Quasi Real time Data Analysis for Air Quality Monitoring with an Electronic Nose

H. Zhou, A. V. Shevade, C. C. Pelletier, M. L. Homer, M. A. Ryan

Jet Propulsion Laboratory
California Institute of Technology
Pasadena CA 91109

Outline

- JPL Enose program overview
- Data analysis task Overview
- Data analysis approach
- Data Preprocessing
  - Baseline drift estimation
  - Humidity subtraction
- Sensor calibration and selection
- Functional group identification
- Model building

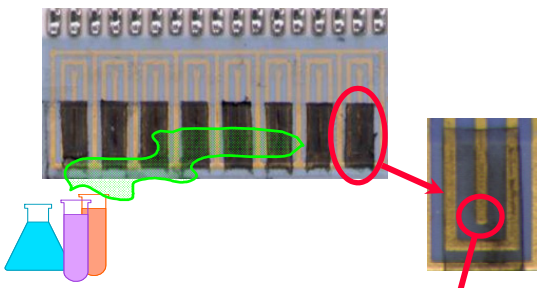
JPL Enose Program Overview

WHAT IS AN ELECTRONIC NOSE?

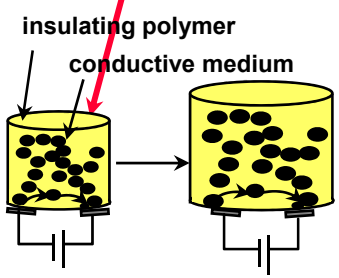
An array of non-specific, conductometric chemical sensors which mimics the mammalian nose by recognizing patterns of response.

1. Measure baseline resistance in each polymer-carbon composite sensor to establish R_0.

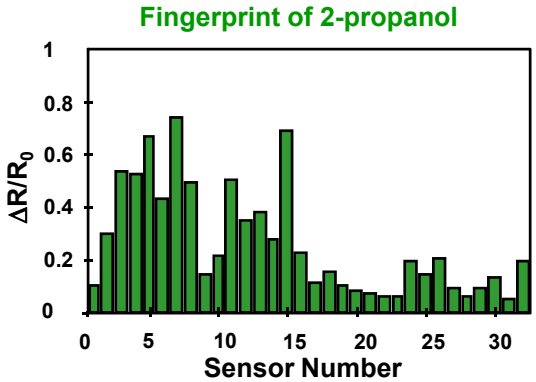
2. Compound from a leak or a spill comes into contact with sensors on the sensing head.



3. The sensors change resistance ΔR in response to change in composition of the air



4. Resistance is recorded, $\Delta R/R_0$ is computed, and the distributed response pattern generated.



5. Responses of the sensor array are analyzed and quantified using software developed for the task.

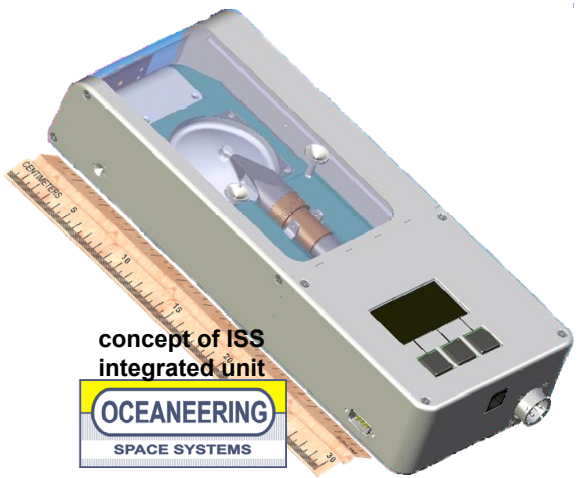
2-propanol - 300 ppm

WHY USE AN ENose for Cabin Air Quality Monitoring?

- **Fills gap between an alarm (no id or quantification) and complex analytical instruments**
  - ID and quantification of trained-for chemical species
  - Wide dynamic range: fractional ppm to 10,000 ppm
  - Array based sensing; can be trained to detect new species and training data uplinked
  - Runs continuously (30 to 360 data points/hr) and autonomously
- **Simple design, robust, rugged, microgravity-insensitive**
- **Minimal crew interaction required**
- **Requires no consumables**
- **Low mass, volume, power**
- **Readily integrated with larger devices and with monitoring/control systems**
- **Can analyze volatile aerosols as well as vapors**

JPL Enose Program Overview

THREE GENERATIONS OF ENoses:



Generation 1
Experiment on STS-95, TRL 6-7
funded by AEMC

Volume: 2000 cm^3 inc. computer

Mass: 1.4 kg including computer

Power: 1.5 W ave., 3 W peak

Detect/ID/Quant 10 compounds at 1 hour SMAC. No real-time data analysis; data acquisition and device control with HP 200LX computer.

6 day flight experiment successful.

Generation 2
Ground Testing, TRL 5-6
funded by AEMC

Volume: 750 cm^3 w/o computer

Mass: 800 g w/o computer

Power: 1.5 W ave., 3 W peak

Detect/ID/Quant 21 compounds at 24 hour SMAC. Data acquisition and device control possible with PDA computer; real time data analysis with ultra micro computer.

Extensive ground testing in environmental chamber.

Generation 3
Tech. Demo. on ISS, TRL 4
funded by AEMC

Volume: 3-4 L inc. computer

Mass: 3-4 kg inc. computer

Power: 5-8 W ave., ~ 14 W peak

Detect/ID/Quant 10 compounds at defined concentrations, including Hg, SO_2. Deconvolute mixtures, id unknowns by functional group.

Data acquisition, device control, real time data analysis, display included.

Extensive ground testing in environmental chamber; then six month test on-orbit.

JPL Enose Program Overview

CABIN AIR QUALITY MONITORING

FUNCTIONS

- **Incident monitor for targeted contaminants exceeding targeted concentrations. Identify and quantify.**
- **Monitor for presence of compounds associated with fires or overheating electronics**
- **Monitor clean-up process**

CHARACTERISTICS

- **Low mass, low power device**
- **Requires little crew time for maintenance and calibration**
- **Detects, identifies and quantifies selected chemical species at or below 24 hour SMAC**

Enose Data Analysis - Overview

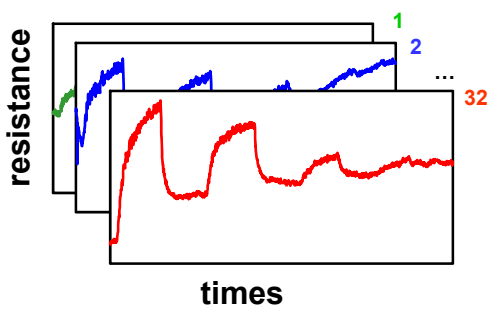
- ***Challenges:***
  - Quasi real time, autonomous detection and analysis (Id &Qn)
  - Deconvolution of mixtures of up to three
  - Functional group identification

  - Poor consistency in sensor responses
  - Mixed linear and nonlinear sensor responses
  - Similar (correlated) response patterns among target analytes
  - Large dynamic response range (magnitude difference in the order of 4)
  - Baseline drift

  - Only partially controlled environment: humidity, temperature, pressure

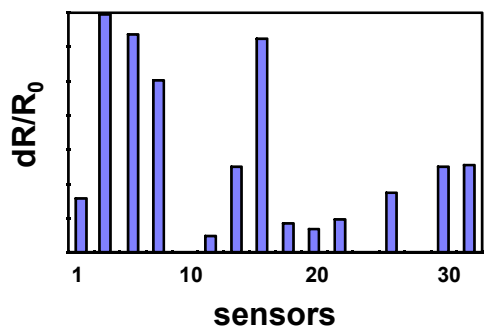
  …

Multi-channel time-series resistance data



==>

Extracted response pattern at a given time



Software anslysis result

==> **35 ppm toluene+ 50 ppm benzene**

Enose Data Analysis - Overview

Development History -1st Gen

- Investigated /compared various approaches
  - PCA, DFA, NNBP, LA, GA, LM NLS
  - Off-line data analysis (ID and Qn only, no event detection)

- Selected LM –NLS (Levenberg Marquart Nonlinear Least Squares) method
  - Adaptive to both linear and nonlinear sensor responses
  - Recognition on untrained mixtures (of trained single analytes)
    - Significantly less training time, less training data needed
  - Robust to noisy data

- Verification w/ 6-day STS-95 flight experiment data (1998)

Enose Data Analysis - Overview

Development History -2nd Gen

- LM-NLS:
  - Expanded analyte list (from 10 to 22)
  - Developed quasi real time data analysis
  - Investigated humidity subtraction technique to uncover small concentrations of analyte in presence of humidity change
  - Investigated temperature effect
  - Investigated func'l group identification w/ sub sensor array

- Extensive Ground Testing

Enose Data Analysis - Overview

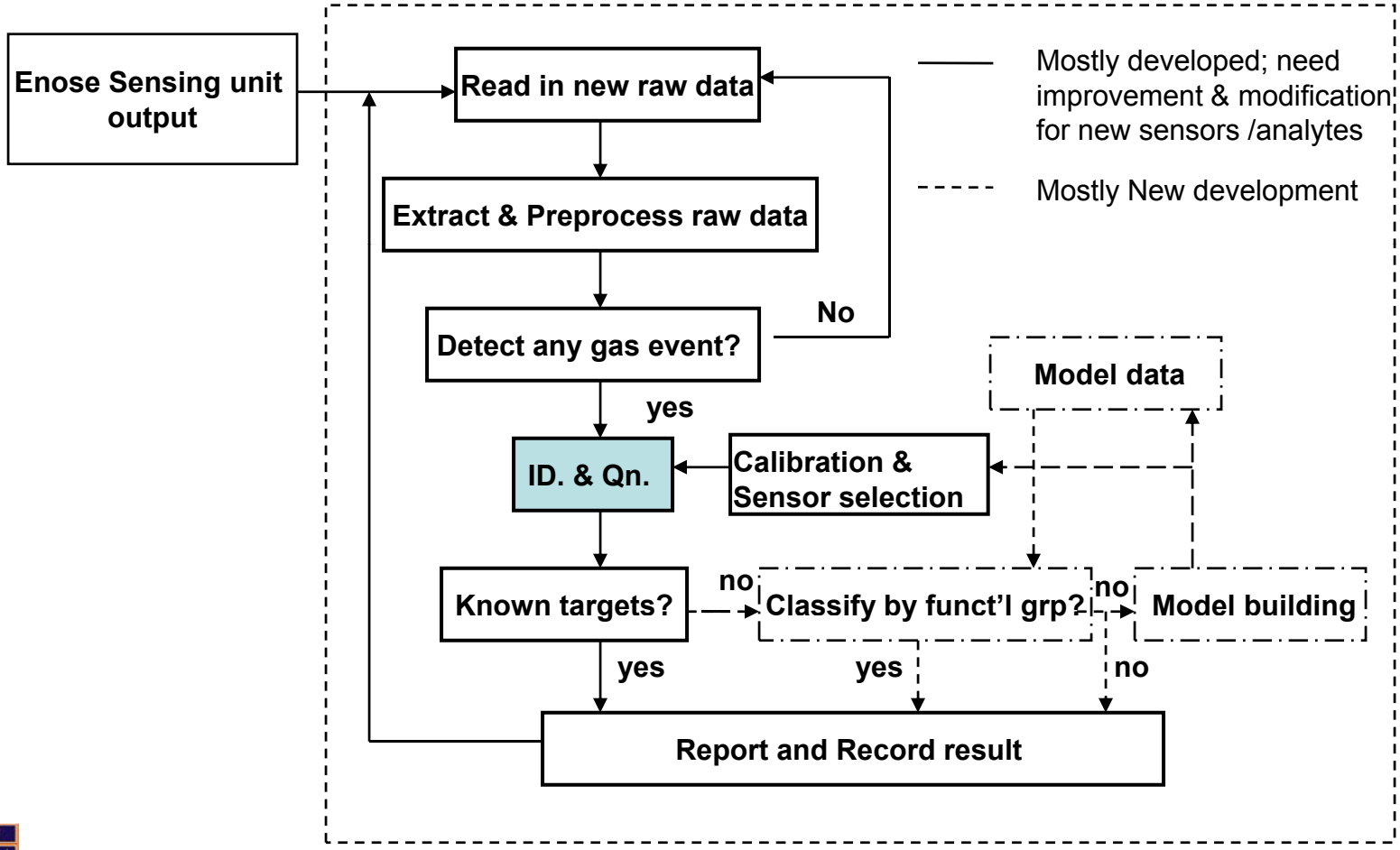
Development Goals/Requirements -3rd Gen

- Third Generation:
  - Quasi real time event detection, identification and quantification
  - Ability to uncover small analyte signals under large humidity
  - Ability to classify unknowns by functional group (alcohols or aromatics)
  - Ability to identify compounds not trained for as “unknown”
  - Improvement on data analysis speed, memory efficiency and performance for embedded implementation

- Will undergo extended (6-month) testing on ISS (2008)

Data Analysis - Approach

Flow Chart (current)



Data Analysis - Approach

Approaches Investigated

- Discriminant Function Analysis (DFA):
  - Attempt to find hyperplanes dividing the groups.
  - Not suitable when groups are not separable (by a hyperplane); not well suited for mixtures.
- Back Propagation Neural Networks (BPNN):
  - Try to find a best-fit function (linear or nonlinear; no models needed) for the training data.
  - Good for generalization of functions to cases outside the training set.
  - Not suitable when data groups overlap;
  - Normally requires “clean” training data; not well suited for mixtures
- Linear Algebra (LA):
  - Try to find the best-fit (in a least square sense) parameters vector x from an observation vector y by pseudo-inverse y=Ax. A: system characteristics matrix from training data.
  - Suitable only for linear model.
- Differential Genetic Algorithm (DGA) :
  - Imitates principles of genetics & natural evolution: e.g., recombination, inversion, mutation and selection. Various paths to the optimum are checked; info about them exchanged.
  - Simple concept, easy to use, fast convergence
  - Limited control ability

Data analysis Approach

Approaches Investigated

- Partial least squares (PLS) †
  - Multivariate signal averaging; inverse linear regression
  - Non-iterative, rapid calculation
  - Can be robust to environmental variations with properly designed calibration
- Support Vector Machine (SVM)*
  - Decomposing N-classes classification into multiple 2-classes tasks
  - Need training data of every possible combination of gas compounds
  - Good result on trained data, but poor on untrained ones (poor generalization)
- Maximum likelihood*
  - Use covariance of error matrix to construct likelihood function
  - Two step (classification and quantification) procedure
- Probabilistic Neural Network*
  - Soft decision (PDF) based classification
  - Much less successful than SVM

* Independent work by I-A-I, Inc., on preprocessed JPL 1st Generation ENose data

† Still under investigation

Data Analysis - Approach

Approach Selected

- LM-NLS
  - Global & local; classification & quantification - simple code and tree structure
  - Recognition on untrained mixtures (of trained single analytes)
    - Significantly less training data needed (than many other approaches)
  - Robust to noisy data
  - Adaptive to both linear and nonlinear sensor responses :
    - increase damping (reduce step) for a highly nonlinear problem
    - decrease damping (increase step) for a linear problem
  - Good generalization (on future data)

Data Analysis - Approach

Approach - LM NLS Method

- Objective
  - Find best-fit concentration vector ***c*** from an observed resistance change vector ***dr/r*,** which is related to ***c*** through a known function, ***dr/r***=f(A, ***c***). i.e., ***dr/r***=A_1***c***+A_2***c***$^{2.}$

    A_1 and A_2 : sensor coefficient matrix pre-established from training data.

- Procedure - Iterative optimization:
  - Establish A_1 and A_2 from training data (curve fitting)
  - From a given starting point of ***c***, calculate discrepancy of the fit:

    *residual=wt *(computed f(A,**c**)-observed **dr/r**)*

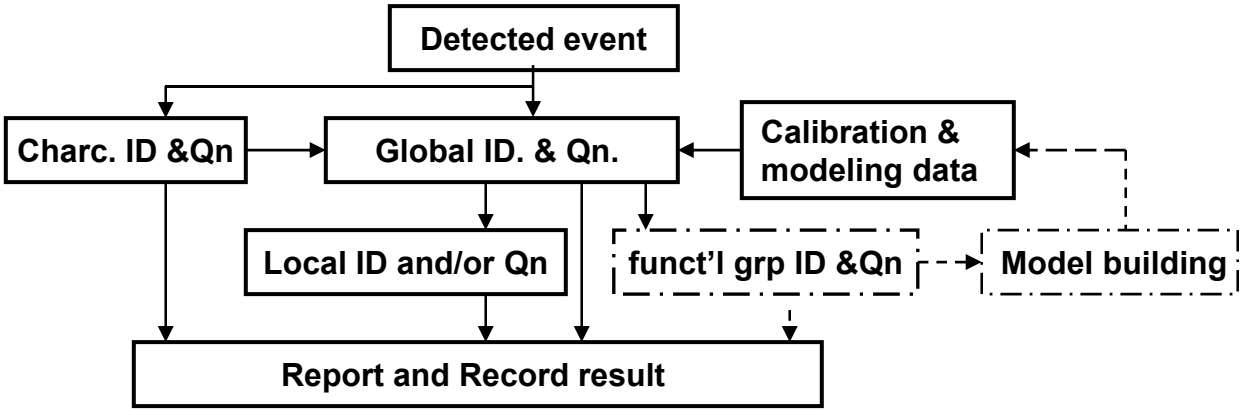
    update ***c*** with a better-fitted (smaller residual) concentration ***c*** at each step till max iteration number reached.

  - Repeat for multiple starting point of ***c***, and update better-fitted *c* that favors less number of significant elements

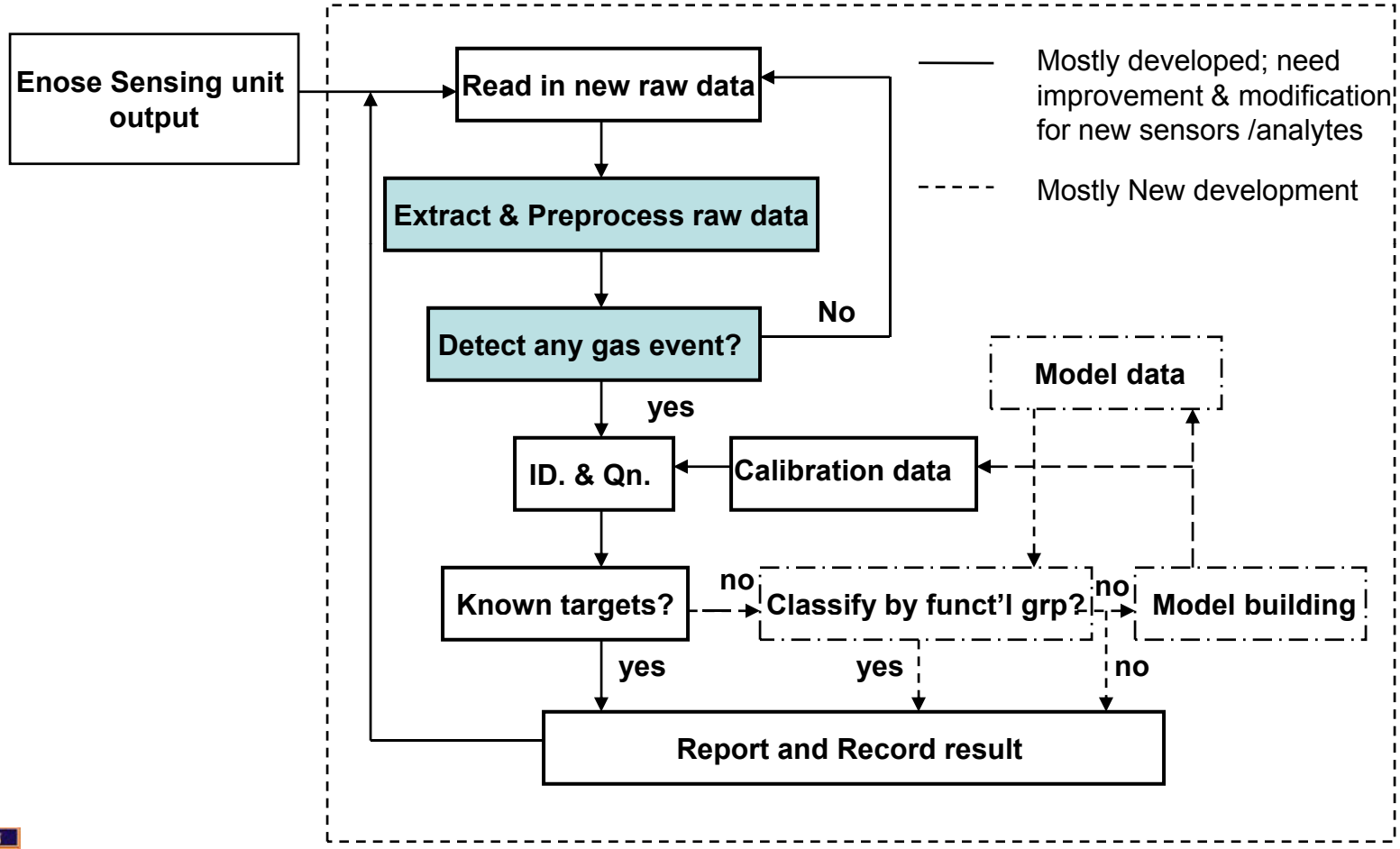
Data Analysis - Approach

Flow chart

- ID and Qn based on characteristic response if any
- Global (all analytes) ID and Qn
  - For most gases, one step global ID & Qn w/ LM-NLS
- Second step local (selective) ID and/or Qn
  - For easily confused gases, after the first step
  - For analytes w/ only few reliable responsive sensors
- Functional group check
  - When residual is too large to be reliable for first step result
- ID of unknowns based on model building



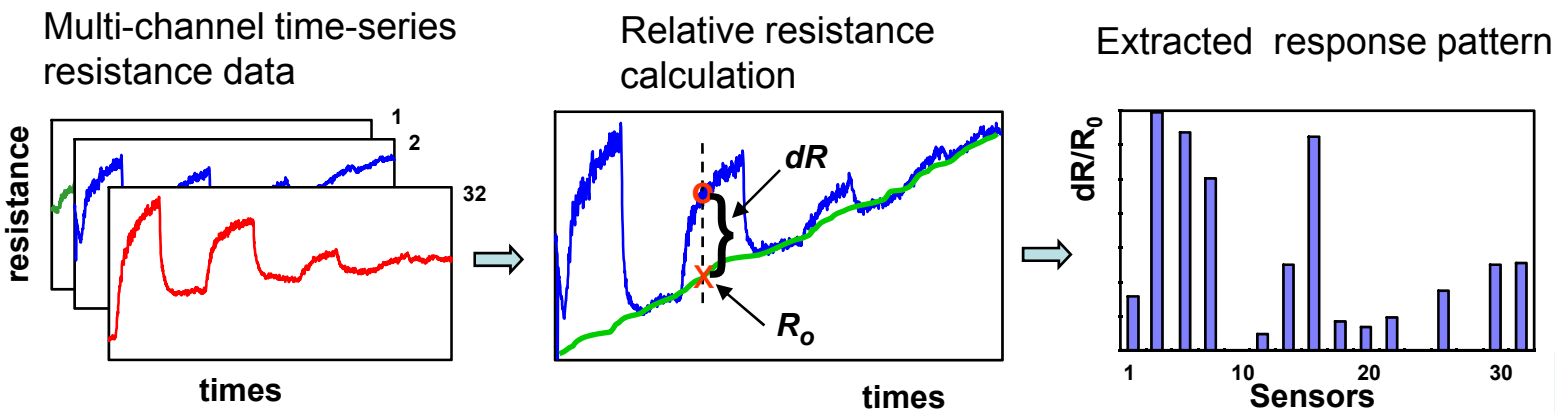
Data Preprocessing



Data Preprocessing

Converting time-series R to resistance response pattern dR/R_o

- General Noise removal
  - Appropriate filtering
- Baseline drift estimation
  - Multiple causes, e.g., changes in temperature, humidity, flow pressure, etc.
  - Need fast and accurate estimation for real time signal detection
- Calculation of resistance change
  - Dynamic relative resistance change: *dR/Ro*
- Event detection



Preprocessing –Baseline Estimation

- Baseline drift estimation
  - Investigate temperature and/or humidity caused baseline drift estimation
  - Use appropriate digital filtering (w/ known events if any) to estimate slowly varying baseline drift
  - Additionally,
    - Use clean air reference cycle info to establish piecewise baseline
    - Correct abnormal baseline change due to sudden operating temperature change

Preprocessing –Baseline Estimation

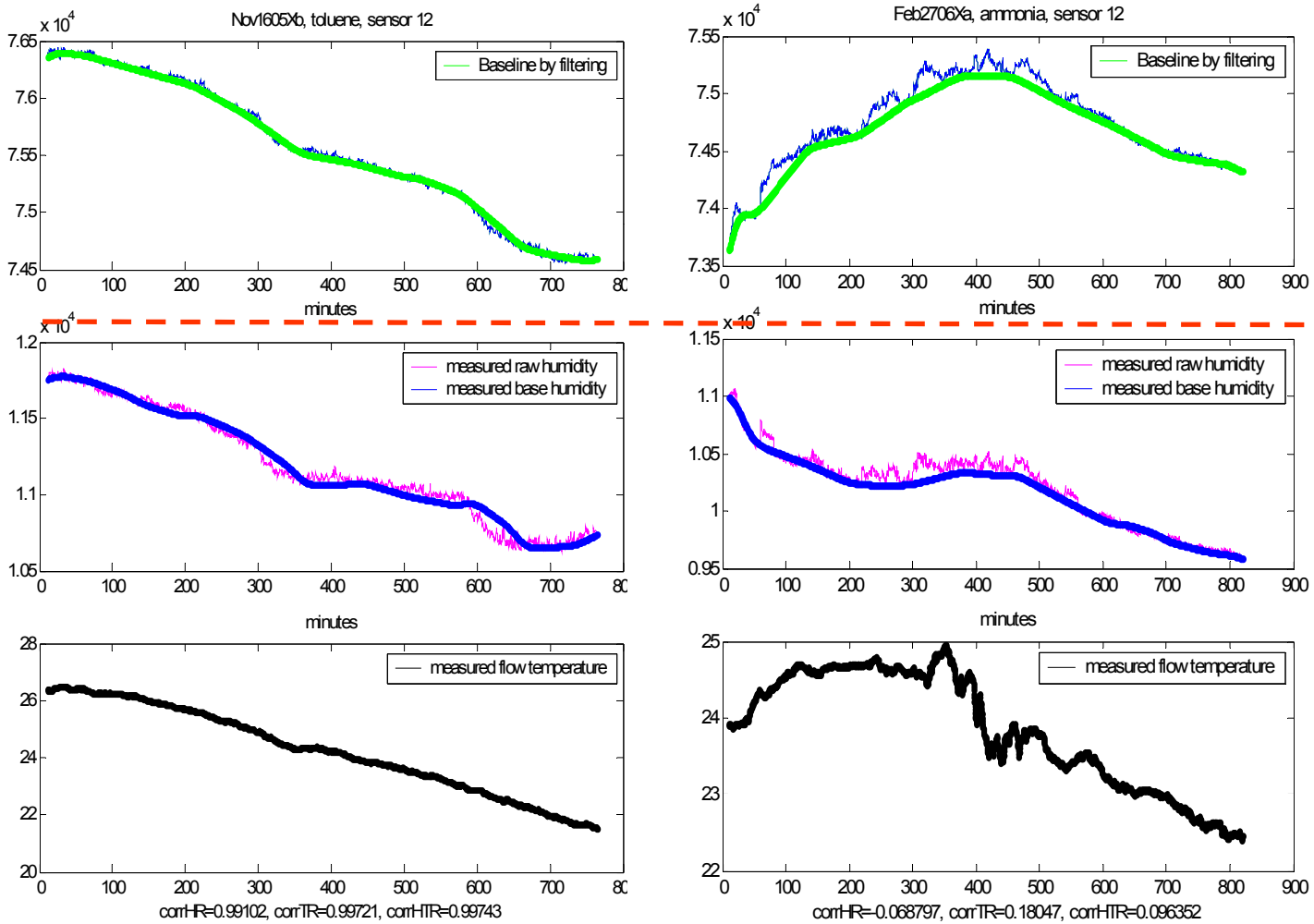
Investigation of baseline estimation using temperature & humidity info:

- Correlation (bwtn humidity or temperature or their combination and the actual baseline drift) exists but…
  - The degree of correlation varies greatly among different sensors
  - The degree of correlation also varies greatly among different files (for same or different sensors/ analytes)

- Temperature & humidity info (or their combination) not reliable enough to establish actual baseline drift

Preprocessing –Baseline Estimation

- Correlation exists but… varies from time to time
  - Same sensor (sensor #12), different analytes (ammonia, toluene)



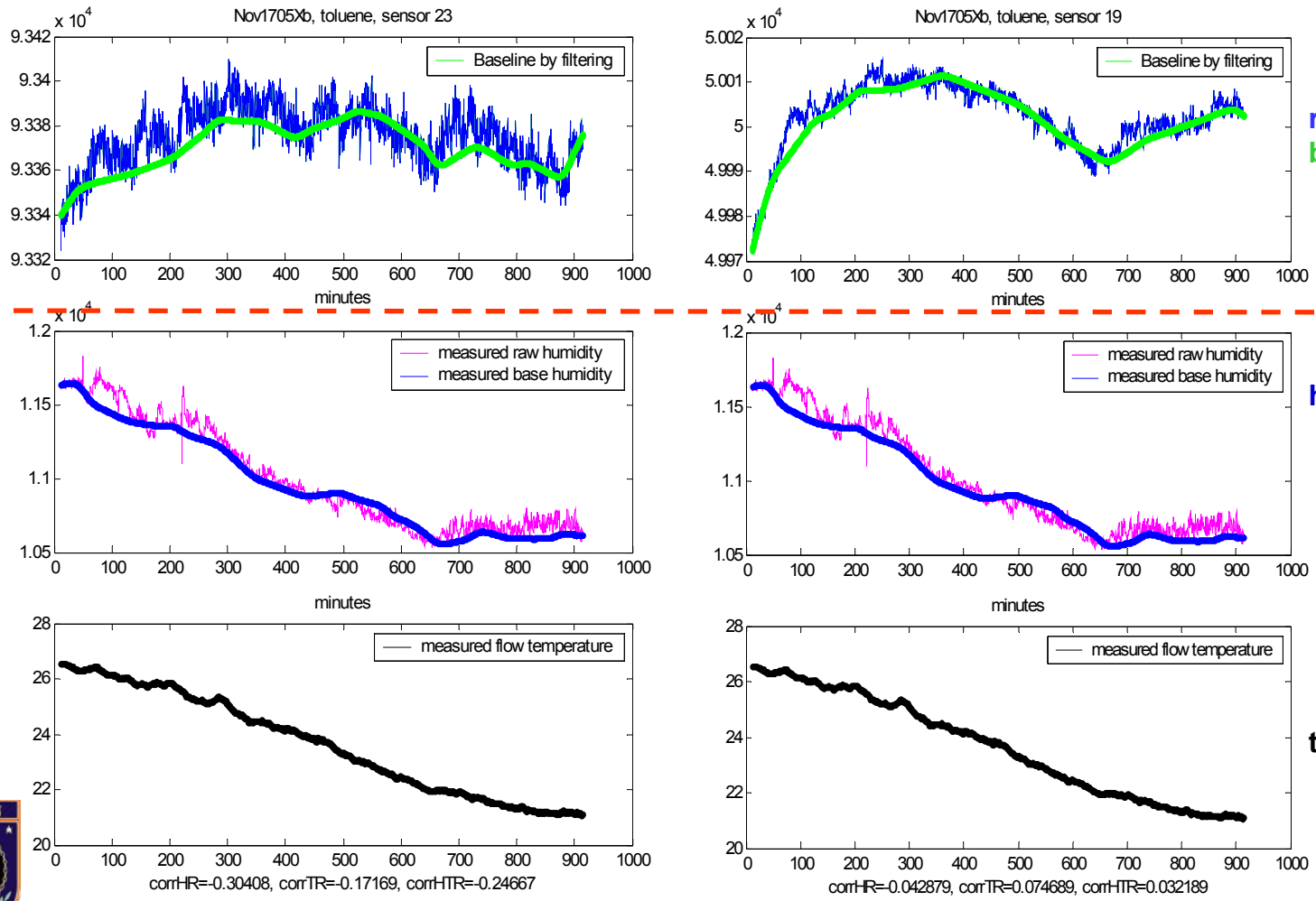
raw resistance baseline (green)

humidity (blue)

temperature

Preprocessing –Baseline Estimation

- Also varies among different sensors
  - Different sensors (# 23, # 19), same analyte, same data files



raw resistance
baseline (green)

humidity (blue)

temperature

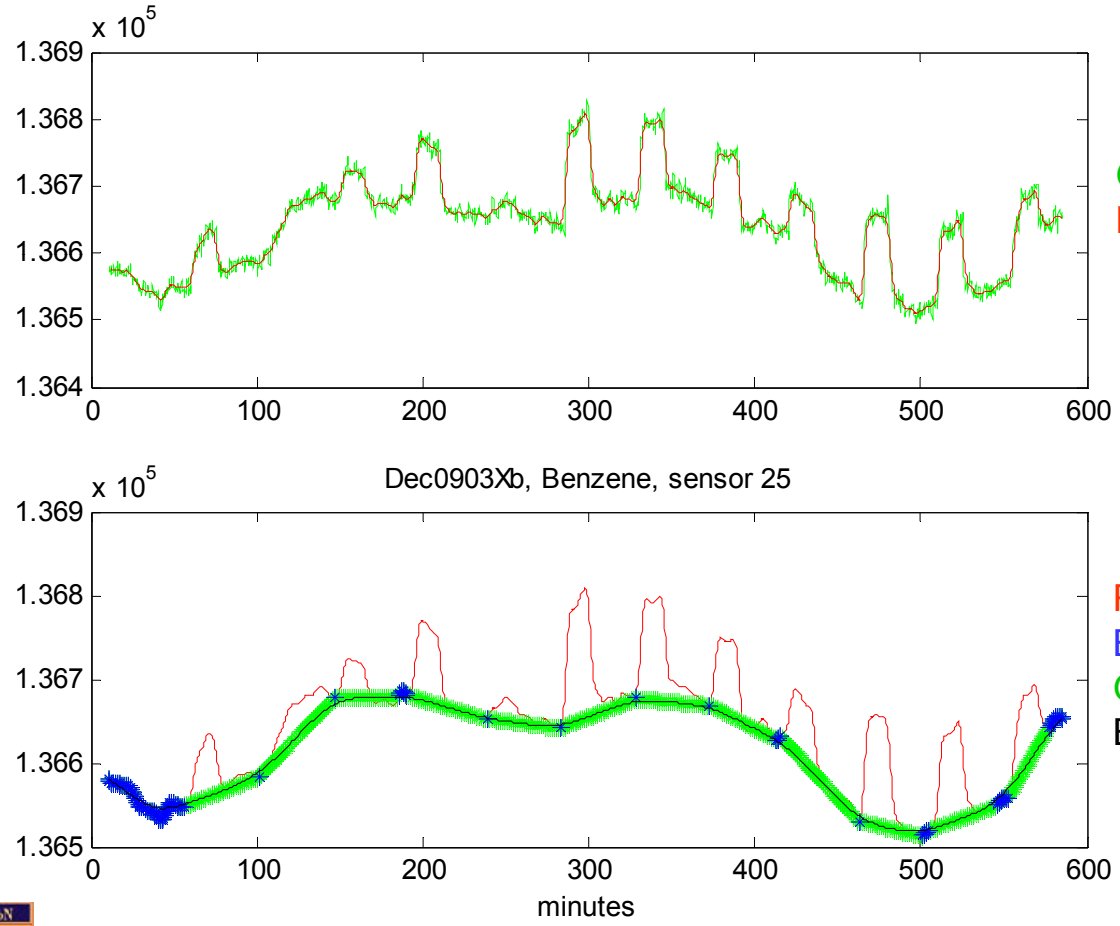
Preprocessing –Baseline Estimation

- Digital filtering w/ automatic peak rejection
  - Use known (calibration mode) or detected (real time mode) event info to identify peak regions
    - Blank out peak/event regions
    - Retain non-peak regions as baseline points (A)
    - Fill in peak regions with piece-wise Linear fit (B)
  - Estimate baseline by filtering points made of (A+B)

Preprocessing –Baseline Estimation

Jet Propulsion Laboratory
California Institute of Technology

Baseline estimation -- calibration mode



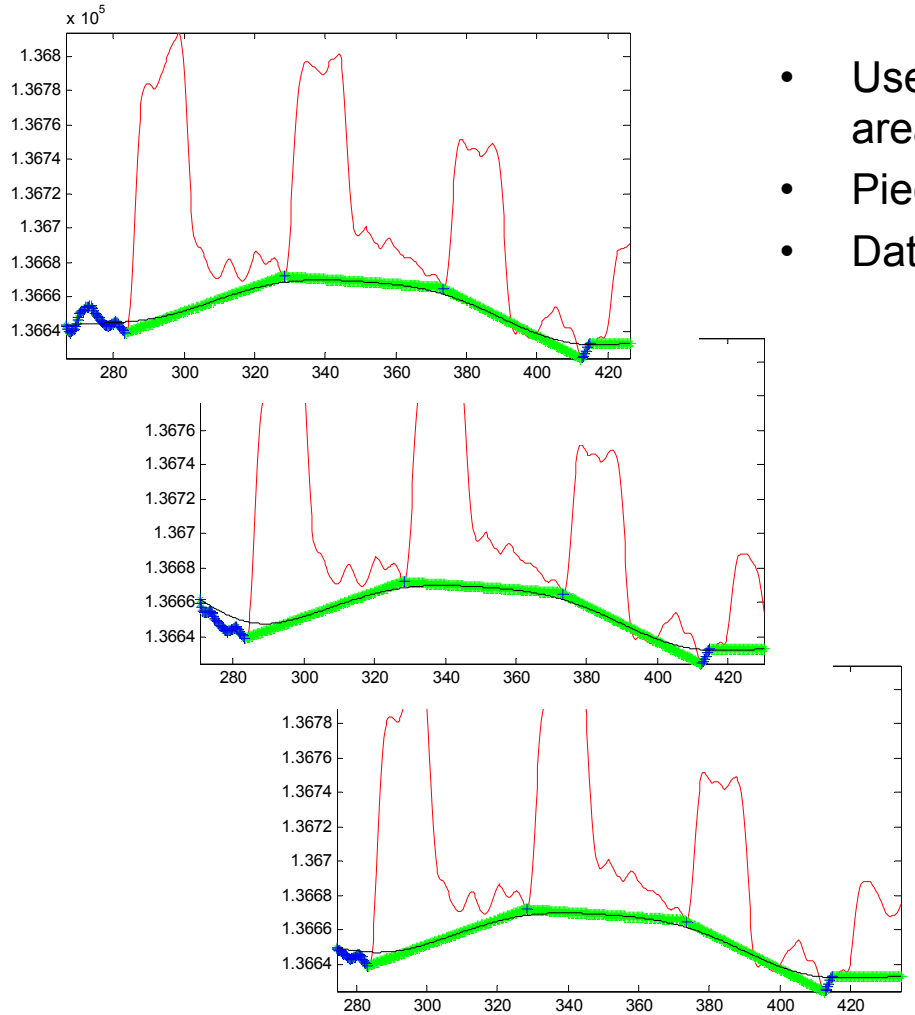
Green – raw resistance data R
Red – smoothed R

Red -- – smoothed R
Blue + – points retained from R
Green – points filled in for peak areas
Black – baseline estimation by filtering

Preprocessing –Baseline Estimation

Baseline estimation -- Real time mode

…



Dec0903Xb, Benzene, sensor 25

minutes

- Use detected events info to block peak areas (recent past events)
- Piece-wise Linear fit for all peak region
- Data of 2 ~4hrs retained

Red - – smoothed R
Blue + – points retained from R
Green – points filled in for peak areas
Black – baseline estimation by filtering

…

Preprocessing – Event Detection

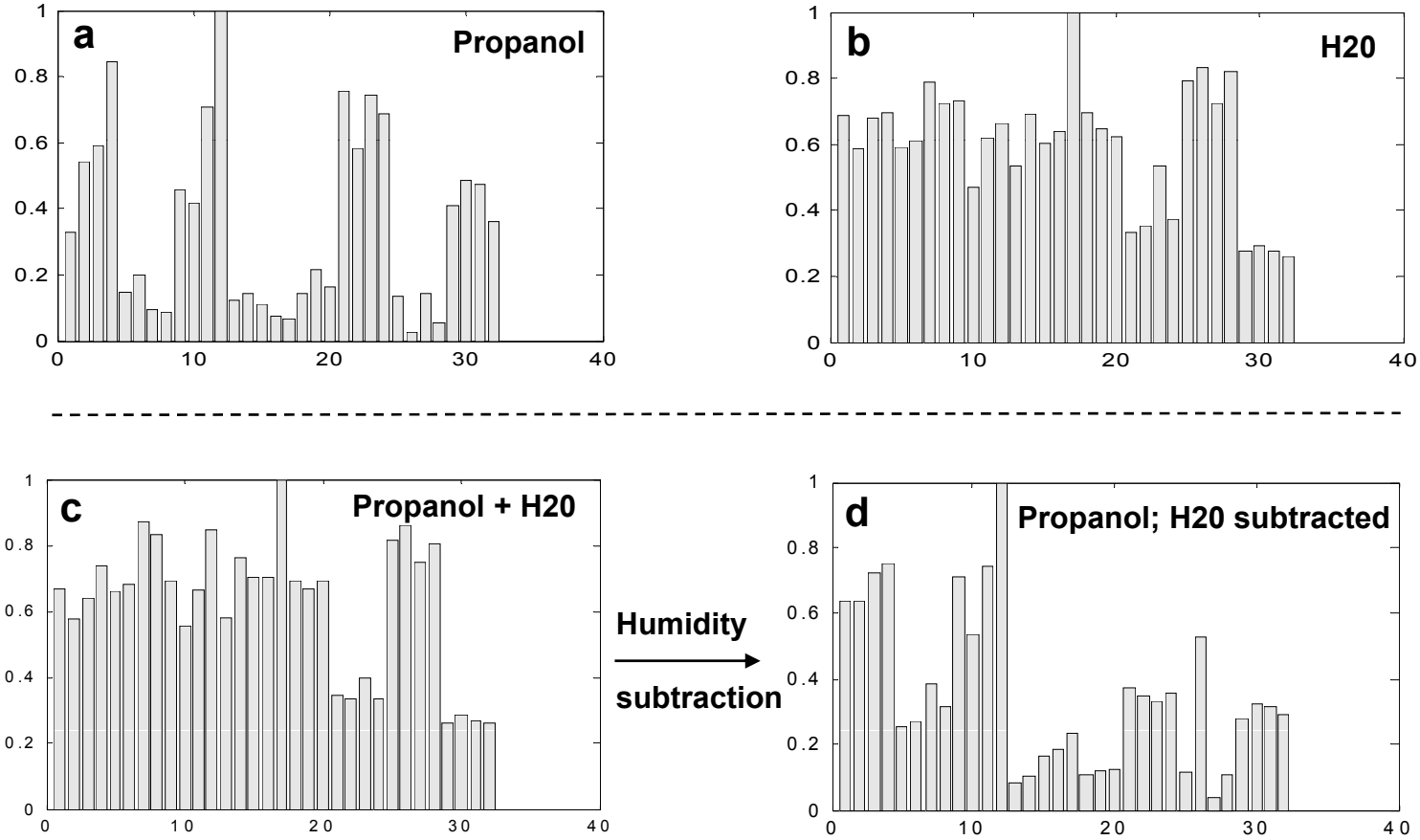
- Event detection & Pattern Extraction
  - Threshold test of sensors response strength
    - Use selected sensors based on sensor's limit-of-detection (LOD)

      LOD= 3*σ/ sensor sensitivity

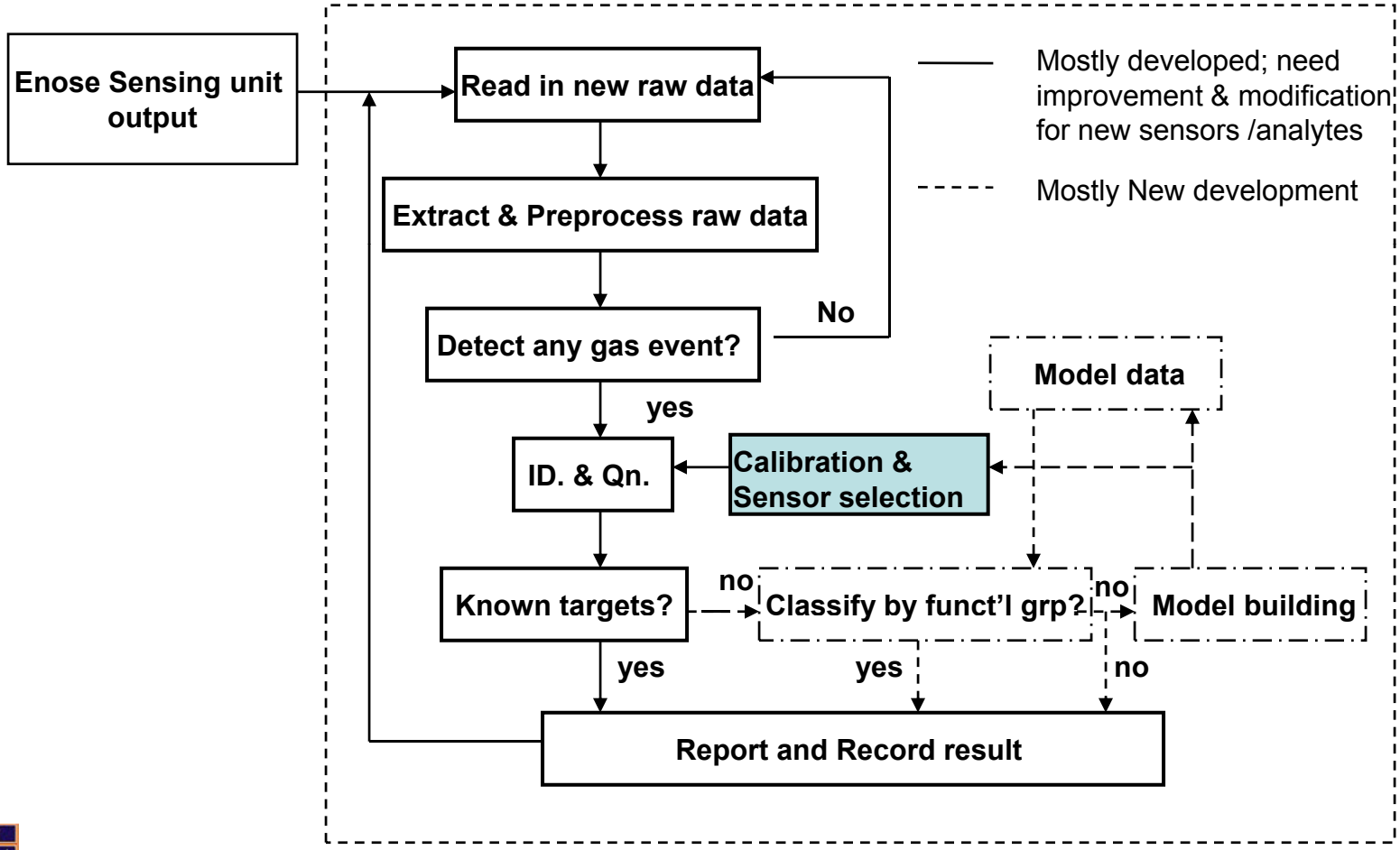
      σ – standard deviation of a sensor's response to an analyte
  - Humidity subtraction
    - Humidity can easily overwhelm weak signals
      - 10^4 vs single digit in response magnitude
    - Improve pattern extraction by subtracting large humidity response using measured humidity info

Preprocessing – Humidity Subtraction

- Subtracting humidity response using measured humidity info
- Improved detection/identification of weak analyte signal accompanied with large humidity changes by about 50% in a test done with 30 propanol events



Sensor Calibration & Selection

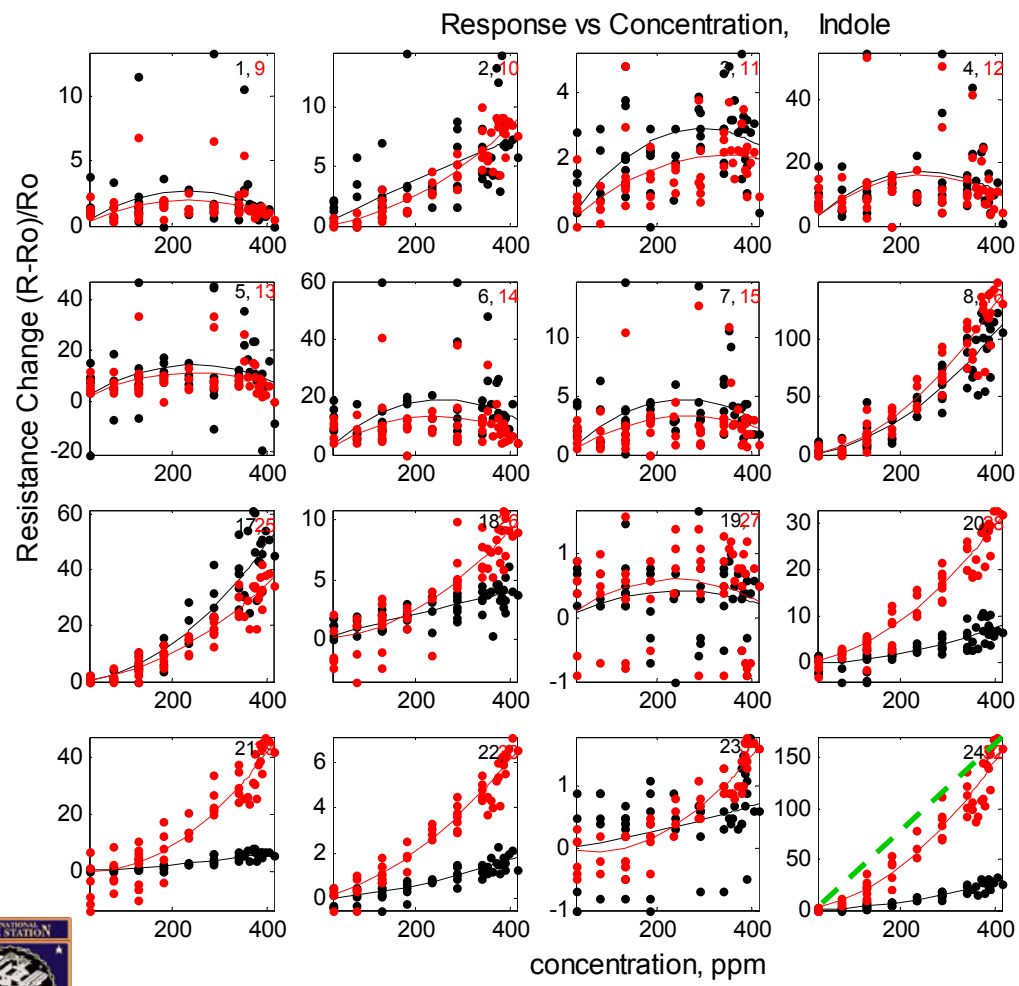


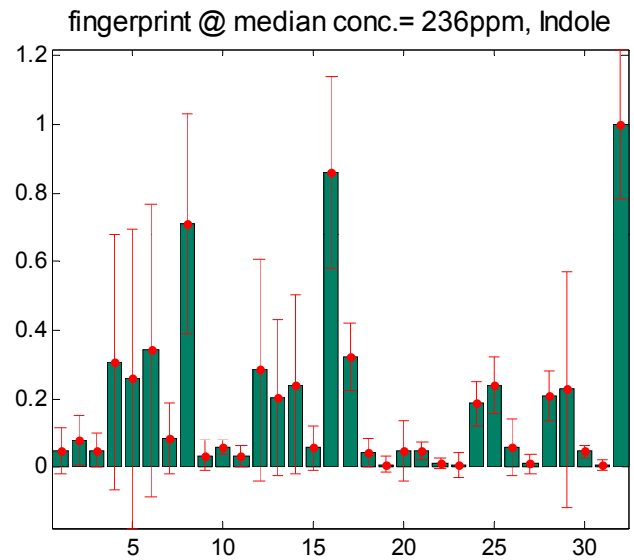
Sensor Calibration - Singles

- Single gases
  - Outliers removal
  - Model fitting and coefficients calculation for each analyte, each sensor
    - e.g., Least square fit of 2nd order polynomial, passing (0,0)
  - Statistics:
    - Reliability (inconsistency): variation between actual response vs. fitted response (scattering within an analyte)
    - Selectivity: diversity to different analytes
    - Sensitivity: responsiveness to different analytes
    - Correlation: btwn concentration and response strength
    - (Euclidean) Fingerprint distances between analyte pair
  - Calibration for different environment conditions
    - Base humidity, temperature, pressure
    - Experimental design?

Sensor Calibration - Singles

Resistance change v.s. Concentration, **Indole;** 2nd order polynomial fit : $dr/r = A_1 c + A_2 c^2$





• , • Experimental data

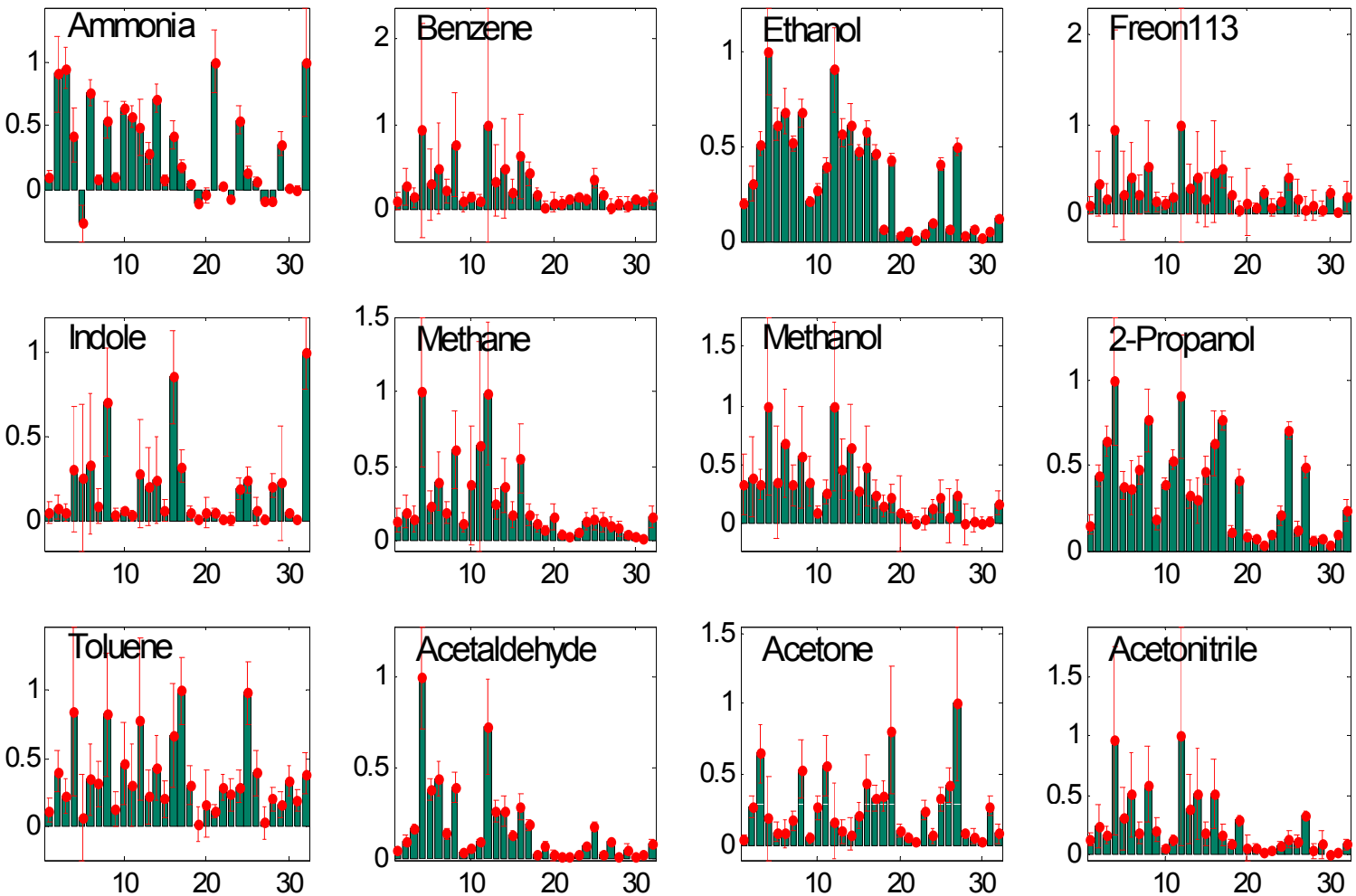
Polynomial fit

If linear fit

Sensor Calibration - Singles

Jet Propulsion Laboratory
California Institute of Technology

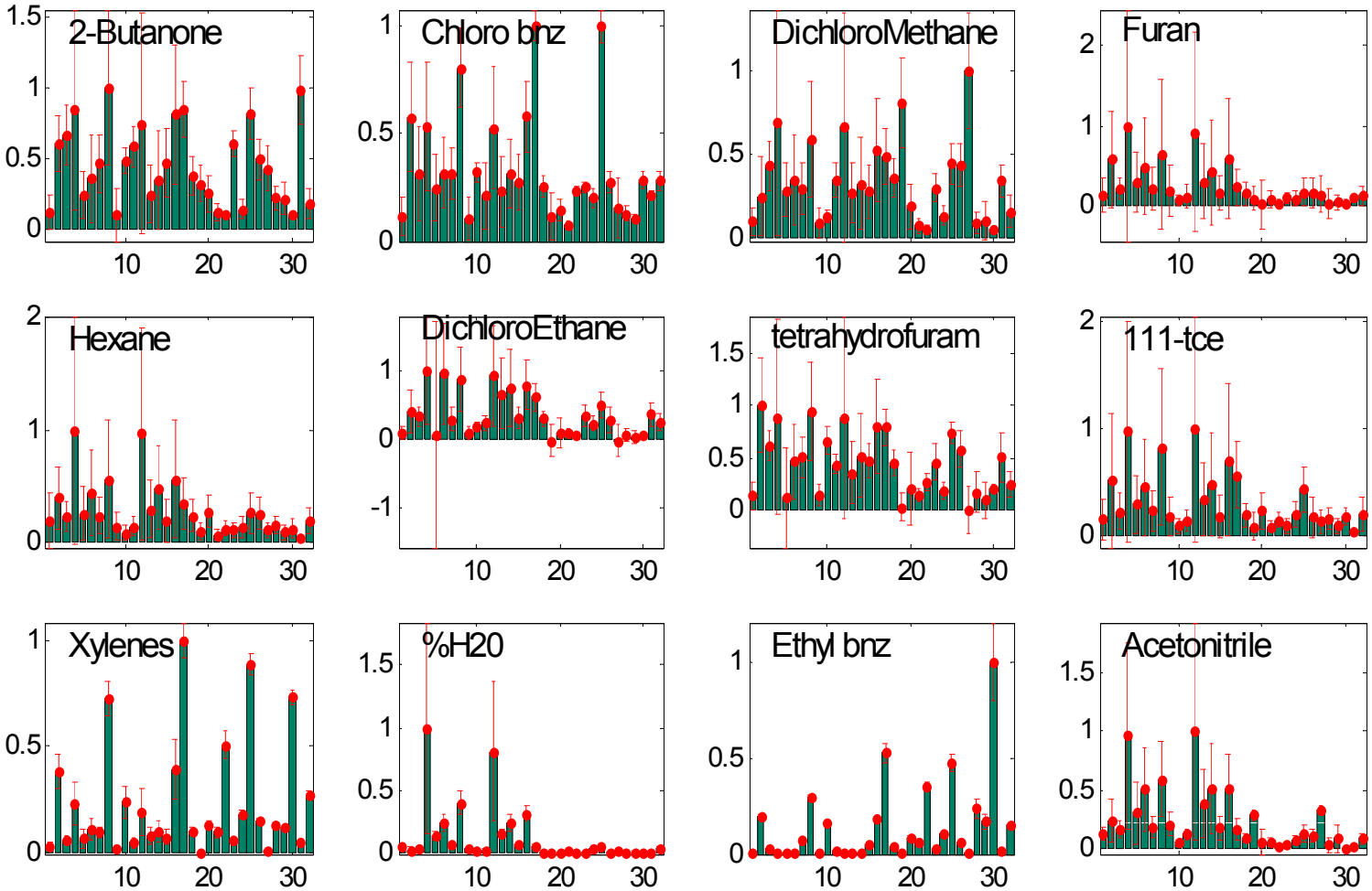
Fingerprints of 2nd Gen Analytes



Sensor Calibration - Singles

Jet Propulsion Laboratory
California Institute of Technology

Fingerprints of 2nd Gen Analytes



Sensor Calibration - Singles

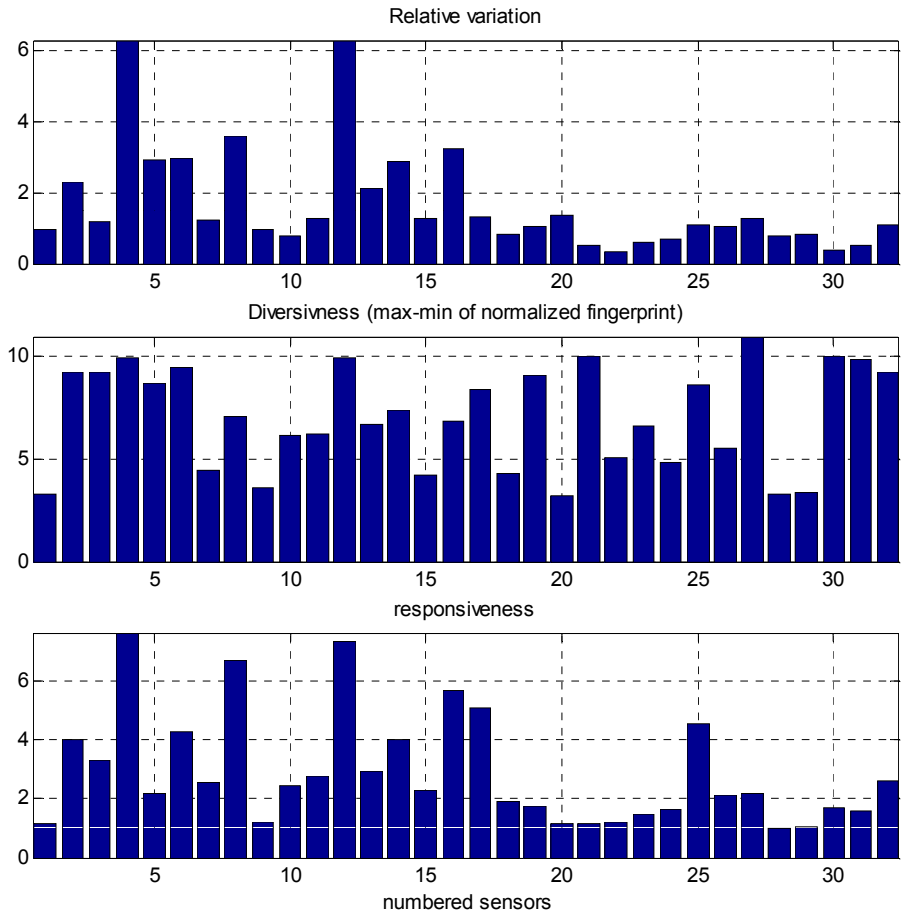
Variation, Diversity and Responsiveness of Sensors in Array

Averaged over all analytes at target concentration

Reliability:
Measure of scatter
Variation < 1 preferred

Selectivity (Diversity)
(max-min of norm. fingerprint)
Higher number preferred

Responsiveness (Sensitivity)
Relative response strength
Higher number preferred



Numbered Sensors

Sensor Calibration - Singles

	Wthn-	Cross- Analyte fingerprint distance																						
analytes #:		1	2	3	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	21	22	23	24	25
Ammonia(1)	11																							
Benzene (2)	34	31																						
Ethanol(3)	4	28	15																					
Freon113(5)	34	31	7	17																				
Indole(6)	19	29	23	31	25																			
Methane(7)	18	27	11	16	11	25																		
Methanol(8)	25	28	13	10	13	29	13																	
2-Propanol(9)	7	29	15	9	16	29	15	15																
Toluene(10)	20	29	16	22	14	27	19	22	15															
Acetaldehyde(11)	8	35	17	22	17	26	15	18	24	29														
Acetone(12)	14	35	28	25	28	35	27	29	23	28	36													
Acetonitrile(13)	21	31	10	13	12	27	10	9	16	23	14	29												
2-Butanone(14)	16	30	21	19	22	32	22	23	11	13	32	23	25											
Chloro bnz(15)	10	30	17	22	15	26	21	22	14	8	29	25	24	14										
DichloroMethane(16)	15	33	17	16	17	30	19	19	15	18	26	12	19	15	16									
Furan(17)	42	29	9	15	11	26	10	11	16	20	13	31	7	24	20	20								
Hexane(18)	27	30	7	16	7	24	10	10	16	16	17	28	10	20	16	16	8							
DichloroEthane(19)	28	28	13	15	15	29	19	15	16	14	27	28	20	17	16	17	17	15						
tetrahydrofuram(21)	20	27	18	19	19	32	22	23	13	11	32	28	25	9	14	19	22	18	13					
111-tce(22)	31	30	7	15	7	24	13	13	13	13	20	30	12	18	13	17	9	6	12	15				
Xylenes(23)	6	38	26	34	24	27	32	35	27	18	37	31	35	29	17	31	32	28	29	27	25			
%H20(24)	15	39	23	30	23	30	21	26	31	33	12	42	21	38	35	33	20	24	32	37	27	40		
Ethyl bnz(25)	4	47	37	43	33	36	42	44	41	35	44	40	44	42	35	40	44	38	41	41	36	21	49	
analytes #:		1	2	3	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	21	22	23	24	25

Avg distance: 19 | **23** **min btwn-class distance: 6**

Sensor Calibration - Mixtures

- Mixture gases
  - Additive linearity (superposition) validation

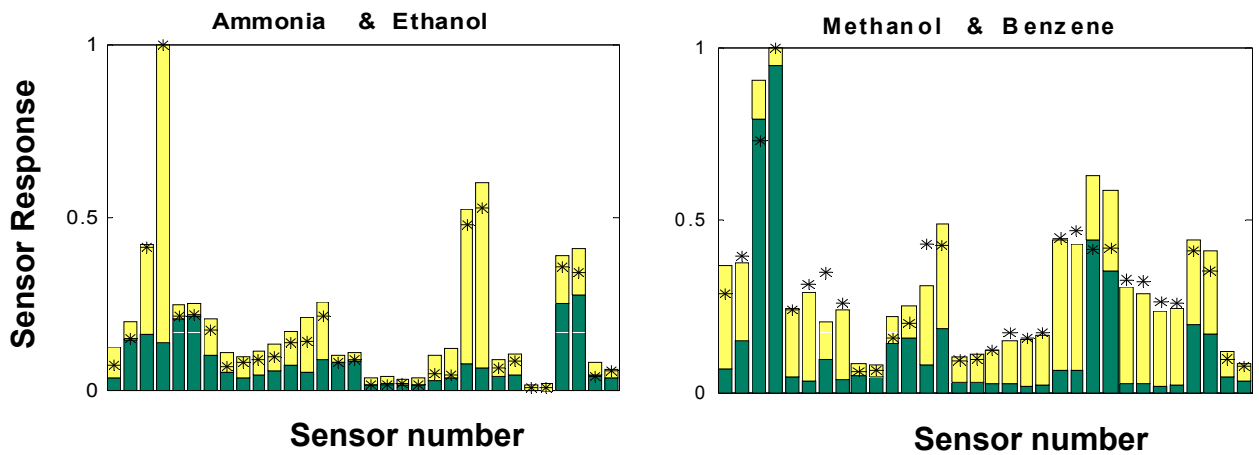
    polynomial *with no cross interaction effect*: $dr/r = A_{1a}c_a + A_{2a}c_a^2 + A_{1b}c_b + A_{2b}c_b^2$

  - Alternative mixture model if necessary:

    e.g., polynomial *with 2-way cross interaction effect*: $dr/r = A_{1a}c_a + A_{2a}c_a^2 + A_{1b}c_b + A_{2b}c_b^2 + A_{ab}c_ac_b$

  - Dominant effect of one component?
    - Subtraction technique

Additive linearity holds reasonable well in a previous study



Sensor Calibration - Other Conditions

- Calibration for different environment conditions:

  Nominal ISS Atmosphere conditions:
  - Relative Humidity: 25% -75%
  - Temperature: 20°C -34°C
  - Partial Pressure: ~15mmHg

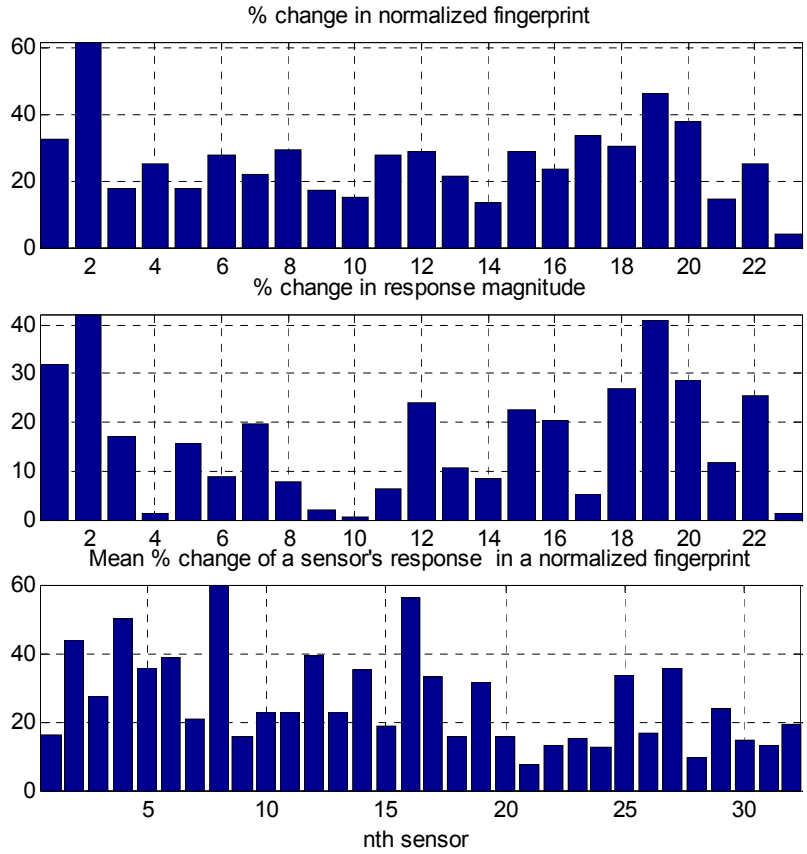
- Investigation of sensor calibration needs at different RH conditions
  - Difference of fitted normalized fingerprints at 5k ppm, 18k ppm, or 23k ppm, against 10k ppm with 2nd gen data

Sensor Calibration - Other Conditions

Changes for diff humidity: 10k vs 5k

Analyte	Fngr chg	Mag chg	Calib?
Ammonia	33	32	x
Benzene	61	11	yes
Ethanol	17	17	no
Freon113	25	1	x
Indole	18	16	no
Methane	28	9	x
Methanol	22	20	no
2-Propanol	30	8	x
Toluene	17	2	no
Acetaldehyde	15	0	no
Acetone	21	1	x
Acetonitrile	29	24	x
2-Butanone	22	11	no
Chloro bnz	14	9	no
DichloroMethane	29	23	x
Furan	23	20	no
Hexane	33	5	x
DichloroEthane	30	27	x
tetrahydrofuram	46	41	x
111-tce	38	29	x
Xylenes	15	12	no
%H20	25	25	x
Ethyl bnz	4	1	no

yes – need recalib
x – borderline
no – no need recalib



Avg=25%

Avg=15%

* **Change in average success rate if use 10k coefficient for 5k data: - 4%**

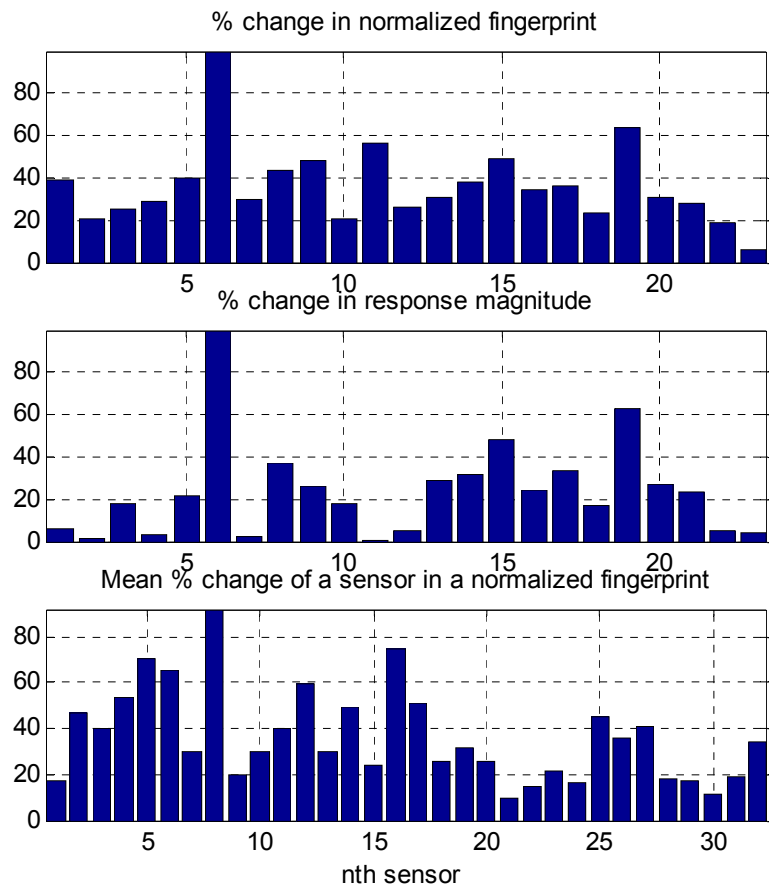
Most...Least changed Sensors:	8	16	4	2	12	6	27	5	14	25	17	18	23	30	22	31	24	28	21
(Mean % change in fngrprnt):	60	55	48	44	39	39	36	35	35	34	33	16	15	15	13	13	12	10	7

Sensor Calibration - Other Conditions

Changes for diff humidity: 10k vs 18k

Analyte	Fngr chg	Mag chg	Calib?
Ammonia	34	1	x
Benzene	20	1	no
Ethanol	25	18	no
Freon113	29	3	x
Indole	40	22	x
Methane	99	99	yes
Methanol	30	3	x
2-Propanol	43	36	x
Toluene	48	26	x
Acetaldehyde	21	18	no
Acetone	27	26	x
Acetonitrile	26	5	x
2-Butanone	31	28	x
Chloro bnz	38	31	x
DichloroMethane	49	47	x
Furan	34	24	x
Hexane	36	34	x
DichloroEthane	23	17	no
tetrahydrofuram	78	40	yes
111-tce	31	27	x
Xylenes	28	23	x
%H20	19	5	no
Ethyl bnz	6	5	no

yes – need recalib
x – borderline
n – no need recalib



Avg=35%

Avg=24%

* **Change in Average success rate if use 10k coefficient for 18k data: -15%**

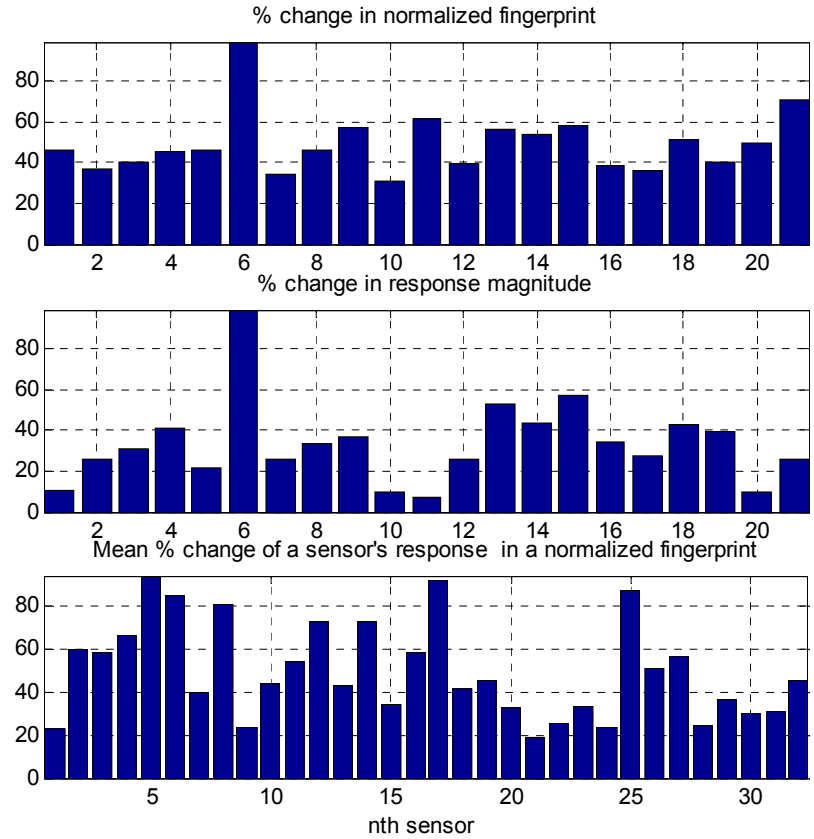
Most...Least changed Sensors:	8	16	5	6	12	4	17	14	2	25	27	31	28	1	29	24	22	30	21
(Mean % change in fngrprnt):	92	74	70	65	59	54	51	49	46	45	40	19	18	17	17	16	15	12	10

Sensor Calibration - Other Conditions

Changes for diff humidity: 10k vs 23k

Analyte	Fngr	Mag	Calib?
Ammonia	46	10	x
Benzene	37	26	x
Ethanol	40	31	x
Freon113	45	41	x
Indole	46	22	x
Methane	100	100	yes
Methanol	34	26	x
2-Propanol	46	34	x
Toluene	58	37	yes
Acetaldehyde	31	9	x
Acetone	61	7	yes
Acetonitrile	40	26	x
2-Butanone	56	53	yes
Chloro bnz	54	44	yes
DichloroMethane	58	58	yes
Hexane	39	34	x
DichloroEthane	36	27	x
tetrahydrofuram	51	43	yes
111-tce	40	40	x
Xylenes	50	9	x
Ethyl bnz	71	26	yes

Yes –need calib
X – borderline
N – no need calib



Avg=50%

Avg=33%

*** Change in average success rate if use 10k coefficient for 23k data: - 20%**

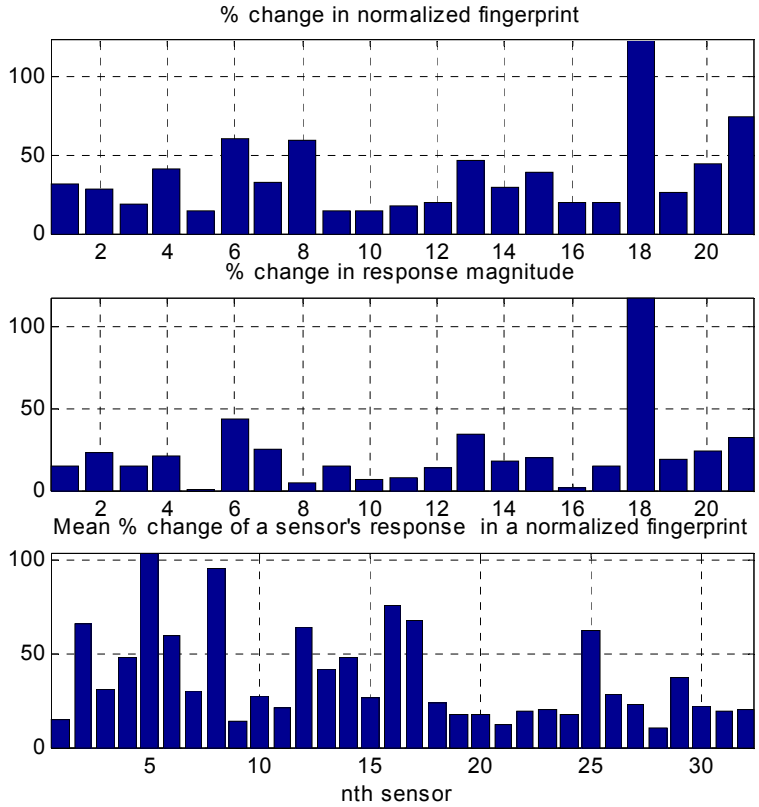
Most...Least changed Sensors:	5	17	25	6	8	12	14	4	2	3	16	31	30	22	28	24	9	1	21
(Mean % change in fngrprnt):	94	92	87	85	80	73	72	67	60	58	58	31	30	25	24	24	24	23	19

Sensor Calibration - Other Conditions

Changes for diff humidity: 18k vs 23k

Analyte	Fngr	Mag	Calib?
Ammonia	31	15	x
Benzene	28	23	x
Ethanol	19	15	no
Freon113	41	21	x
Indole	14	0	no
Methane	61	44	yes
Methanol	33	25	x
2-Propanol	60	5	yes
Toluene	15	15	no
Acetaldehyde	14	7	no
Acetone	17	8	no
Acetonitrile	20	14	no
2-Butanone	47	35	x
Chloro bnz	29	18	x
DichloroMethane	39	20	x
Hexane	20	1	no
DichloroEthane	20	15	no
tetrahydrofuram	123	118	yes
111-tce	26	19	x
Xylenes	45	24	x
Ethyl bnz	75	32	yes

Yes – need calib
X – borderline
no – no need calib



Avg=37%

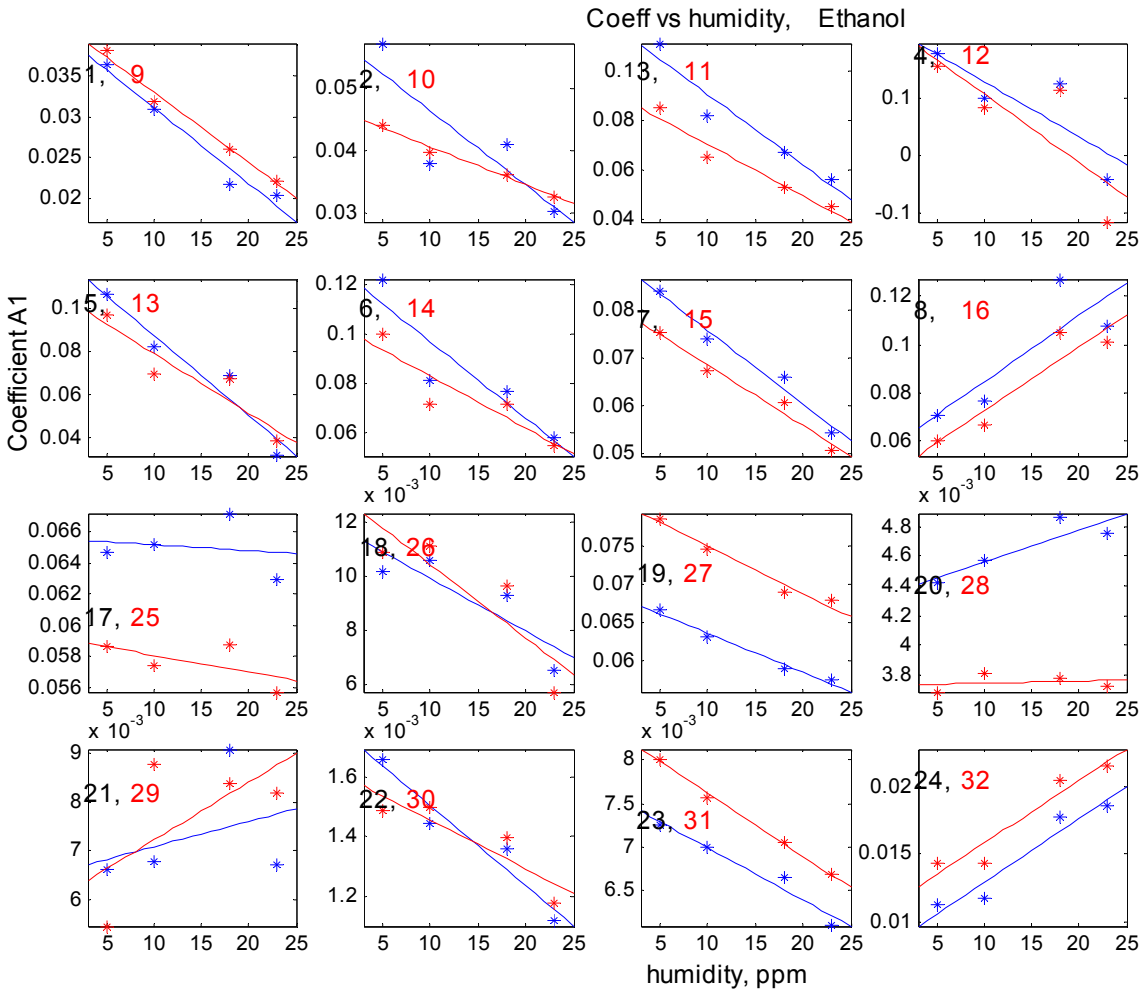
Avg=23%

Most...Least changed Sensors: 5 8 16 17 2 12 25 6 14 4 13 .... 31 20 19 24 1 9 21 28
(Mean % change in fngrprnt): 104 95 76 67 66 64 63 59 48 48 41 .... 19 18 17 17 15 14 12 10

Sensor Calibration - Other Conditions

Jet Propulsion Laboratory
California Institute of Technology

Sensor Coeff vs Base humidity – Linear approximation



Ethanol

Avg correlation:

r^2=0.95

Sensor Calibration - Other Conditions

Summary:

- Base humidity change affects data analysis results
  - Within 5k are almost tolerable; need calibration for bigger change
- Sensor coefficient vs base humidity can be roughly modeled as linear
  - Correlation >0.80 mostly; could be higher if we have better calibration data set
- Calibration needs / strategy:
  - Calibration at two base humidity levels, e.g., 10k & 20k (or 18k)
  - Interpolate /Extrapolate to get the rest
  - Similarly, calibration at two temperature levels

Sensor Selection

- Poorly behaved sensors can be more detrimental to data analysis than being helpful.

- Optimal sensor array selection and weighting
  - For global classification and quantification
  - For local classification and/or quantification
  - For functional group classification and quantification
  - For event detecting
  - *Assistance in sensor material selection*

- ENose sensor selection challenge:
  - Selectivity is important while reliability and sensitivity can be major limiting factors in detecting target compounds reliably
  - Meaningful selection that will actually improve data analysis result

Sensor Selection

- Sensor selection approaches
  - PCA, Cluster Analysis
  - ANOVA-based approach
  - Genetic algorithm optimization w/ appropriate statistic measures

- Sensor selection by ANOVA-based technique:

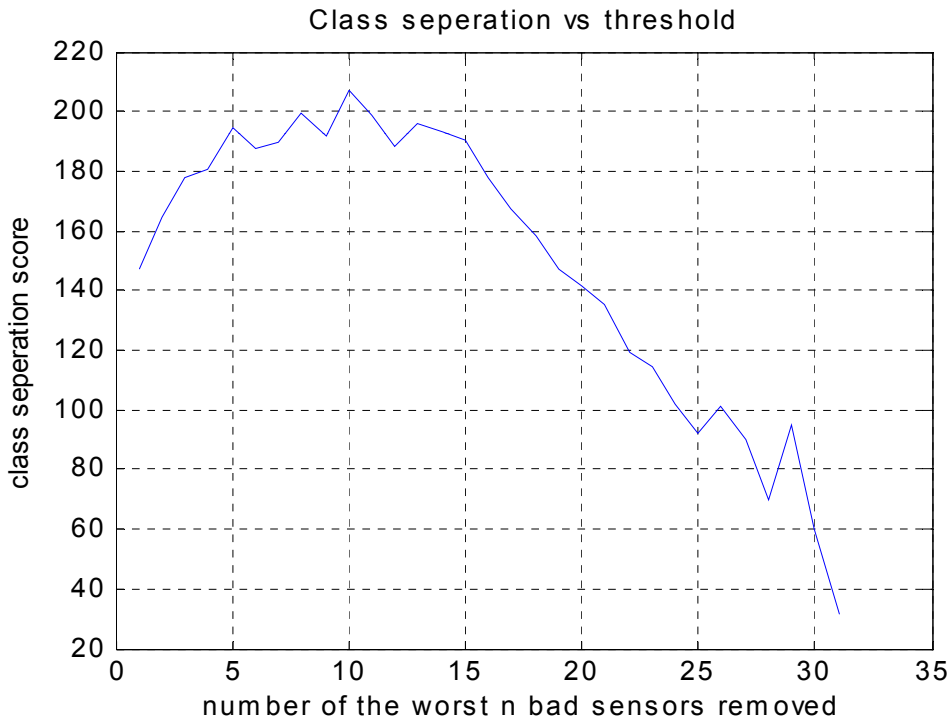
  f =(selectivity / reliability) *sensitivity*correlation

  - Selectivity (*cross-class variation*):
    - Difference bwtn max, min relative response strengths among all analytes
    - Mean of differences of relative response strengths among analyte pairs
  - Reliability (*within-class variation*): mean variation btwn relative actual vs. fitted responses
  - Sensitivity: mean of a sensor's relative response strengths to all analytes
  - Correlation: btwn concentration and response strength

Sensor Selection

- Sensor array optimization:
  - Class separation vs threshold of *f* (worst sensors removed)

    Class separation = cross-class fingerprint distance / within-class fingerprint variation



2nd generation sensors:

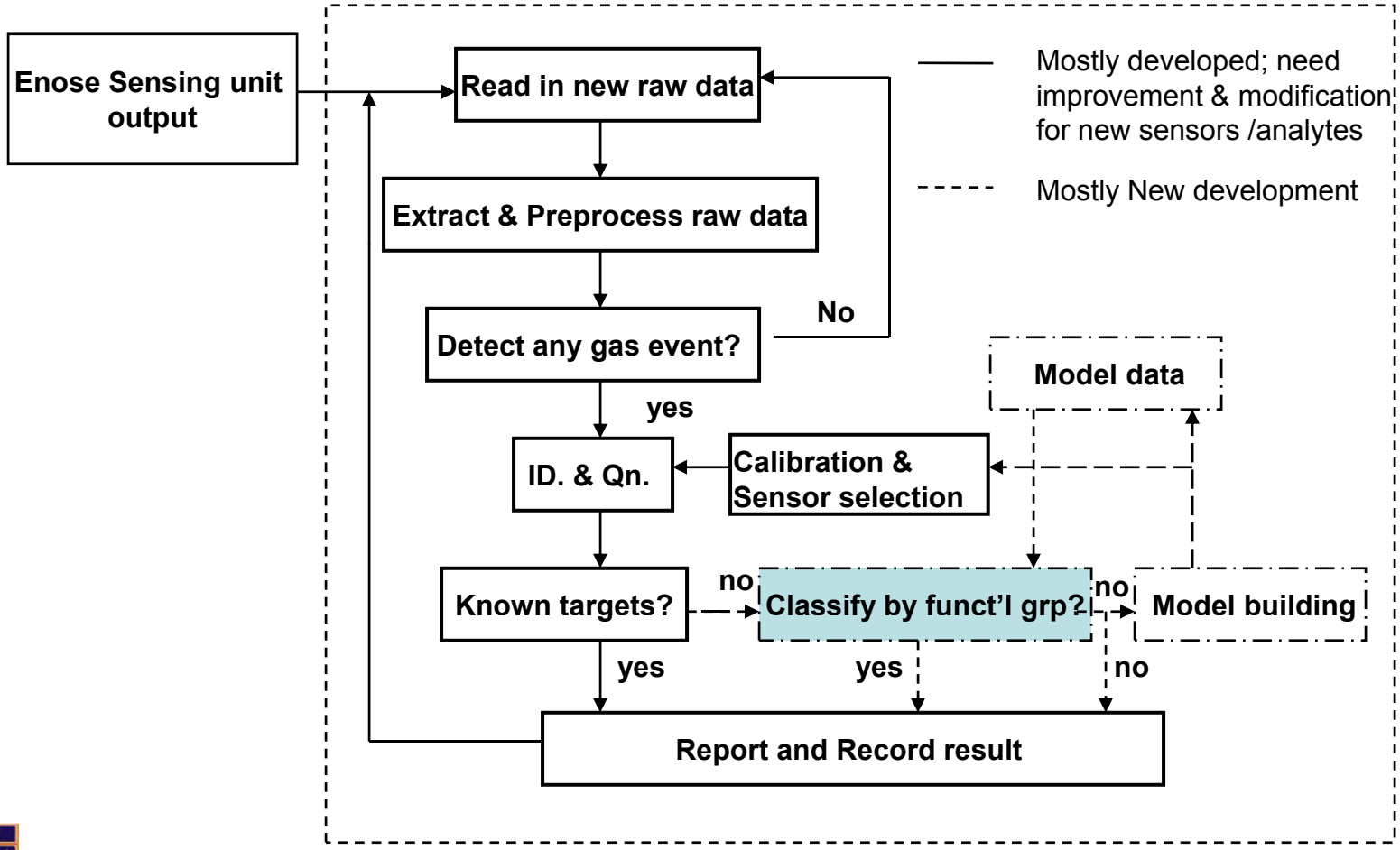
The Best sensors:

30 31 22 25 23 10 17 27…

Best class separation w/ these worst sensors removed:

4 12 8 5 16 20 13 14 1 6

Functional Group Classification



Functional Group Classification

- Analytes can be grouped into “families” or functional groups
  - Based on similarities in physical structure /externally observed properties, e.g.,
    - Alcohol: ethanol, 1-, 2- Butanol, 1-, 2- propanol, etc.,
    - Aromatic: benzene, toluene, chlorobenzene, furan, etc.,
    - One nitrogen: ammonia, acetonitrile, indole, etc.
    - Ketone: acetone, 2-butanone, etc.

    …

  … But not strictly member of one and only one functional group

- Desire to identify an “unknown” (not-trained-for) analyte according to its functional group
  - e.g., identify methanol as an alcohol without being trained for

Funct'l Group Classification - Approach

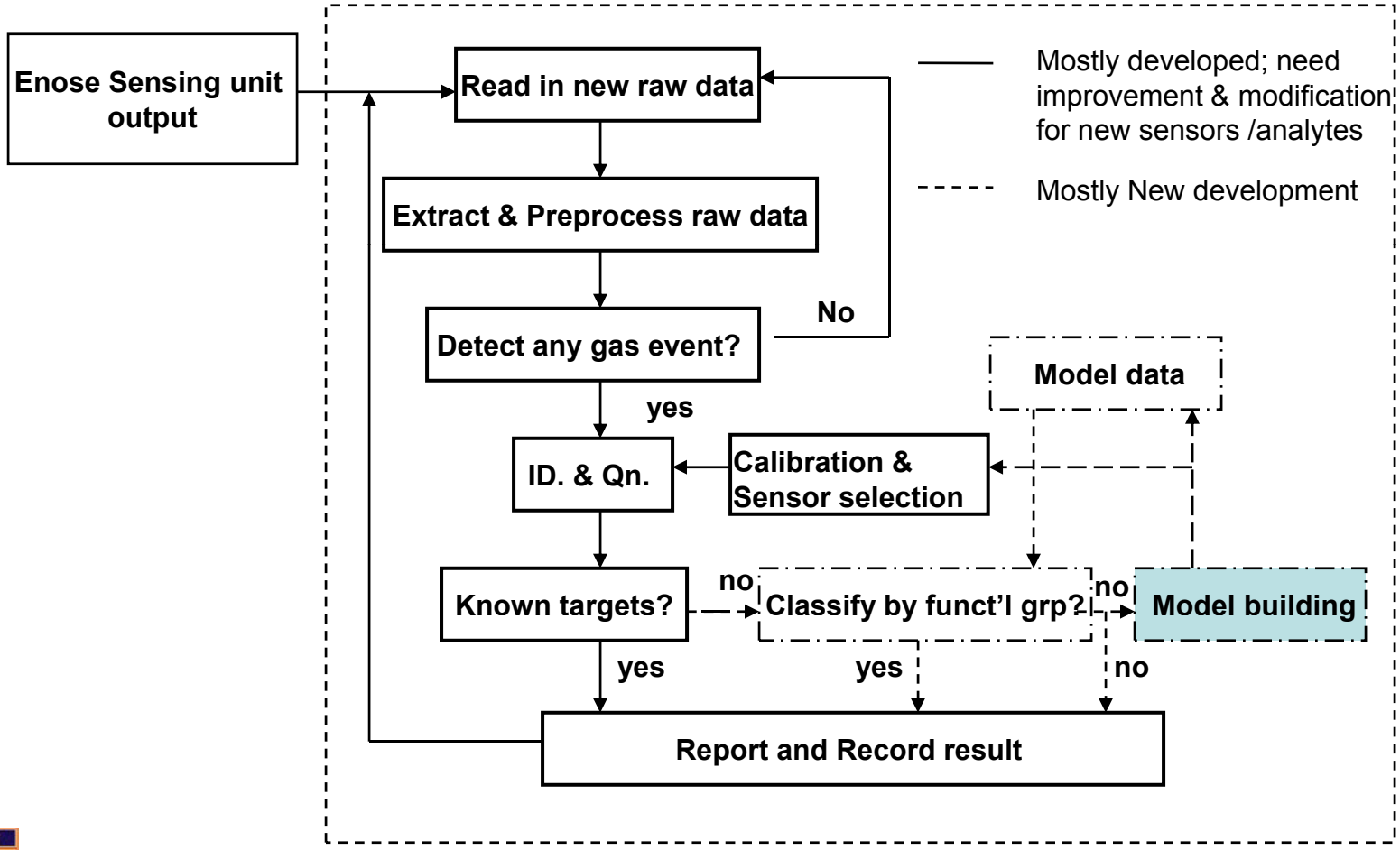
NLS with sensor subset optimized for functional groups

- Select optimum subset sensors that minimize the distance within a funct'l grp and maximize the distance btwn the rest of analytes
- Result in the 1st (regular Id&Qn) step be given preference in the 2nd (funct'l group Id) step
  - e.g., If the 1st step results in "methanol +freon" w/ an unacceptably large residue, then in the 2nd step, check first if "alcohol+freon" would result in an acceptable residue .
- Other possibilities will also be checked and the result will be accepted only if the associated residue is less than that in the above
- if all residues are greater than acceptable levels → unknowns

Funct'l Group Classification –Case Study

- Case studied:
  - Use simulated sensor response (from molecular modeling)
    - **10 trained target analytes**: methanol, ethanol, 1-proponal, methane, ammonia, benzene, formaldehyde, freon 113, indole, toluene
    - **4 untrained alcohols**: 2-propanol, butanol, pentanol, phenol
    - **5 untrained aromatic**: anthracence, ethyl benzene, p-xylene, and m-xylene
  - 170 simulated binary mixture test events:
    - One trained analyte+ one untrained analyte, or two untrained analytes
    - Different concentration combinations
  - Two possible functional groups
    - alcohol and aromatic
  - Average success rate ~ 84%
    - 2-propanol (86%), butanol (67%), pentanol (76%), phenol (67%),
    - Anthracene (86%), Ethyl benzene (95%), p-xylene (95%), o-xylene (86%), and m-xylene (100%).

Model Building



Model Building

- Can we determine the identity of an unknown chemical which causes the sensor array to respond by models built from simulation?
  - Sensor modeling support

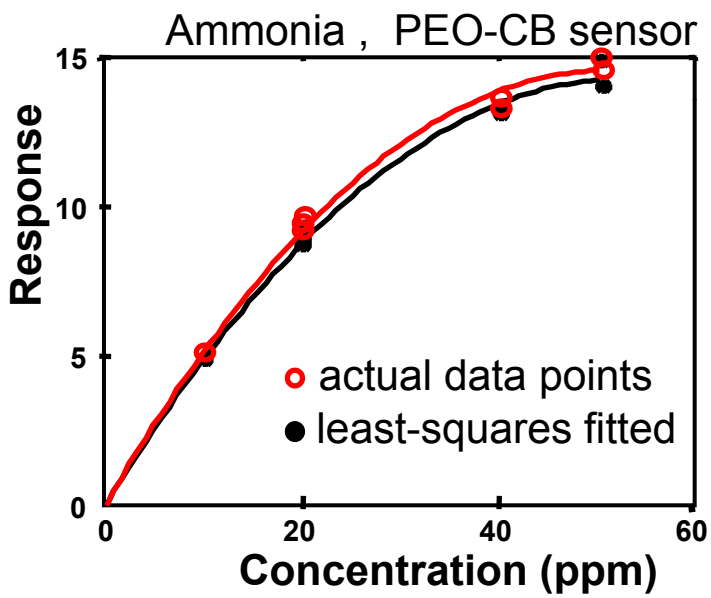
Sensor Modeling Support

- Objectives
  - Develop the ability to determine the identity of an unknown chemical which causes the sensor array to respond.
  - Provide characteristic sensor response of untrained analytes for **functional group identification** based on known responses to trained ones
  - Predict sensor's response under different base humidity given response under nominal humidity

- Approach
  - Quantitative Structure Activity Relationships (QSAR)

MODEL OF SENSOR RESPONSE

- Goal
  - Develop representative response equation (using known experimental data) for each polymer-carbon sensing film in the array (16 equations for the array) to generate virtual training sets for any given sensor array



Experimental sensor response fit:

Resistance change (y) = A_1 x+ A_2 x^2

A1 and A2 : constant; x: concentration

Correlate the constant A_1 from experimental response fit with molecular descriptors using Quantitative Structure-Activity Relationships (QSAR)

MODEL OF SENSOR RESPONSE

QSAR METHODOLOGY

Generate molecular models for polymer-carbon sensing film
Calculate analyte interactions energies w/ polymer-carbon sensing film

Generate QSAR table
Calculate molecular descriptors (analyte and sensor response)

Correlate sensor response to molecular descriptors

- **Generate QSAR equations using Genetic Algorithms**
- **Select the statistically most significant equation containing the polymer-analyte interaction term**

Experimental sensor response (training set)

MODEL OF SENSOR RESPONSE

QSAR DESCRIPTORS

Intrinsic analyte descriptors

molar Refractivity	MR
size of the molecule	V_m
H-bond donor/acceptor	HB_D, HB_A
dipole	
principal moment of inertia	

Sensor interaction energy descriptors

(Electrostatic, vdW, H-bond contributions)

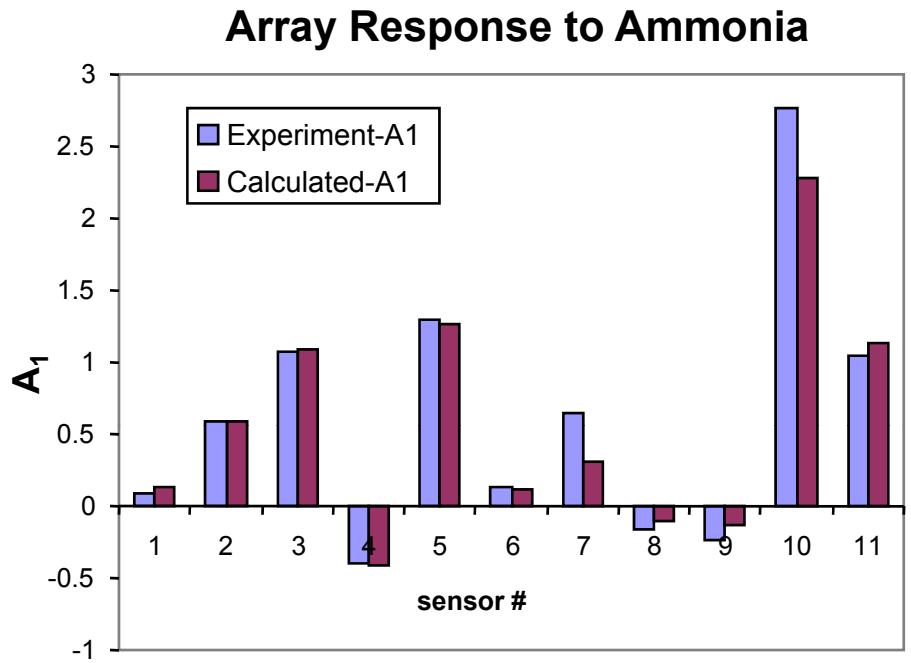
polymer-analyte	E_{pa}
polymer-water	E_{pw}
analyte-water	E_{aw}
analyte-analyte	E_{aa}
polymer-carbon black	E_{pcb}
analyte-carbon black	E_{acb}

Descriptor calculations

Sensor interaction energy : *Monte Carlo technique (sorption calculation)*

Intrinsic analyte descriptors : *Quantitative Structure-Property Relationships (QSPR)*

MODEL OF SENSOR RESPONSE



Polymer list (sensors #)

1 Poly (4-vinylphenol-co-methyl methacrylate)
2. Poly (ethylene-co-acrylic acid)
3. Poly (styrene-co-maleic acid)
4. Poly (2-vinyl pyridine)
5. Poly (4-vinylpyridine)
6. Vinyl alcohol/vinyl butyral copolymer
7. Cyanoethyl hydroxyethyl cellulose
8. Soluble polyimide
9. Polyepichlorohydrin
10. Poly (epichlorohydrin-co-ethylene oxide)
11. Polyethylene oxide

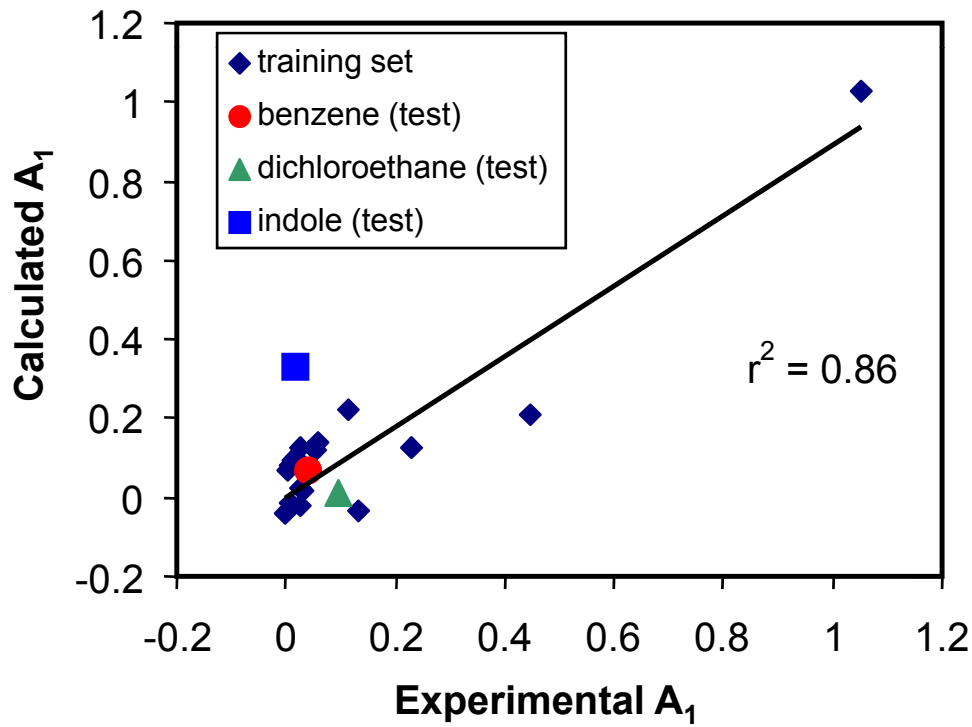
MODEL OF SENSOR RESPONSE

QSAR SENSOR RESPONSE MODEL

Predictive Model

Correlate polymer and analyte interactions to A_1 from analysis algorithm

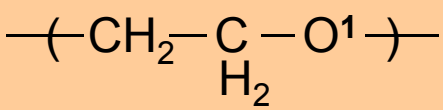
PEO-carbon composite film



Calculated Coefficient A_1=

+	0.116	HB_D^2
+	2.4E-03	MR^2
+	0.152	E_{pa}

monomer



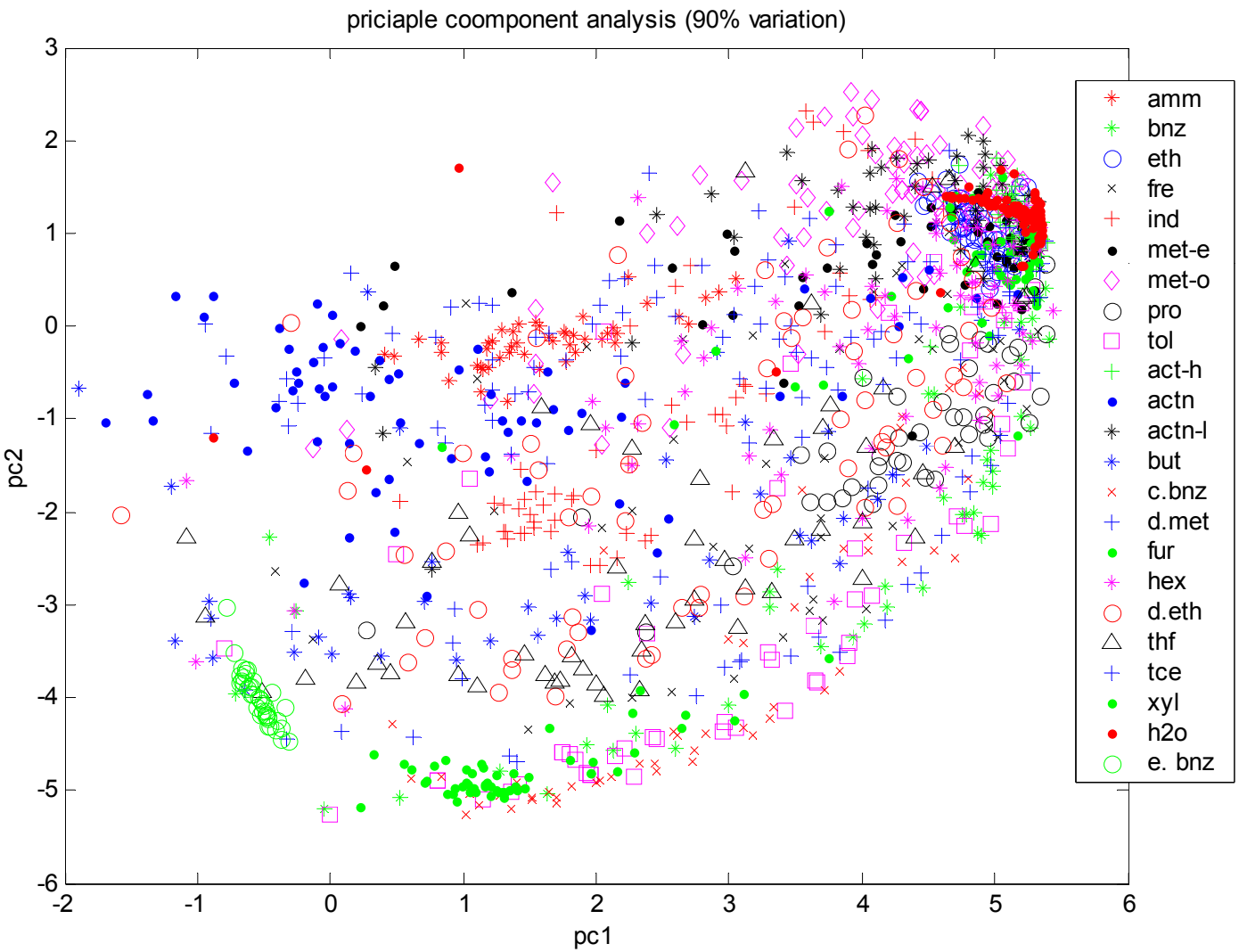
pKa (site1, 25°C) = 15.24 ± 0.10

pKa calculations:
ACD/Lab online software

Backup Material

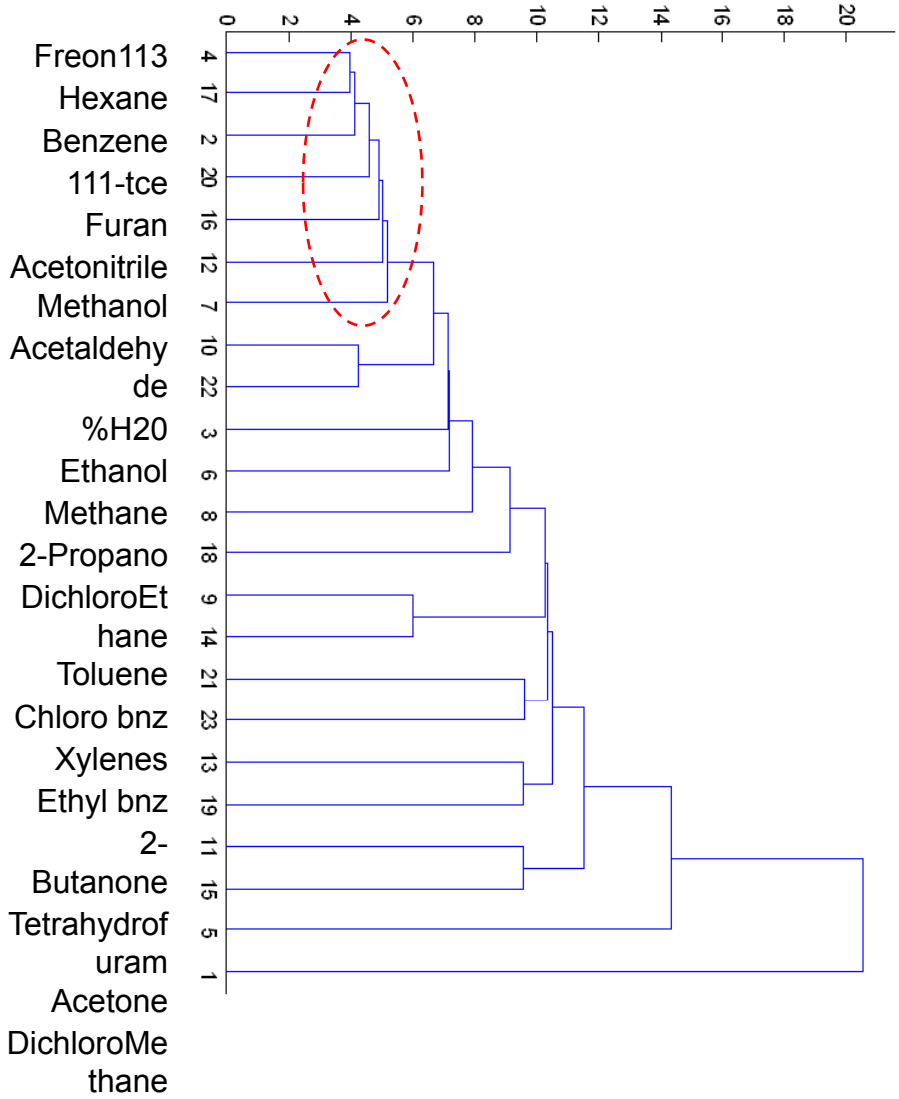
Data Analysis – PCA analysis

- PCA plot of 2nd generation Enose data on 22 analytes



Data Analysis – Cluster Analysis

- Cluster Analysis of 2nd generation Enose data, Single linkage



Data Analysis Results -1st generation

Singles:

- Overall success rate ~ 85%.
  - "Success" is correct identification and quantification (within ± 50%).
- Most failed cases (~10%) are no-event identification -- concentration below identifiable threshold

Gas Compound	Conc Range tested (ppm)	Success Rate (%)
Ammonia	10 - 50	100
Benzene	20 - 150	88
Ethanol	10 - 130	87
Freon 113	50 - 525	80
Formaldehyde	50 - 510	100
Indole	.006 - .06	80
Methane	3000 - 7000	75
Methanol	10 - 300	63
Propanol	75 - 180	80
Toluene	30 - 60	50
%Relative Humidity	5 - 65	100
Wipe	500 - 4000	100

Data Analysis Results -1st generation

Mixtures

- Moderate overall success rate ~60%.
- No training data on mixtures used; total ~75 testing events.
- Sensor loosely follows additive linearity property for the tested binary mixtures.
- Those of low success rates are largely due to low concentrations of one of the mixture components (for which detecting even the single gases is difficult)

Gas Compound 1	Conc. range (ppm)	Gas Compound 2	Conc. range (ppm)	Success Rate (%)
ammonia	20-50	ethanol	100	100
ammonia	10-20	benzene	20-60	65
benzene	50-150	methanol	50	100
benzene	30-100	propanol	75 – 100	60
benzene	10-30	formaldehyde	85 – 250	20
ethanol	40-130	wipe	60-185	90
methanol	25	propanol	80-185	25
methanol	50-75	toluene	30-60	40

Data Analysis Results -1st generation

FLIGHT DATA ANALYSIS

- Every daily marker was correctly identified as 2-propanol. In some cases it was identified as 2-propanol plus a humidity change
- All peaks other than daily markers were identified as humidity change
- In every case where the independent humidity monitor in the cabin recorded a humidity change, it was detected by the ENose. Humidity change quantification by the ENose corresponded to the change recorded by the cabin monitor.
- Independent air samples analyzed by Gas Chromatography/Mass Spectrometry showed no presence of contaminants that should have been detected by the ENose.

Data Analysis Results -2nd generation

- Ground Testing

	Miss	False	Id	Id &Qn
Case 1: all/all:	4%	9%	78%	73%
Case 2[a]: half/all:	4%	10%	78%	73%
Case 3[b]: third/all:	4%	13%	78%	72%
Case 4[c]: 1st half / 2nd half:	4%	11%	77%	73%
Case 5[d]: all /2nd half:	4%	11%	77%	74%
Case 6[e]: 10k coeff/5k data:	3%	11%	73%	67%

a. sort events from all data files by conc for each analytes, then use odd cased for training and all cases for testing
b. Similar as a, but use every other 2 cases for training and all cases for testing
c. Training on 1st (early) half data files; testing on (later) 2nd half data.
d. Testing on 2nd half data which are part of training data
e. Training data are of RH @10kppm, test data are of RH @5kppm

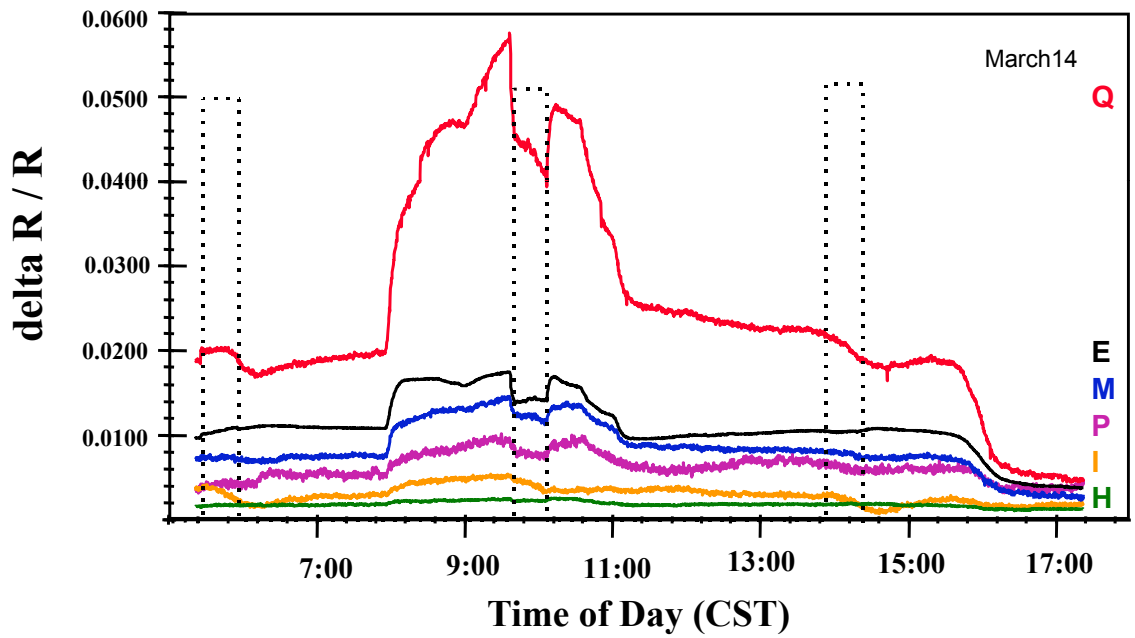
Data Analysis Results -2nd generation

- Breakdown success rates:

Ammonia	(6 - 60)	miss= 0%, false= 5%,	Id=100%,IQ=100%,
Benzene	(3 - 75)	miss= 4%, false= 3%,	Id= 79%, IQ= 79%,
Ethanol	(200-6000)	miss= 0%, false= 1%,	Id=100%,IQ=100%,
Freon113	(15 - 500)	miss=10%,false=12%,	Id= 63%, IQ= 56%,
Indole	(25 -415)	miss=13%,false=42%,	Id= 79%, IQ= 79%,
Methane	(1550 -15000)	miss= 6%, false=22%,	Id= 64%, IQ= 36%,
Methanol	(6 -100)	miss= 5%, false=13%,	Id= 55%, IQ= 45%,
2-Propano	(30 - 400)	miss= 0%, false= 1%,	Id= 96%, IQ= 96%,
Toluene	(5 - 50)	miss= 0%, false= 6%,	Id= 86%, IQ= 76%,
Acetaldehyde	(4 -20)	miss= 4%, false= 5%,	Id= 88%, IQ= 88%,
Acetone	(65 - 600)	miss= 3%, false= 1%,	Id= 91%, IQ= 91%,
Acetonitrile	(4 -25)	miss= 0%, false=11%,	Id= 85%, IQ= 81%,
2-Butanone	(15 - 150)	miss= 4%, false= 8%,	Id= 75%, IQ= 71%,
Chloro bnz	(6 -30)	miss= 8%, false=17%,	Id= 50%, IQ= 50%,
DichloroMethane	(10 - 150)	miss= 7%, false=18%,	Id= 86%, IQ= 79%,
Furan	(1 - 35)	miss= 0%, false= 9%,	Id= 60%, IQ= 53%,
Hexane	(16 -150)	miss= 0%, false=11%,	Id= 71%, IQ= 63%,
DichloroEthane	(10 -100)	miss= 0%, false= 6%,	Id= 44%, IQ= 42%,
Tetrahydrofuram	(13 - 120)	miss= 0%, false= 4%,	Id= 71%, IQ= 71%,
111-tce	(7 -200)	miss=13%,false=18%,	Id= 49%, IQ= 47%,
Xylenes	(33 - 300)	miss= 0%, false=24%,	Id= 96%, IQ= 96%,
%H20	(1000 - 13400)	miss= 7%, false= 4%,	Id= 89%, IQ= 89%,
Ethyl bnz	(20 - 185)	miss= 0%, false= 2%,	Id=100%,IQ=100%,
	(conc range)	Avg= **miss= 4%, false= 11%,**	**Id= 77%, IQ=73%,**

Enose Data Analysis - Overview

Sensor response recorded in an Early Human Testing (EHT) chamber test



When the EHT chamber door was opened, there is significant change in resistance in Enose sensors. Dotted vertical lines show 30 minute reference cycles with a corresponding dip in sensor response.

JPL Electronic Nose (ENose) Comparison w/State of Art

Relative to current ISS monitoring

- No real-time event monitor currently exists aboard ISS.

Relative to higher-end analytical instruments (*e.g.* VOA, VCAM):

- Low mass, small volume, low power, rugged.
- No consumables required.
- Runs continuously (30 to 360 data points/hr).
- Excellent complement to these instruments - can serve as trigger to activate them.

Relative to other ENoses, which are based on pure conducting polymers:

- Much wider range of usable polymers - superior fingerprinting.
- Quantification included in analyte identification.
- Superior environmental stability.
- Tunable resistance properties - facilitates readout.

Relative to materials-specific sensors:

- Can be trained to recognize new materials and unknowns.
- Can be used to deconvolute and identify simple mixtures of knowns.
- One array serves to detect and identify a wide variety of analytes.

Second Generation JPL ENose: 21 Analytes

tested concentrations

	24 hr SMAC (ppm)	1/3 SMAC (ppm)	3x SMAC (ppm)	low (ppm)	high (ppm)
acetaldehyde (1500ppm gas)	6	2.00	18	2	20
acetone	200	66.67	600	65	600
acetonitrile (1500 ppm gas)	4	1.33	12	1	25
ammonia - NH3	20	6.67	60	6	60
benzene (1510 ppm gas)	3	1.00	9	3	75
2-butanone	50	16.67	150	15	150
chlorobenzene	10	3.33	30	3	30
dichloromethane	35	11.67	105	11	150
dichloroethane	0.4	0.13	1.2	0.12	98
ethanol	2000	666.67	6000	665	6000
ethyl benzene	60	20.00	180	16	184
freon 113	50	16.67	150	15	150
hexane	50	16.67	150	16	150
indole	0.3	0.10	0.9	0.10	0.90
methane	5300	1766.67	15900	3000	40000
methanol	10	3.33	30	3	100
isopropanol	100	33.33	300	30	600
tetrahydrofuran	40	13.33	120	13	120
toluene	16	5.33	48	5	50
1,1,1-trichloroethane	11	3.67	33	5	200
o,p-xylenes	100	33.33	300	33	300

JPL Enose Program Overview and Background

Jet Propulsion Laboratory
California Institute of Technology

Analytes for Technical Demonstration Onboard ISS

	ANALYTE	Target Con (ppm)	JUSTIFICATION	QUANT. RANGE (ppm)	ODOR THRESHOLD* (ppm)
TIER 1	Ammonia	5	Coolant; 24 hr SMAC 20	1.6 – 15	50
	Mercury	0.01	Hg vapor in lighting	.003 - .03	odorless
	Sulfur Dioxide	1	Thionyl chloride batteries	.3 – 3	3
TIER 2	Acetone	270	24 Hr SMAC	90 – 810	20 - 600
	Dichloromethane	10	24 Hr SMAC	3 – 30	200 - 300
	Ethanol	500	Solvent; 24 hr SMAC 2000	166 – 1500	10
	Freon 218	20	Russian AC coolant	6 – 60	na
	Methanol	10	24 Hr SMAC	3 – 30	10 - 100
	2-Propanol	100	24 Hr SMAC	30 – 300	20 - 40
	Toluene	16	24 Hr SMAC	5 – 50	0.2 - 2
TIER 3	Formaldehyde	TBD	24 Hr SMAC	TBD	0.5 - 1

* SOURCE: US Coast Guard, Chemical Hazards Response Information System; http://www.chrismanual.com/
National Institutes of Health, Hazardous Substance Data Bank http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB