Fault Detection in Distributed Climate Sensor Networks using Dynamic Bayesian Networks

George Chin

Sutanay Choudhury

Lars Kangas

Andres Marquez

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Distributed Climate Sensor Validation

- The Atmospheric Radiation Measurement (ARM) climate research facility is a Department of Energy scientific user facility that provides continuous field observations of cloud, aerosol, and radiative processes to advance global climate models
- ARM instruments run unattended 24 hours, 7 days a week
- Each site has > 20 instruments and is measuring hundreds of data variables
- Data is examined by the ARM Data Quality Office who manually examine hundreds of data plots a day
- Although there are a few redundant measurements, most instruments are measuring unique variables
- Subtle instrument problems (such as sensor drift) can easily go undetected
- However, the atmosphere is a connected system – combining measurements of different atmospheric parameters from multiple instruments can identify unphysical or problematic data

Simple Climate Sensor Validation Example

- A simple example involves the Optical Rain Gauge (ORG) and the Microwave Radiometer (MWR)
- The ORG uses an infrared laser to measure rain rate; very often it is not raining so long periods of 0 values are not surprising
- The MWR measures microwave radiation emitted by water vapor and liquid in the atmospheric column
- During rain events, water droplets land on the MWR radome and produce high microwave signals
- → Very high MWR signals with no corresponding ORG values > 0, could indicate problem with ORG

More Complex Climate Sensor Validation Example

- When a new instrument is installed, incorrect calibration factor is sometimes assigned to the instrument logger
- IRT (infrared thermometer) measures amount of infrared radiation emitted by the atmosphere – function of temperature, water vapor, and cloud height and thickness
- Combining information on cloud height (ceilometer), sky cover (total sky imager), temperature and humidity (surface meteorology) could indicate calibration errors in IRT measurements

Distributed Climate Sensor Validation

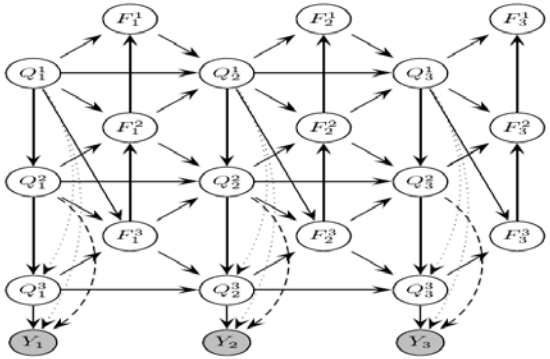
- ARM instruments run unattended 24 hours, 7 days a week
- Each site has > 20 instruments and is measuring hundreds of data variables
- Data is examined by the ARM Data Quality Office who manually examine hundreds of data plots a day
- Although there are a few redundant measurements, most instruments are measuring unique variables
- Subtle instrument problems (such as sensor drift) can easily go undetected
- However, the atmosphere is a connected system – combining measurements of different atmospheric parameters from multiple instruments can identify unphysical or problematic data

Dynamic Bayesian Network Model for Climate Sensor

- Bayesian networks are represented as direct acyclic graphs (DAGs)
  - Nodes = random variables
  - Arcs = conditional dependencies
- Probabilistic inferencing is process of computing conditional probability for query variables given evidence or values of observed variables
- Dynamic Bayesian networks model stochastic evolution of set of variables often over time
- With DBNs, discrete time is introduced and conditional distributions are related to parent variable values of the previous time point
- Applications: speech recognition, bionetwork analysis, social networks, computer networks

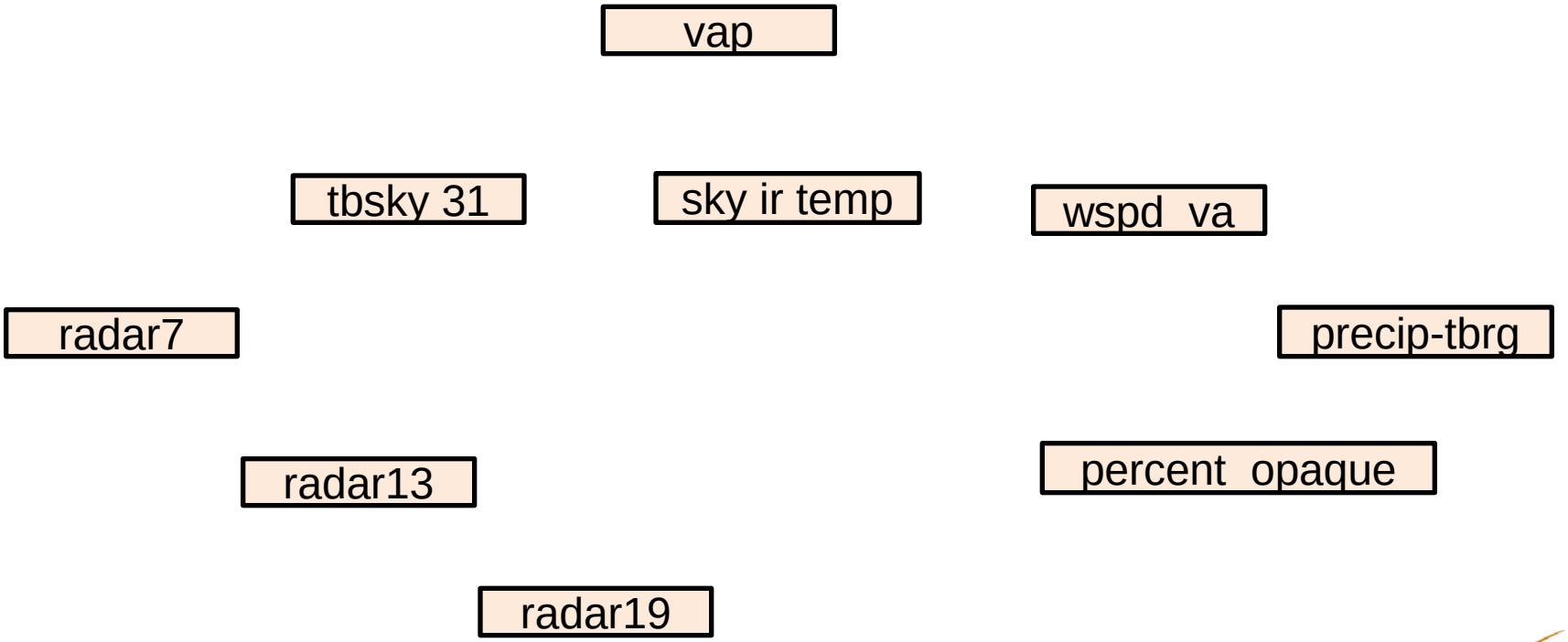
Conditional Probability Table (CPT)

E	B	P(A | E,B)	
e	b	0.9	0.1
e	¬b	0.2	0.8
¬e	b	0.9	0.1
¬e	¬b	0.01	0.99



'Hel' - 'lo' 'World'

Dynamic Bayesian Network Model for Climate Sensor Network Fault Detection



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Dynamic Bayesian Network Model for Climate Sensor Network Fault Detection

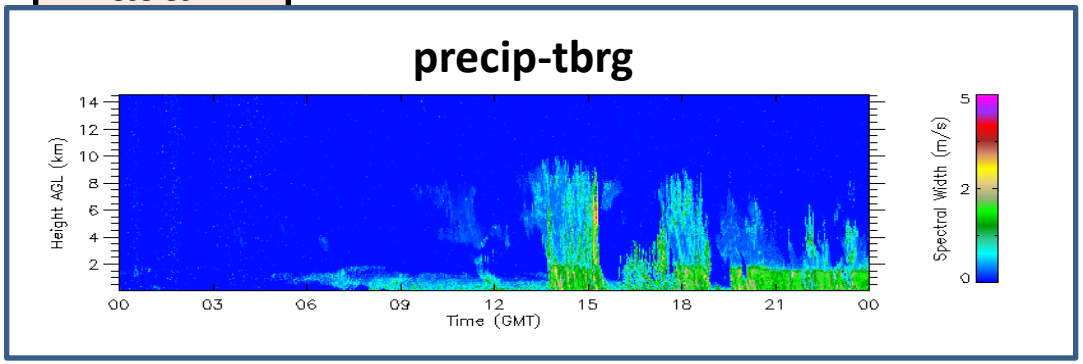
vap

tbsky 31

sky ir temp

< 0.001 – No rain
0.001-0.032 – Light rain
0.032-0.200 – Moderate rain
> 0.200 – Heavy rain

radar7



precip-tbrg

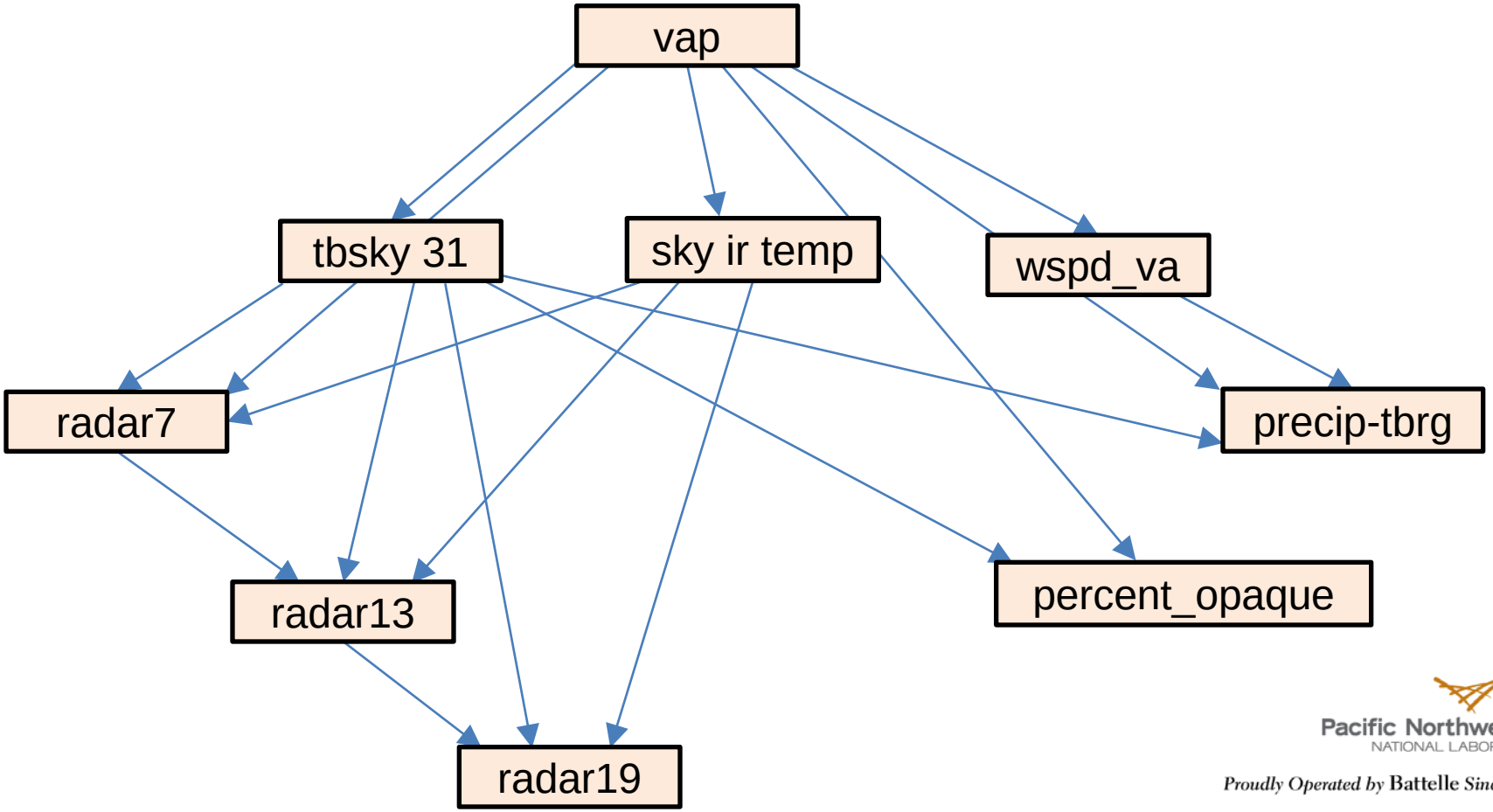
percent_opaque

Dynamic Bayesian Network Model for Climate Sensor Network Fault Detection

	Radar7	Radar13	Radar19	Radar25	Radar31	Radar37	Radar43	Radar49	Radar55	Lidar3	Lidar6	precip_tbrg	tbsky31	vap	sky_ir_temp	bar_pres	rh	temp	vap_pres	wspd_va	percent_opaque
Radar7	1.0000	0.5114	0.3876	0.2871	0.2197	0.1832	0.1102	0.0681	0.0481	0.0229	0.0047	0.0860	0.2820	0.1869	0.1973	0.0291	0.0028	0.0187	0.0493	0.0738	0.0783
Radar13	0.5114	1.0000	0.6990	0.5243	0.4091	0.3426	0.2125	0.1409	0.0945	0.0534	0.0249	0.0410	0.2599	0.1639	0.2813	0.0013	0.0600	0.0063	0.0303	0.0200	0.1436
Radar19	0.3876	0.6990	1.0000	0.7084	0.5472	0.4602	0.3063	0.1992	0.1303	0.0922	0.0543	0.0145	0.1573	0.1221	0.2230	0.0001	0.0689	0.0141	0.0227	0.0062	0.1220
Radar25	0.2871	0.5243	0.7084	1.0000	0.6778	0.5150	0.3591	0.2362	0.1613	0.1465	0.0973	0.0064	0.0980	0.0972	0.1882	0.0001	0.0497	0.0103	0.0172	0.0060	0.1085
Radar31	0.2197	0.4091	0.5472	0.6778	1.0000	0.6149	0.4086	0.2784	0.1794	0.1719	0.1377	0.0034	0.0718	0.0818	0.1726	0.0001	0.0411	0.0102	0.0117	0.0056	0.1172
Radar37	0.1832	0.3426	0.4602	0.5150	0.6149	1.0000	0.5784	0.3736	0.2318	0.1560	0.1473	0.0030	0.0531	0.0827	0.1644	0.0022	0.0341	0.0034	0.0210	0.0060	0.1147
Radar43	0.1102	0.2125	0.3063	0.3591	0.4086	0.5784	1.0000	0.5353	0.3375	0.1224	0.1322	0.0020	0.0274	0.0630	0.1270	0.0022	0.0180	0.0005	0.0204	0.0073	0.1037
Radar49	0.0681	0.1409	0.1992	0.2362	0.2784	0.3736	0.5353	1.0000	0.5493	0.1165	0.1361	0.0019	0.0202	0.0680	0.1295	0.0099	0.0190	0.0000	0.0355	0.0051	0.1111
Radar55	0.0481	0.0945	0.1303	0.1613	0.1794	0.2318	0.3375	0.5493	1.0000	0.1041	0.1146	0.0017	0.0151	0.0717	0.1155	0.0194	0.0156	0.0016	0.0520	0.0054	0.0839
Lidar3	0.0229	0.0534	0.0922	0.1465	0.1719	0.1560	0.1224	0.1165	0.1041	1.0000	0.8144	0.0004	0.0100	0.0339	0.0620	0.0006	0.0022	0.0014	0.0005	0.0000	0.0311
Lidar6	0.0047	0.0249	0.0543	0.0973	0.1377	0.1473	0.1322	0.1361	0.1146	0.8144	1.0000	0.0025	0.0020	0.0226	0.0406	0.0021	0.0078	0.0042	0.0016	0.0008	0.0251
precip_tbrg	0.0860	0.0410	0.0145	0.0064	0.0034	0.0030	0.0020	0.0019	0.0017	0.0004	0.0025	1.0000	0.5483	0.1256	0.1398	0.0043	0.0656	0.0090	0.0305	0.1004	0.0753
tbsky31	0.2820	0.2599	0.1573	0.0980	0.0718	0.0531	0.0274	0.0202	0.0151	0.0100	0.0020	0.5483	1.0000	0.3557	0.2831	0.0249	0.1126	0.0130	0.0589	0.0590	0.2237
vap	0.1869	0.1639	0.1221	0.0972	0.0818	0.0827	0.0630	0.0680	0.0717	0.0339	0.0226	0.1256	0.3557	1.0000	0.4175	0.2648	0.0379	0.0945	0.4608	0.0358	0.2047
sky_ir_temp	0.1973	0.2813	0.2230	0.1882	0.1726	0.1644	0.1270	0.1295	0.1155	0.0620	0.0406	0.1398	0.2831	0.4175	1.0000	0.0686	0.2104	0.0135	0.1641	0.0328	0.4489
bar_pres	0.0291	0.0013	0.0001	0.0001	0.0001	0.0022	0.0022	0.0099	0.0194	0.0006	0.0021	0.0043	0.0249	0.2648	0.0686	1.0000	0.0016	0.2666	0.4426	0.0523	0.0108
rh	0.0028	0.0600	0.0689	0.0497	0.0411	0.0341	0.0180	0.0190	0.0156	0.0022	0.0078	0.0656	0.1126	0.0379	0.2104	0.0016	1.0000	0.5595	0.0497	0.0450	0.1096
temp	0.0187	0.0063	0.0141	0.0103	0.0102	0.0034	0.0005	0.0000	0.0016	0.0014	0.0042	0.0090	0.0130	0.0945	0.0135	0.2666	0.5595	1.0000	0.2016	0.0727	0.0383
vap_pres	0.0493	0.0303	0.0227	0.0172	0.0117	0.0210	0.0204	0.0355	0.0520	0.0005	0.0016	0.0305	0.0589	0.4608	0.1641	0.4426	0.0497	0.2016	1.0000	0.0223	0.0225
wspd_va	0.0738	0.0200	0.0062	0.0060	0.0056	0.0060	0.0073	0.0051	0.0054	0.0000	0.0008	0.1004	0.0590	0.0358	0.0328	0.0523	0.0450	0.0727	0.0223	1.0000	0.0228
percent_opaque	0.0783	0.1436	0.1220	0.1085	0.1172	0.1147	0.1037	0.1111	0.0839	0.0311	0.0251	0.0753	0.2237	0.2047	0.4489	0.0108	0.1096	0.0383	0.0225	0.0228	1.0000

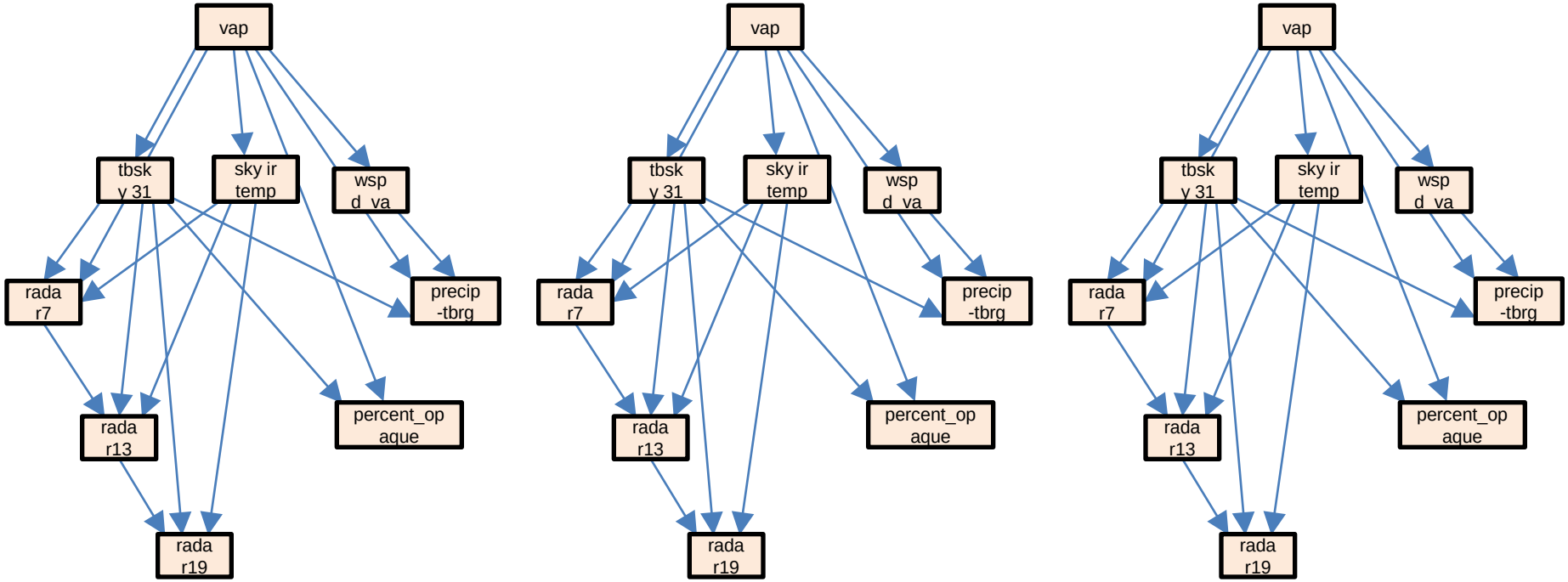
- Compute Pearson Correlation among all pairs of sensor measurements
- Add expert knowledge

Dynamic Bayesian Network Model for Climate Sensor Network Fault Detection



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

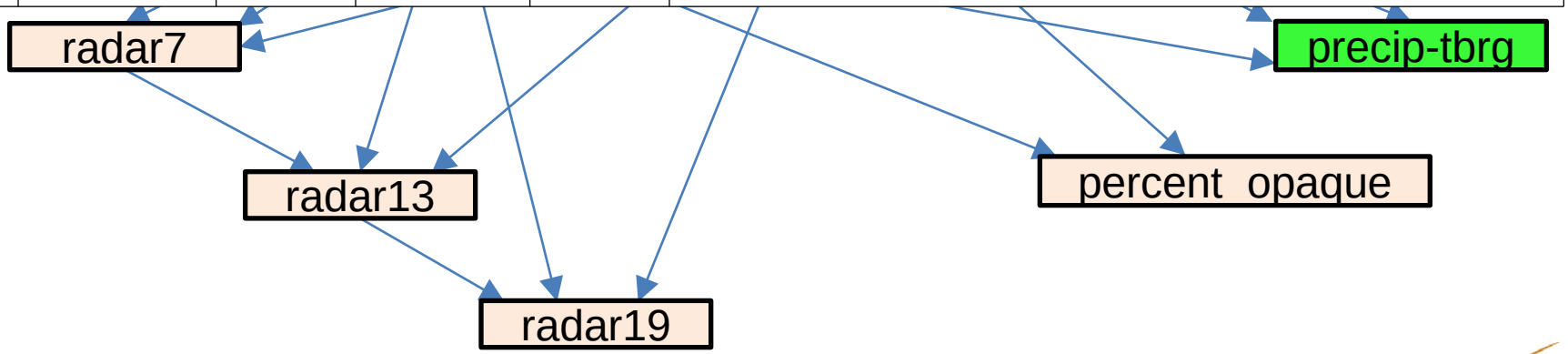
Dynamic Bayesian Network Model for Climate Sensor Network Validation



- Replicate per time step
- Add dependencies across time steps (not shown)

Dynamic Bayesian Network Model for Climate Sensor Network Validation

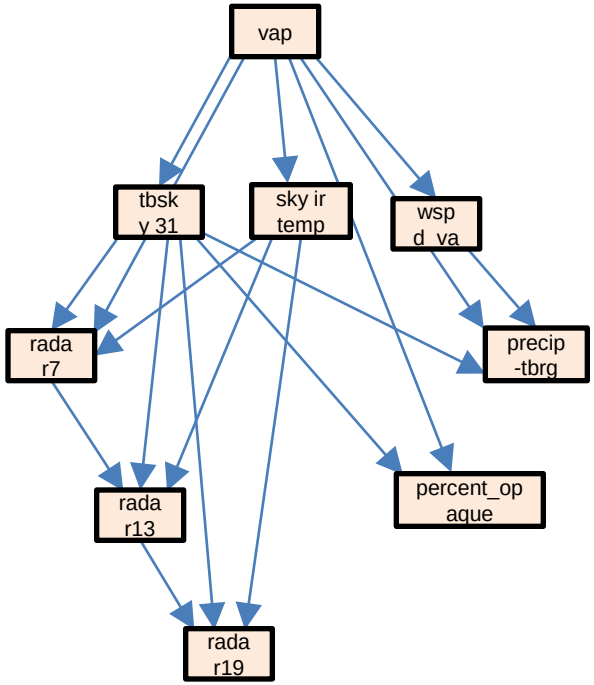
tbsky31	sky_ir_temp	wspd_va	prev precip_tbrg	…	P(precip_tbrg | tbsky31, sky_ir_temp, wspd_va, prev_precip_tbrg, prev_tbsky31, prev_sky_ir_temp, prev_wspd_va)
clear	low	none	none		none: 82.12%; low: 13.21%; med: 4.66%; high: 0.01%
partially	low	none	none		none: 82.32%; low: 13.73%; med: 3.94%; high: 0.01%
mostly	low	none	none		none: 83.86%; low: 14.17%; med: 1.96%; high: 0.01%
full	low	none	none		none: 83.95%; low: 14.55%; med: 1.51%; high: 0.01%
clear	medium	none	none		none: 87.12%; low: 12.09%; med: 0.78%; high: 0.01%
…	…	…			…

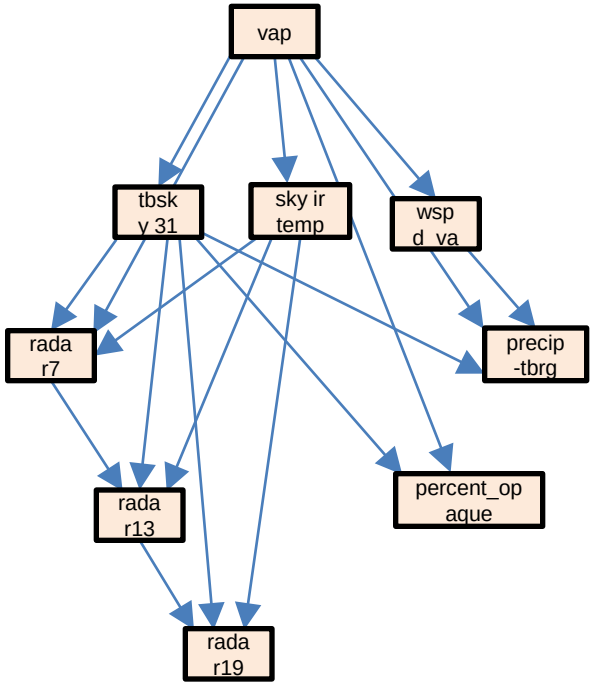


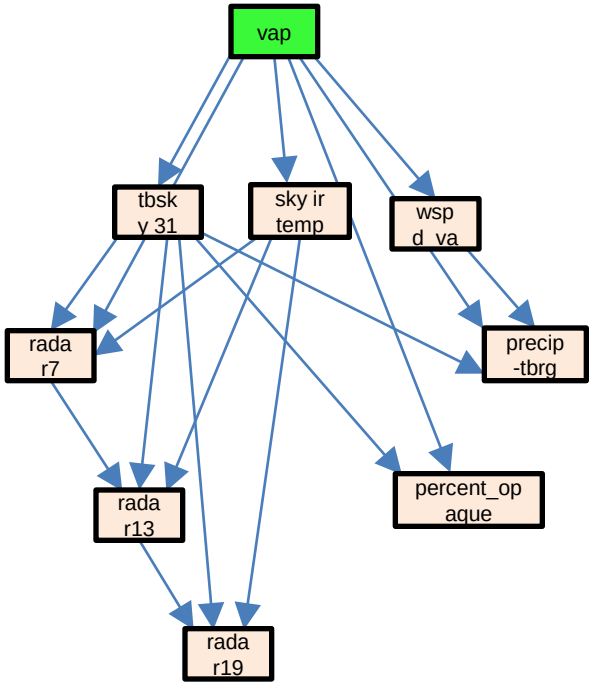
- Compute conditional probability tables (CPTs) using expectation-maximization algorithm

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

DBN Inferencing for Climate Sensor Network Fault Detection



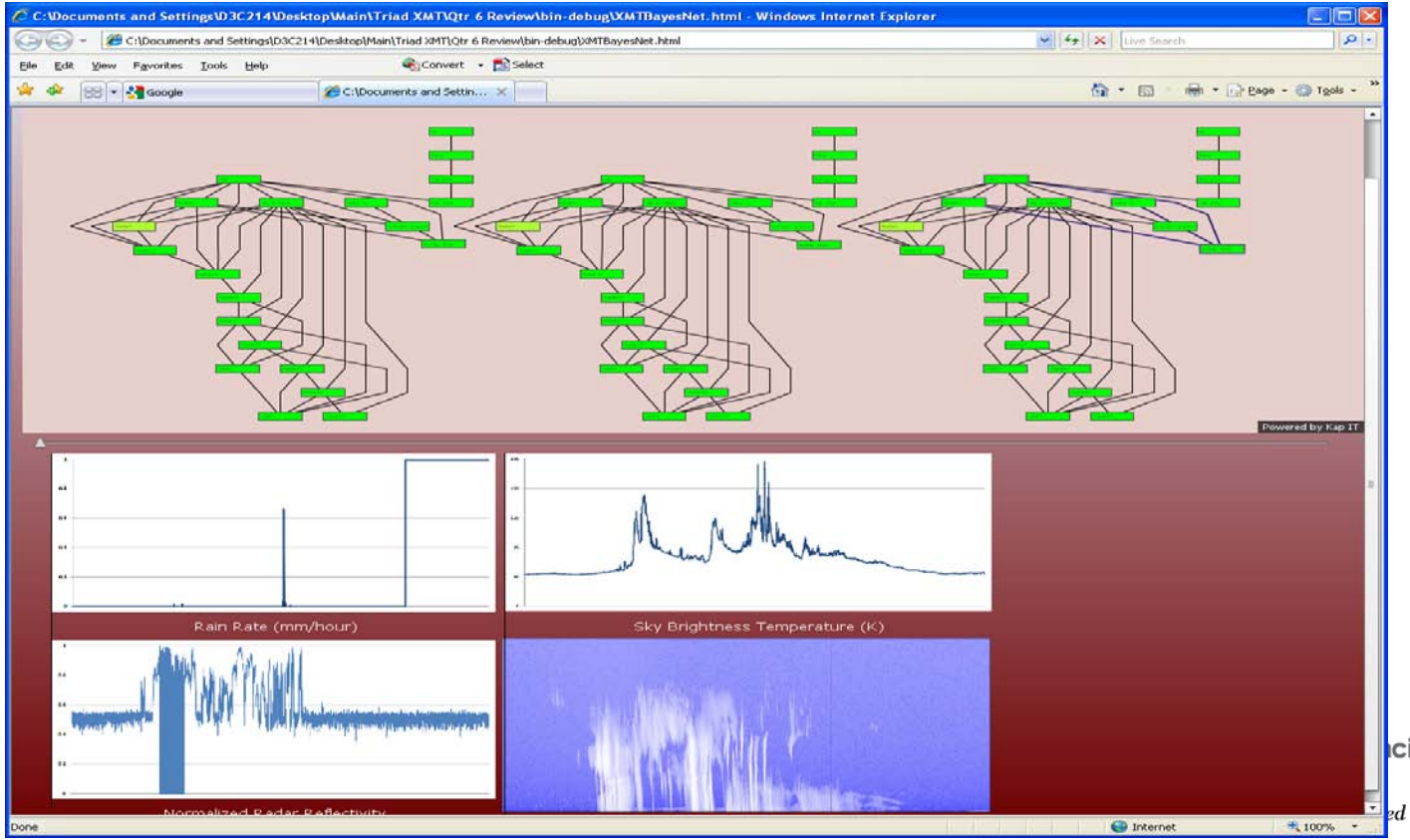




- Query each variable one at a time in last time step

DBN Inferencing for Climate Sensor Network Fault Detection

Climate Sensor Network Validation Demo



Dynamic Network Algorithms on the Cray XMT

- Center for Adaptive Supercomputing Software (CASS) Program
- Exploring different classes of dynamic network algorithms on the Cray XMT
- Inferencing, structure learning, and parameter learning of Bayesian networks are NP-hard
- Developing optimized exact inferencing dynamic Bayesian network algorithm on Cray XMT
- Applying XMT-based dynamic Bayesian network algorithm to detecting fault conditions in the ARM sensor network

Center for Adaptive Supercomputing Software (CASS)

- Cray XMT – Massive Multithreading
- 128 Processors, 1 TB memory
- 128 Hardware Threads/Processor
- Shared-Memory Architecture,
  Fine-grained Parallelism,
  Memory Latency Tolerance
- High Performance for Irregular Applications

Continuing Research and Development

- Extend dynamic Bayesian model to include more measurements and timesteps
- Enable near real-time analysis on data streams
- Utilize standardized combined data formats as they become available
- Investigate n-order dynamic Bayesian models for potentially better model accuracy
- Develop multiple concurrent models covering diverse time intervals
- Explore move from discrete to continuous variables to better support uncertainty quantification