzy Clustering/Fuzzy Inference System Model Generation (FCM,FIS)
• Neural Network Model Training
• Principal Component Analysis (PCA)
• Regression Analysis

Can use CI (computational intelligence) and statistical tools of the “Data Analytics” as they apply to any type of data, here we use photovoltaic (or wind) energy forecasting problem for a micro-grid SoS.
Micro-Grid SoS

This micro-grid represents a facility scale a SoS consisting of a building with:

- A solar photovoltaic system
- A load demand in the form of overall energy consumption
- A reconfigurable control and acquisition system

Manjili, et al., 2012
Source of Data

• Location
  – Three data sets acquired for the area surrounding Golden, CO were combined for to produce the training data for this project.

• Source 1: Solar Radiation Research Laboratory (SRRL)/National Renewable Energy Laboratory (NREL)
  http://www.nrel.gov/midc/srrl_bms/
  – Uses over 70 instruments to measure solar conditions and environmental parameters.
  – Also includes 180° images of the sky that can be used to determine current cloud conditions directly

• Source 2: SOLPOS Data, Measurement and Instrumentation Data Center (MIDC)
  http://www.nrel.gov/midc/solpos/solpos.html
  – Contains information on solar position and available solar energy

• Source 3: Automated Surface Observing System (ASOS), the Iowa Environmental Mesonet (IEM)
  http://mesonet.agron.iastate.edu/ASOS/
  – Contains weather parameters in the set
PV Data Analysis Goals

• Final Data Set
  – Consists of data points for half the month of October 2012
  – Contains approximately 250 variables for each data point
  – Data points containing invalid variable values removed prior to analysis

• Objective of Analysis
  – Three key parameters representing total received solar energy were selected as the output of the analysis:
    • **Global Horizontal Irradiance (GHI)**
    • **Direct Horizontal Irradiance (DHI)**
    • **Direct Normal Irradiance (DNI)**
  – The goal of the models generated was to predict values for these three parameters 60 minutes in advance.
PV Data Analysis Setup

– Due to execution time and memory limitations, the initial set of variables used as the input data set was 13 parameters including:

• Time of Day
• Cloud Levels
• Humidity
• Temperature
• Wind Speed
• Current values of GHI, DHI, and DNI

– Data set further narrowed down to exclude points with very low values of Irradiance - GHI, DHI, and DNI

Output Parameters for a Clear Day
PV Data Analysis Tools

• Non-Parametric Model Generation Tools
  – Two tools used to generate models for this SoS:
    • GENFIS3: Fuzzy Inference System Generation Based on Fuzzy C-Means Clustering
    • NFTOOL: Back-Propagation Trained Feed Forward Neural Network
  – GENFIS3 (MATLAB Fuzzy Logic Toolbox)
    • All input and output variables clustered using Fuzzy C-Means Clustering to generate the fuzzy membership functions
    • Rules then generated to best match the training data set
    • Membership functions and rules can be viewed using the MATLAB FIS GUI tools, such as ruleview
PV Data Analysis Tools

- Non-Parametric Model Generation Tools (cont.)
  - GENFIS3 (MATLAB Fuzzy Logic Toolbox) (cont.)
    - Membership functions and rules can be viewed using the MATLAB FIS GUI tools, such as ruleview
    - This sample shows the output produced with five input variables and three output variables.
    - Four fuzzy membership functions produced per variable.
PV Data Analysis Tools

Evaluation of GHI Solar Irradiance Predicted using GENFIS3

Observed W/m²

GHI Shifted (Actual)

GHI Shifted (Predicted)

Time (Daytime Minutes)

Sample Output Predictions, 13 Input Parameters
PV Data Analysis Tools

- **Non-Parametric Model Generation Tools (cont.)**
  - **NFTOOL (MATLAB Neural Network Toolbox)**
    - By default, uses the Levenberg-Marquardt back propagation method to select weights and biases to minimize the mean squared error performance of a the feed forward network.
    - For this project, the feed forward network contains one hidden layer consisting of ten neurons.

![Initial Feed Forward Network Architecture](image-url)
PV Data Analysis Tools

- Non-Parametric Model Generation Tools (cont.)
  - \textit{NFTOOL} (MATLAB Neural Network Toolbox) (cont.)
  - Performance Training Graphs for 13 Input Parameters
PV Data Analysis Tools
• Non-Parametric Model Generation Tools (cont.)
  – *NFTOOL* (MATLAB Neural Network Toolbox) (cont.)
    • Performance Training Graphs for 13 Input Parameters

Sample Output Predictions, 13 Input Parameters
PV Data Analysis Tools

• **Additional Pre-Processing Efforts**

  – To improve the performance of the generated models, two different paths were investigated
    • Nonlinear Derived Parameter Generation
    • Larger Training Data Sets
  – Nonlinear Parameter Generation
    • Objective: To calculate nonlinear parameters from the original set of input parameters
    • Calculated parameters include:

      \[
      \begin{align*}
      x(t)^2 & \quad \text{slope}(x(t-60):x(t)) \\
      \sin(x(t)) & \quad \text{mean}(x(t-60):x(t)) \\
      \cos(x(t)) & \quad \text{stdev}(x(t-60):x(t)) \\
      \text{slope}(x(t-1):x(t)) &
      \end{align*}
      \]
PV Data Analysis Tools

- Additional Pre-Processing Efforts
  - Nonlinear Parameter Generation (cont.)
    - This graph reflects the performance improvement after incorporating derived nonlinear components into the training data set.
  - **Note** that the calculation of nonlinear components excludes data points for which the derived calculations are invalid (missing data points, etc.)

Regression Performance Improvement with Nonlinear Input Data Set
PV Data Analysis Tools

• Additional Pre-Processing Efforts
  – Larger Training Data Sets
    • Objective: To improve trained model performance by including a larger number of input parameters
    • First, the performance of the initial 13 parameters was compared to a 32 parameter data set.
      – As expected, performance increased accordingly
    • Next, the training of models using the entire input set (approximately 250 variables) was attempted, but neither model generation tool would execute with this number of variables.
    • Principal Component Analysis (PCA) was then used to reduce the dimension of the full initial data set
    • PCA was also conducted on the data set that had been expanded to include all calculated nonlinear parameters
PV Data Analysis Tools

- Additional Pre-Processing Efforts
  - Principal Component Analysis (cont.)
    - PCA used to reduce the dimension of the data while minimizing the information lost

Graph Showing Information Contained In Principal Components

Data Recovery Using Only 50 out of ~250 Principal Components
Results

- The best performing model was generated using \textit{NFTOOL} with an input data set generated as follows:
  1. Start with all initial $\sim250$ variables
  2. Calculate nonlinear components for each variable, expanding the data set dimension to $\sim1950$
  3. Perform PCA to drop the dimension of the data to 150
- The best performing \textit{GENFIS3} generated model used the same input data set, but with a post-PCA dimension of 50
Results

• Discussion of Results (cont.)
  
  • Best Model Performance

Best Linear Regression Performance

Best Model GHI Prediction Error
Results

- Discussion of Results (cont.)
- Best Model Performance (cont.)

Evaluation of DHI Solar Irradiance Predicted using NFTOOL

Evaluation of DNI Solar Irradiance Predicted using NFTOOL

Best Model DHI Prediction Error

Best Model DNI Prediction Error
Final Comment

– This work describes the steps necessary to use “Big Data” acquired using instrument stations to predict received solar power using Data Analytics tools to support a micro-grid

– The use of Fuzzy C-Means Clustering and Feed Forward Neural networks in conjunction with Principal Component Analysis successfully yielded predictions of GHI, DHI, and DNI
  
  • The best performing model used a Neural Network trained using preprocessed input data shrunk to a usable dimension using PCA
  
  • Both non-parametric models performed better than a simple, sub-optimal predictor
  
  • The training time was significantly better for the neural network

– Future work is planned to include
  
  • Optimal dataset reduction using Genetic Programming
  
  • Using cloud computing to train models using large datasets
  
  • Design and evaluate a controller to buy or sell energy to the grid based on demand and predictions of received solar energy.
Vehicle Electrification

1 Level 3 Fast DC Charging Stations

1 Level 2 Charging Stations
Sensor Network Architecture at University Center III

UCIII Building

1. Irradiance sensors (2)
2. Solar panel temperature thermocouples (4)
3. Wind Speed (1)
4. Wind Direction (1)
5. Rain Fall (1)
6. Ambient temperature (1)
7. Power Quality (5)
8. Voltages (53)
9. Currents (53)
Open Source Cloud computing will meet the needs of public and private Cloud providers by being simple and massively scalable.

- Software to build a cloud anywhere
- An ecosystem devoted to innovation
- Flexibility in deployment/features
- Standards for broad deployment
- No fear of lock-in
A common platform is here.
OpenStack is open source software powering public and private clouds.

Private Cloud: Run OpenStack software in your own corporate data centers

Public Cloud: OpenStack powers some of the world's largest public cloud deployments.

OpenStack enables cloud federation
Connecting clouds to create global resource pools

Common software platform making Federation possible

1. Virtualization
2. Cloud Data Center
3. Cloud Federation

Automation & Efficiency
## OpenStack @ Research Institutes and Academia

A short list of research institutes and universities:

<table>
<thead>
<tr>
<th>Research Institute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Diego Supercomputer Center</td>
<td>Unleashes the Value of its User Data by leveraging SWIFT</td>
</tr>
<tr>
<td>Argonne National Laboratory</td>
<td>Building High Performance Clouds</td>
</tr>
<tr>
<td>CERN</td>
<td>Overlay opportunistic clouds in CMS/ATLAS at CERN</td>
</tr>
<tr>
<td>National Security Agency – NSA</td>
<td>OpenStack at the National Security Agency</td>
</tr>
<tr>
<td>Brookhaven National Laboratory</td>
<td>US Department of Energy</td>
</tr>
<tr>
<td>MIT CISAL</td>
<td>OpenStack for scientific research</td>
</tr>
<tr>
<td>Massachusetts' open cloud (MOC)</td>
<td>HPC Cloud for Scientific Research and BigData</td>
</tr>
<tr>
<td>University of Texas at San Antonio</td>
<td>Build Advanced OpenStack-based Software Open Cloud Platform for Academic Research</td>
</tr>
<tr>
<td>University of Victoria</td>
<td>Clouds in High Energy Physics</td>
</tr>
<tr>
<td>University of Melbourne Australian Government</td>
<td>Building a Cloud for National Collaborative Research NeCTAR</td>
</tr>
<tr>
<td>UCLA</td>
<td>High Performance Computing Cloud</td>
</tr>
<tr>
<td>Purdue University</td>
<td>Cloud Infrastructure</td>
</tr>
<tr>
<td>Industrial Technology Research Institute (ITRI) Taiwan</td>
<td>Cloud Infrastructure</td>
</tr>
<tr>
<td>University of Southern California – USC</td>
<td>Information Sciences Institute</td>
</tr>
</tbody>
</table>
OpenStack Clouds for Scientific Research

Create a scalable and collaborative hybrid cloud environment accessible for scientific applications to enable Open Science:

- OpenStack-based Private Cloud among universities for hosting, deploying and sharing research applications
- Enable their OpenStack-based Private Cloud to bursting to Open Cloud through Internet2
- Sharing code, data, and platform openly for research empowerment

<table>
<thead>
<tr>
<th>Institution</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet2</td>
<td>Stephen Wolff, Khalil Yazdi</td>
</tr>
<tr>
<td>University of Texas at San Antonio</td>
<td>Mo Jamshidi, Raj Boppana, Daniel Pack</td>
</tr>
<tr>
<td>Massachusetts' Open Cloud (MOC)</td>
<td>Chris Hill (MIT), Orran Krieger (Boston), Michael Goroff (UMass)</td>
</tr>
<tr>
<td>Argonne National Lab</td>
<td>Narayan Desai</td>
</tr>
<tr>
<td>Notre Dame</td>
<td>Paul Brenner (Notre Dame)</td>
</tr>
<tr>
<td>Facebook</td>
<td>John Kenevey, Charlie Manese</td>
</tr>
<tr>
<td>Mellanox</td>
<td>Arik Kol, Eli Karpilovski, Marina Lipshteyn</td>
</tr>
<tr>
<td>80/20 foundation</td>
<td>Lorenzo Gomez, Jeff Prevost</td>
</tr>
</tbody>
</table>

Some books below are being offered free of charge by Mo Jamshidi, Ph.D. These books cover Large-Scale Systems, Intelligent Control Systems, Soft Computing Methodologies, Analysis and Design of Linear Time-Delay Systems.

More information is also provided to find books on Amazon.com that are still protected under copyright and cannot be offered for free. ** For a pre-publication free access to moj@wacong.org **

About the Author

Mo. Jamshidi, Ph.D.

Educational Background:
Ph.D., 1971, University of Illinois, Champaign

Editor-in-Chief: AutoSoft Journals
Advisory Committee Chair: IC-ICCE
Autonomous Landing of a quadcopter on a mobile rover
CONCLUSIONS

• SoS (CPS) has been with us for some time Soon … System integration will be a matter of necessity and not Choice

Modeling via Big Data Analytics has the most promise, as scientific principles can not easily be used to model constituent systems.

• Discrete-event simulation has some promise to handle SoS Simulation.

• Renewable energy sources will continue to contribute to the energy picture, currently at only 5% of total energy, but in 20 years it could reach 20%.

• Application areas of SoS will soon be enhanced towards reality via Big Data Analytics.
QUESTIONS ... ?

Merci Beaucoup
To
LABEX Program

Please send your additional questions to moj@wacorg.org
Quadcopter autonomous flight via network control

https://www.dropbox.com/s/5ywyggqynrttz1m/Drone%20Demo.wmv

Introduction to cloud computing

http://wacong.org/ace/research.html

Cloud Technologies Institute (Rackspace, UTSA)

http://www.rackspace.com/blog/how-the-open-cloud-powers-academic-and-scientific-research/
Autonomous Landing of a quadcopter on a mobile rover