Mijn naam is Jim Bezdek

Hallo en welkom !

Ik woon in Pensacola

Amberjack
Pensacola, FL

Anomaly Detection in Wireless Sensor Networks

Kingfish Napier, NZ

- Introduction to Wireless Sensor Networks
- DCAD: Data Capture Anomaly Detection
- ESAD: Elliptical Summary Anomaly Detection
- Four Measures of dissimilarity for Ellipsoids
- Visual assessment : estimating c with iVAT
- Anomaly detection with SL clustering

WSNs

Sensors
Transceivers
Memory
Processor

✖ But … Limited

Power (Battery)
Bandwidth
Memory
Processing capacity

✔ Cheap and small

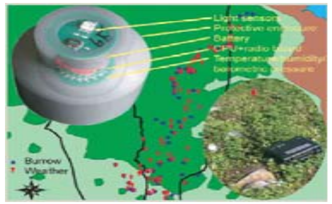
Use in unattended areas
(Environmental Monitoring)

Great Duck Island to monitor the sea bird *Leach's Storm Petrel*

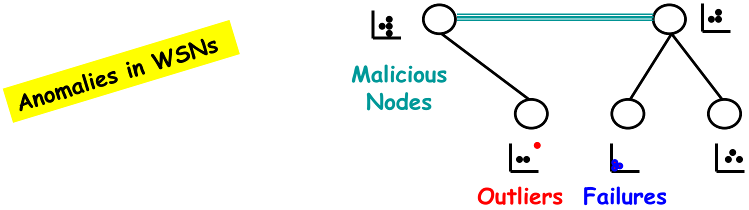
43 Sensor nodes deployed at GDI, 15 km off coast of Maine, USA, summer, 2003

4D Data { Light intensity, Temperature, Humidity, Pressure }

Sensors inside and outside underground nesting burrows



Anomalies : observations inconsistent with the bulk of the data



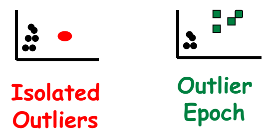
Causes

Sudden environmental change (clouds obscure the sun)
Nodes lose calibration, power failure, physical damage
Malicious attacks (data injection, *angry seabirds* !!)

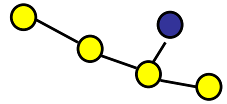
Detection

Analyze measurements/traffic in the network
Model normal behavior to classify anomalies

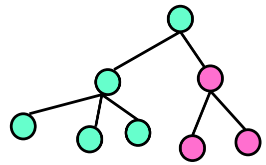
4 types of Anomalies



(1st Order)
Internal Node Anomalies

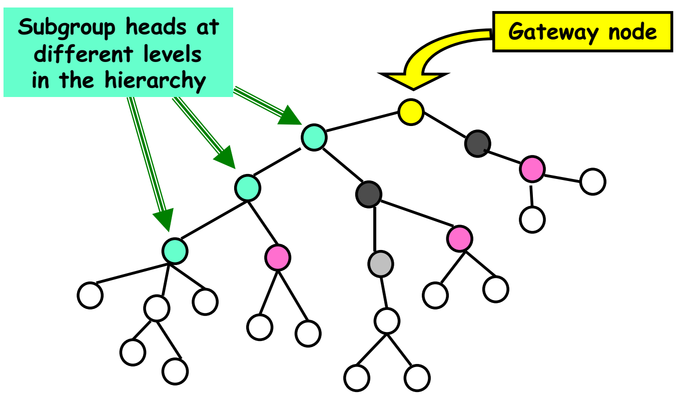


(2nd Order)
Whole Node Anomalies

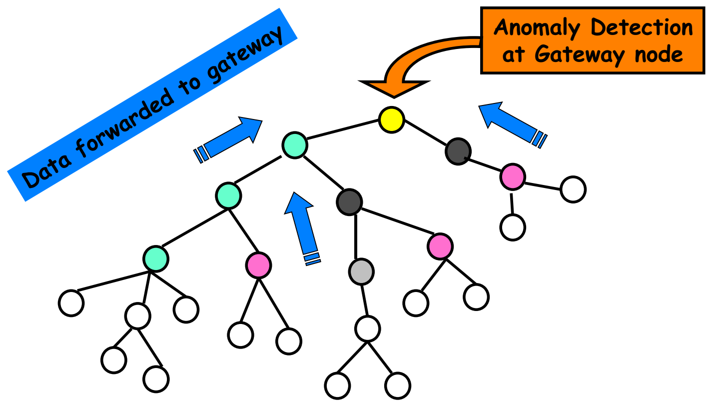


(HO Order)
Anomalous *Subtree*

Hierarchical WSN Topology



The *Centralized Approach* with *sensor data* routed to gateway



Problems with the Centralized *Data* Approach

High energy consumption for data communication

Nodes near the gateway become a bottleneck

Imbalanced load distribution in the network

Reduced physical lifetime of the network

Scalability (*slows down* as nodes are added)

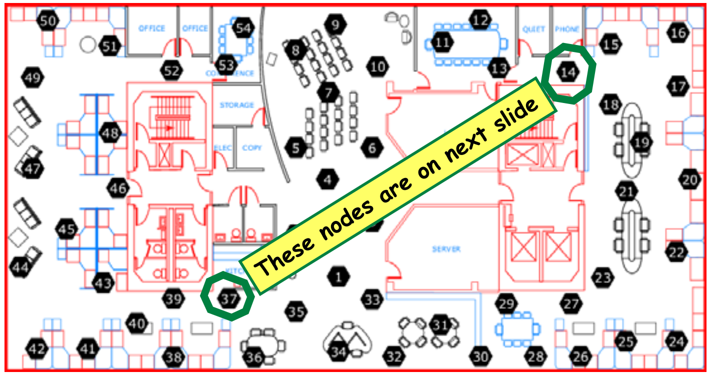
Giant Trevally Maui, Hawaii

DCAD: The Data Capture Anomaly Detection Model

Rajasegarar, Bezdek, Leckie, and Palaniswami (2009). Elliptical Anomalies in Wireless Sensor Networks", *ACM TOSN*, 6(1).



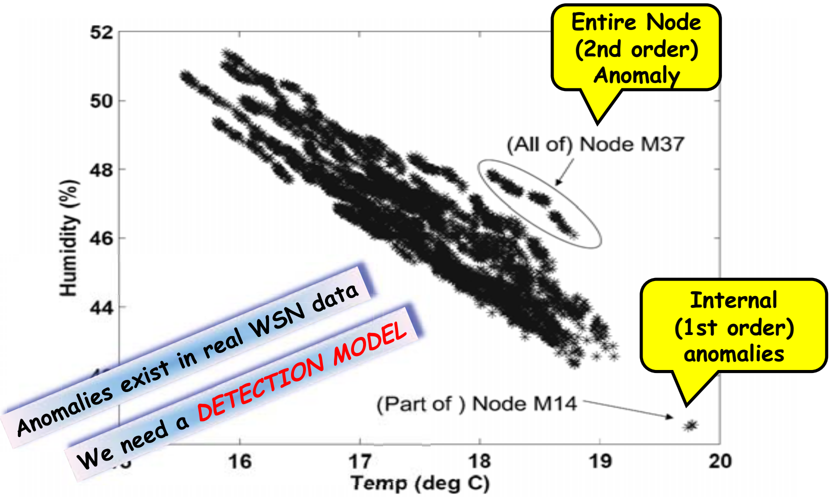

Example : Intel Berkeley Research Laboratory (IBRL) WSN



Period: March 1, 2004
2D Data collected
x = (temp., humidity)

54 nodes deployed
Sampled every 30s
Time window = 4 hr

Scatterplot of all data at the gateway node in the IBRL WSN



A Flashback to … (Hyper) Ellipsoids

Basic notation	$\mathbf{x}, \mathbf{m} \in \Re^p,\ S \in PD \subset \Re^{p \times p}$
Inner product norm induced by S	$\|\mathbf{x}-\mathbf{m}\|_S = \sqrt{(\mathbf{x}-\mathbf{m})^T S (\mathbf{x}-\mathbf{m})}$
Euclidean norm S = pxp identity I_p	$\|\mathbf{x}-\mathbf{m}\|_{I_p} = \sqrt{(\mathbf{x}-\mathbf{m})^T I_p (\mathbf{x}-\mathbf{m})}$
Mahalonobis norm m = mean(X) S = cov (X)	$\|\mathbf{x}-\mathbf{m}\|_{S^{-1}} = \sqrt{(\mathbf{x}-\mathbf{m})^T S^{-1} (\mathbf{x}-\mathbf{m})}$

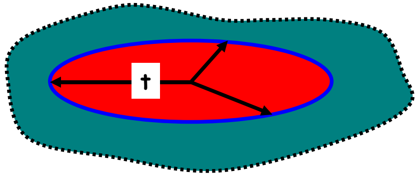
The geometry of Ellipsoids

Ellipsoid $E(S^{-1}, m;\ t)$ induced by S^{-1}, centered at m

Interior E_{int} $\quad E(S^{-1}, \mathbf{m}; t) = \{\mathbf{x} \in \Re^p \mid \|\mathbf{x}-\mathbf{m}\|^2_{S^{-1}} < t^2\}$

Surface E_{surf} $\quad E(S^{-1}, \mathbf{m}; t) = \{\mathbf{x} \in \Re^p \mid \|\mathbf{x}-\mathbf{m}\|^2_{S^{-1}} = t^2\}$

Exterior E_{ext} $\quad E(S^{-1}, \mathbf{m}; t) = \{\mathbf{x} \in \Re^p \mid \|\mathbf{x}-\mathbf{m}\|^2_{S^{-1}} > t^2\}$

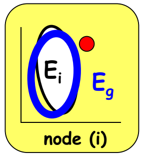


"effective radius" t

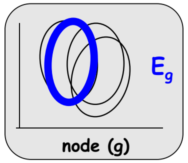
Data Capture Anomaly Detection (DCAD) Model

E_i = Ellipse at node i

$E_g = \Phi(E_i)$ = (Compound) Ellipse at gateway node g



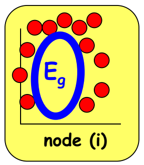
Transmit parameters of E_i's – not DATA



Detection of 1st order anomalies with level sets of E_i

[or backpropped] level sets of E_g on node data

Detection of 2nd/Higher order anomalies with backpropped level sets of E_g on node data

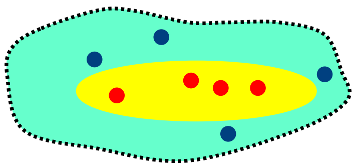


Anomaly Detection by *data capture* with Ellipsoids

Sample $X = \{x_1, \ldots, x_n\} \subset R^p$: m = mean(X); S = cov(X)

"Normal" Points $NP_{X,t} = E_{int} \cup E_{surf}$ $\quad NP_{X,t} = \{\mathbf{x} \in \Re^p \mid \|\mathbf{x}-\mathbf{m}\|^2_{S^{-1}} \le t^2\}$

Anomalous Points $AP_{X,t} = E_{ext}$ $\quad AP_{X,t} = \{\mathbf{x} \in \Re^p \mid \|\mathbf{x}-\mathbf{m}\|^2_{S^{-1}} > t^2\}$



$NP_{X,t} \cap AP_{X,t} = \varnothing$

$NP_{X,t} \cup AP_{X,t} = \Re^p$

$E(S^{-1}, m;\ t)$ has *user-defined* "effective radius" t

Three definitions of EAs depend on *choice of t*

Elliptical *Cardinality* Anomalies (ECAs)

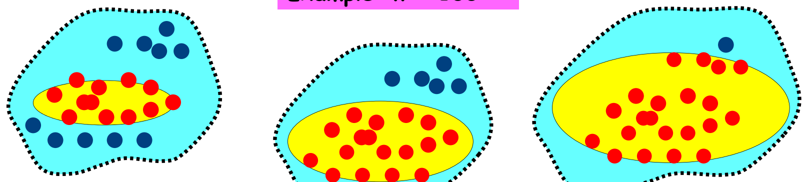
Compute and order n distances $\{d_k\} = \{\|x_k - m\|^2_{S^{-1}}\} \rightarrow d_{(1)} \le \ldots \le d_{(n)}$

Choose capture % c, defines t $c \in \{0.01, \ldots, 1.00\} \rightarrow t = d_{(\lfloor cn \rfloor)}$

$NP_{X,(\lfloor cn \rfloor)}$ 100 c% "normal" data

$AP_{X,(\lfloor cn \rfloor)}$ exterior (anomalous) data

Example: n = 100



Elliptical *Number of Sigmas* Anomalies (ENSAs)

Assume $X \sim n(\mu, \Sigma) \sim$ Gaussian

n_σ = # of Std. Devs. from m

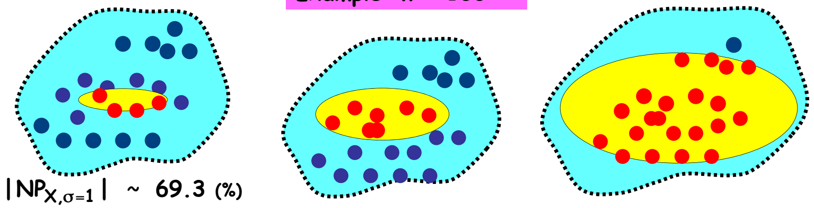
Choose prob(n_σ), define t=n_σ

$n_\sigma \in \{1, 2, \ldots, k\} \rightarrow t = n_\sigma$

$NP_{X,n_\sigma)}$ 100 c% "normal" data

AP_{X,n_σ} exterior (anomalous) data

Example: n = 100



Elliptical *chi-squared* Anomalies (ECSAs)

Assume $X \sim n(\mu, \Sigma)$

Pick $\alpha \in \{0.01, \ldots, 1.00\}$

Choose % (1-α), defines t

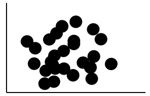
$t = \chi^2_p(\alpha)$

$NP_{X,\chi^2_p(\alpha)}$ 100(1-α) % "normal"

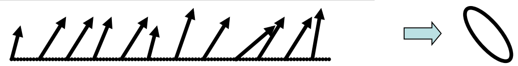
$AP_{X,\chi^2_p(\alpha)}$ exterior anomalies

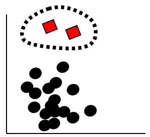
$$\underbrace{\text{limit}}_{n\to\infty} \{ENSA - ECSA\} = 0$$

1st Order (or Type 1) Anomalies: *Internal to a node*

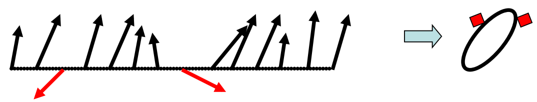


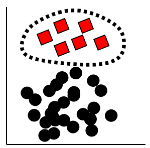
Normal node : all vectors similar



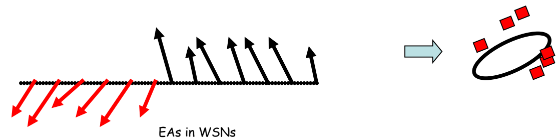


Data anomalies: odd vectors differ

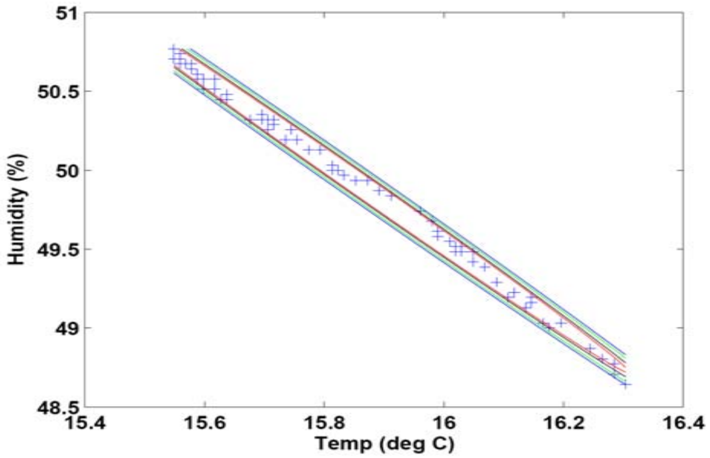




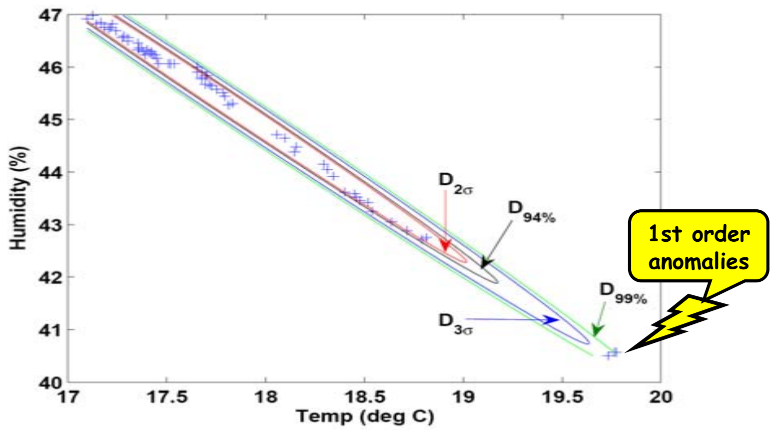
Epoch anomaly: contiguous vectors differ



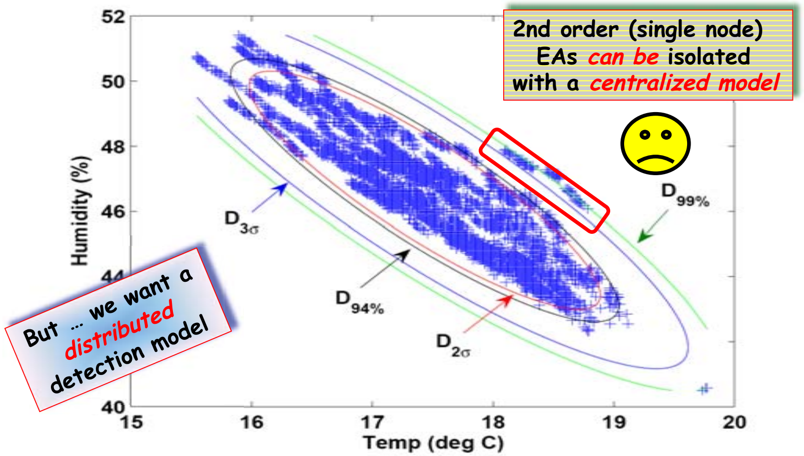
IBRL Node 50 : Example of a normal node



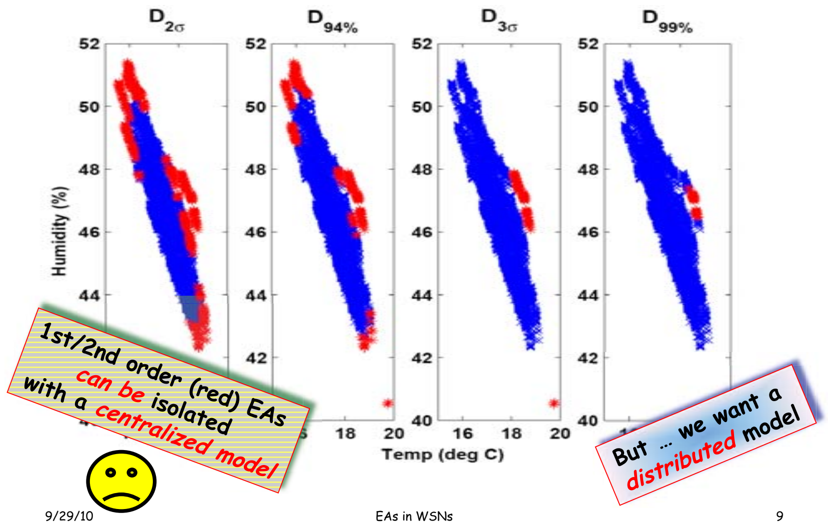
1st order *ECA* and *ENSA* anomalies at IBRL Node 14



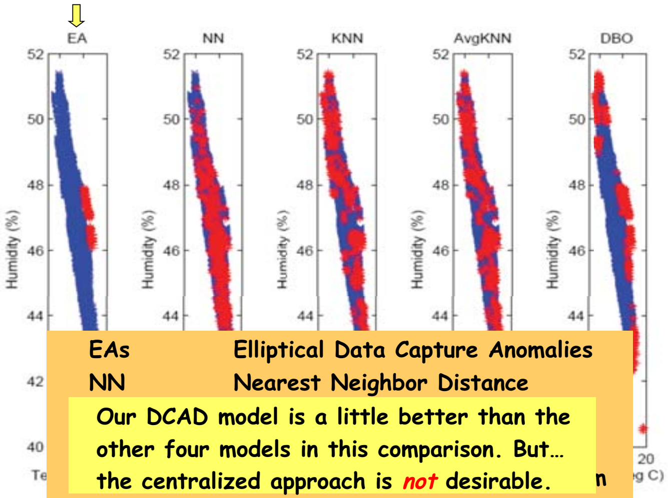
ECA and *ENSA* ellipses based on *all data* at IBRL gateway Node



ECAs and *ENSAs* based on *all data* sent to the gateway node



5 centralized methods on all data at the gateway node



EAs	Elliptical Data Capture Anomalies
NN	Nearest Neighbor Distance

Our DCAD model is a little better than the other four models in this comparison. But… the centralized approach is *not* desirable.

Speckled Trout
Pensacola, FL

ESAD: The Elliptical Summaries Anomaly Detection Model

Bezdek, Havens, Keller, Leckie, Park, Palaniswami, Rajasegarar,(2010). Clustering elliptical anomalies in sensor networks, *Proc WCCI 2010*.

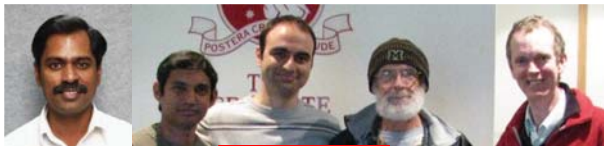
Bezdek, Rajasegarar, Moshtaghi, Havens, Leckie, Palaniswami (2010). Anomaly Detection in Environmental Monitoring Networks, in press, *IEEE CIM*.

Moshtaghi, Havens, Park, Bezdek, Leckie, Rajasegarar, Keller, Palaniswami (2010). Clustering Ellipses for Anomaly Detection, in press, *Patt. Recog*.

Rajasegarar, Bezdek, Moshtaghi, Leckie, Palaniswami, Havens, (2010). Measures for Clustering and Anomaly Detection in Sets of 3D Ellipsoids, in review, *CVGIP*.

Our WSN Gang

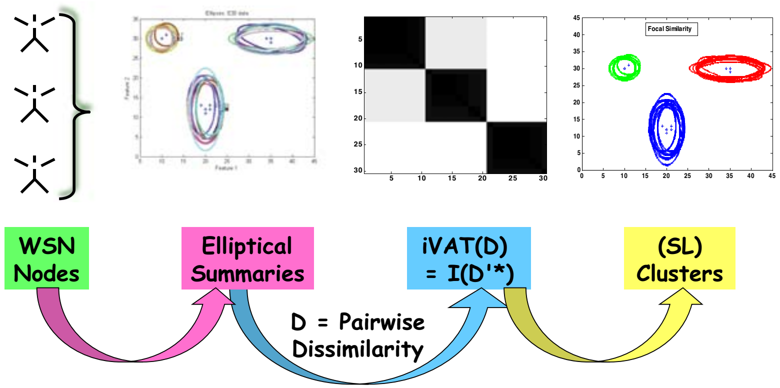
U. of Missouri

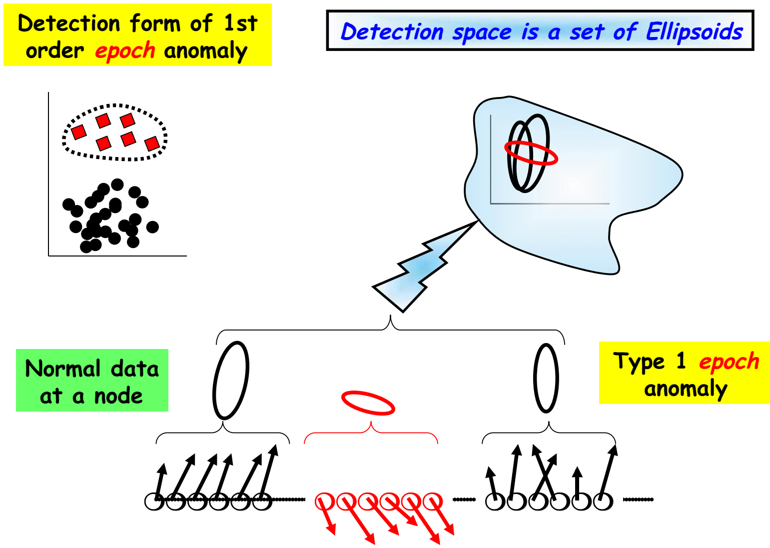


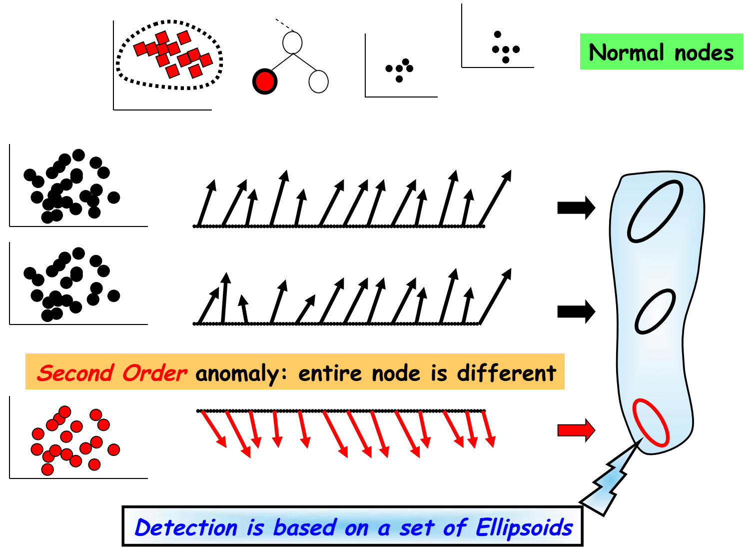
U. of Melbourne

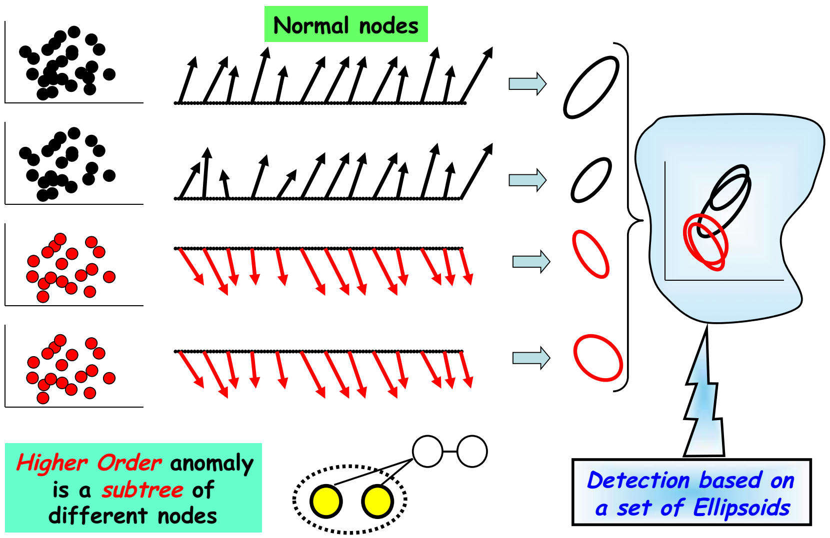
U. of W. Sydney

Elliptical Summary Anomaly Detection (ESAD) Model

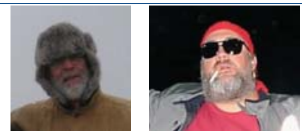








Bezdek, Hathaway (2002). VAT: A tool for visually assessing (cluster) tendency, *Proc. IJCNN*, 2225-2230.



Wang, Leckie, Kotagiri, Bezdek (2010). iVAT : Visual Analysis for cluster tendency assessment, *in press, PAKDD*.

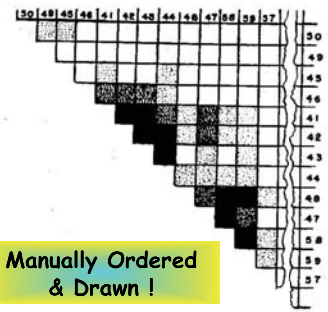
Havens, Bezdek (2010). Recursive (iVAT), *in review, TKDE*.

The First RDI

Cattell, R. (1944) *Psychometrika*, 9(3), 169-184

"With or witho... *ge* lists he then ... f the variabl... ring linkage ... *ge the diagonal* or ... as possible. If the process is successful, the result ... nd strikir ... "

An RDI in 1944… ????… Duhhh !



Manually Ordered & Drawn !

3 intensities

Black : 1 ≥ r > 0.6
Grey : 0.5 ≤ r <0.6
Light : 0 ≤ r < 0.5

Visual Assessment : VAT

Our "Claim" in 2002 — The utility of I(D) for visual assessment of clusters depends on the *ordering* in D

Our "Idea" in 2002

Reorder D → D* so groups of similar objects are close in D*

Nb: *Cattell published both the claim and idea in 1944 !*

"Man learns from history that man learns nothing from history" … Hegel

The *(2002)* VAT Algorithm (n ≤ ~ 5000)

USER FRIENDLY - NO PARAMETERS TO CHOOSE !

Input — Dissimilarity data D_{nxn} (if S, D=[1]-S)

Set — K={1,…,n} : I = J = ∅ : ... … 0]

Find — (i, j)∈ argmax$_{p\in K}$ ... ⇒ P(1) = i ; I = {i} ; J = K-{i}

Loop — ... k = 2 to n

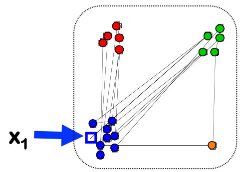
(i, j)∈argmin$_{p\in K, q\in K}$ {D_{pq}}

P(k) = i : I = I ∪ {j} : J = J - {j}

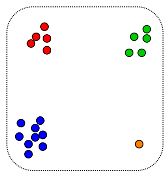
VAT RDI — $[D^*]_{ij}=D_{P(i)P(j)}$ for 1 ≤ i,j ≤ n

Display — D* as Image I(D*)

Example

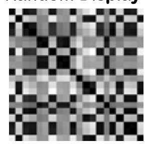


Random Edge Sequence



Data X

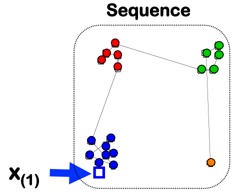
I(R)=



Random Display

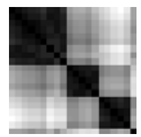
4 clusters ?

1 Largest
2 Medium
1 Singleton



VAT Edge Sequence

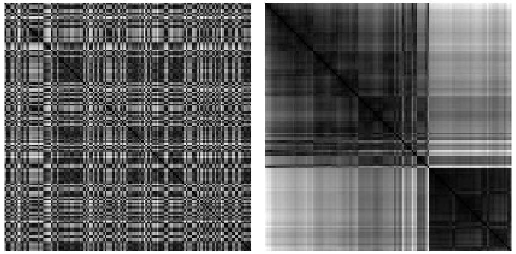
I(R*)=



VAT Display

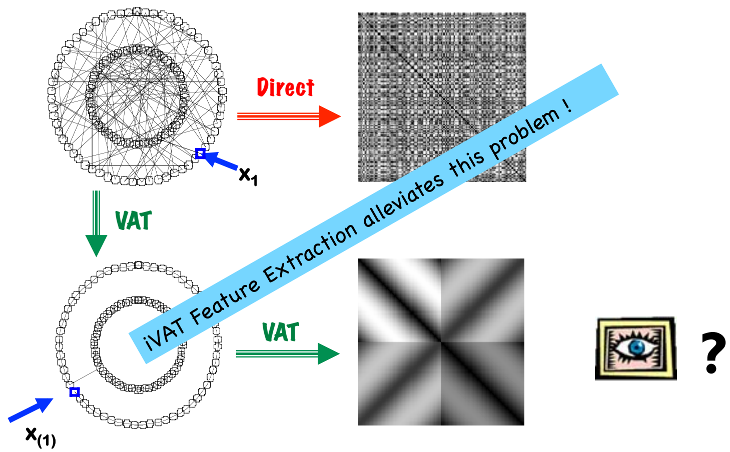
Example : IRIS Data

n = 150 vectors $c_{physical}$ = 3
p = 4 features $c_{geometric}$ = 2



You can 2 clusters
Right Answer !

Example : Two Rings Data



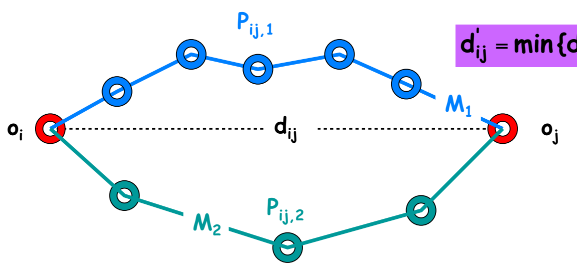
VAT image I(D*) may be blurry, or point to the "wrong" c

Transform D to D' *before* VAT by *Path-based distance* from i to j

P_{ij} = All paths in D from i to j

$$d'_{ij} = \min_{p \in P_{ij}} \{ \max_{1 \le h < |p|} \{ d_{p[h]p[h+1]} \} \} \Rightarrow D \mapsto D'$$

$$d'_{ij} = \min\{d_{ij}, M_1, M_2\} = M_1$$



This distance is well suited to chained (SL~VAT) clusters

The *(2010B) recursive* iVAT Algorithm ($n \lesssim$ ~ 5000)

Input VAT RDI D*

Set D'*=[0]

Loop For k = 2 t...

... min $\{D^*_{...}\}$ $_{k=1,\dots,r-1}$

$D'^*_{r...}$ ... $c = j$

... = max{D'^*_{jc} }; $c = 1, \dots, r-1; c \ne j$

Loop For j = 2 ... < j : $D'^*_{ji} = D'^*_{ij}$

iVA... ... $_{ij}$

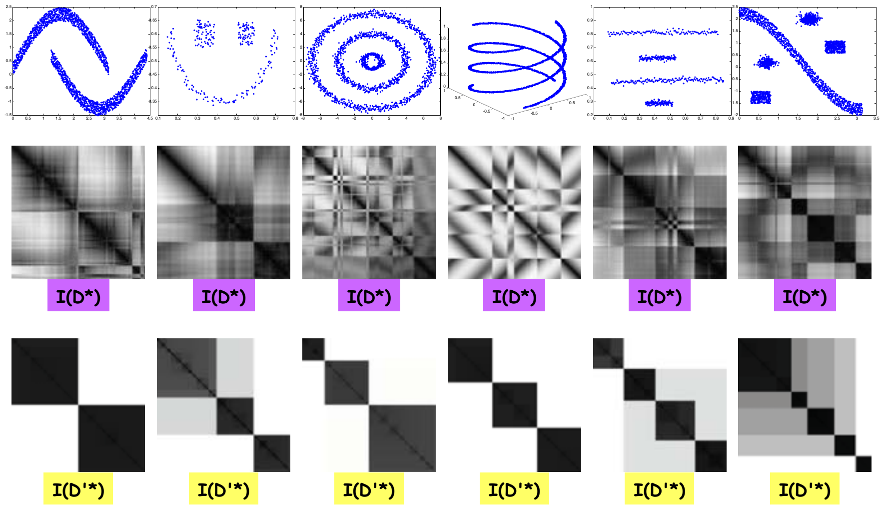
Display D'* as Image I(D'*)

USER FRIENDLY - NO PARAMETERS TO CHOOSE !

2010A iVAT: iterative construction of D', D*=VAT(D') is O(n^3)

2010B iVAT: recursive construction of D* from D* is O(n^2)

iVAT = VAT applied to D' instead of D : *Synthetic Data*



Conclusions : Why Visual Clustering ?

"The preliminary examination of most data is facilitated by the use of diagrams. *Diagrams prove nothing, but bring outstanding features readily to the eye*; they are therefore no substitute for critical tests as may be applied to the data, but are valuable in suggesting such tests, and in explaining conclusions founded upon them."

R. A. Fisher (1924) *Statistical Methods for Research Workers*

"[visual methods] should in no way be regarded as methods to be used to the *exclusion* of other types of multivariate analysis; indeed they will, in general, be *most* helpful when used alongside, and in association with, other forms of analysis."

Brian Everitt (1978) *Graphical Techniques for Multivariate Data*

'What is the use of a book', thought Alice, *'without pictures…'*

Lewis Carroll (1866), *Alice in Wonderland*

Measures of Dissimilarity for Pairs of Ellipsoids

Kawai, Napier, NZ

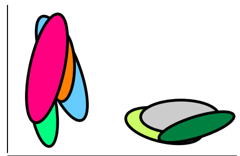
Similarity Measures for Ellipsoids $E_u = \{\cup E_j\}$

$s(E_i, E_j)$ is a (strong) similarity measure on $E_u \times E_u$ $\Leftrightarrow$

- $0 \le s(E_i,E_j) \le 1 \;\; \forall i,j$
- $s(E_i,E_j) = 1 \Leftrightarrow E_i = E_j$
- $s(E_i,E_j) = s(E_j,E_i) \;\; \forall i \ne j$

s is a (weak) similarity measure if $s(E_i,E_j)=1$ but $E_i \ne E_j$

Important properties for pairwise similarity of Ellipsoids



Location (centers of E_i and E_j)

Shape (eigenvalues of E_i and E_j)

Orientation (eigenvectors of E_i, E_j)

Compound Normalized Similarity $s_{cn}(E_i, E_j)$

evs of $E_i = (S_i^{-1}, m_i, 1)$ $\quad \alpha = \{\alpha_1 \le \alpha_2 \le \ldots \le \alpha_p\} \to \alpha^* = (1/\sqrt{\alpha_1}, \ldots, 1/\sqrt{\alpha_p})^T$

evs of $E_j = (S_j^{-1}, m_j, 1)$ $\quad \beta = \{\beta_1 \le \beta_2 \le \ldots \le \beta_p\} \to \beta^* = (1/\sqrt{\beta_1}, \ldots, 1/\sqrt{\beta_p})^T$

(vector) of angles between $\{(\alpha_i, \beta_i)\}$ $\quad \theta = \arccos(\mathrm{diag}(R_i^T R_j))$

$$s_{cn}(E_i, E_j) = e^{-\left(\underbrace{\left\| m_i - m_j \right\|^2_{(S_i + S_j)^{-1}}}_{\text{Location}} + \underbrace{\left\| \alpha^* - \beta^* \right\|}_{\text{Shape}} + \underbrace{\left\| \sin\theta \right\|}_{\text{Orientation}}\right)}$$

Theorem

s_{cn} is a (strong) similarity measure on $E_u \times E_u$

Transformation Energy Similarity $s_{te}(E_i, E_j)$

Basic Idea: Measure the "Energy" needed to transform E_i, E_j into a common space

$E_k = E(R_k S_k^{-2} R_k^T, m_k; 1);\; k = i, j$

$x_j = f(x_i \mid E_i, E_j) = S_j R_j (R_i^{-1} S_i^{-1} x_i - m_i + m_j)$

$\left\| f(E_i, E_j) \right\|_2 = \max\left\{ f(z \mid E_i, E_j) :\; z \in \Re^p;\; \|z\|_2 = 1 \right\}$

$$s_{te}(E_i, E_j) = 1 \Big/ \max\left\{ \left(\left\| f(E_i, E_j) \right\|_2, \left\| f(E_j, E_i) \right\| \right)_2 \right\}$$

Theorem

s_{te} is a (strong) similarity measure on $E_u \times E_u$

Bhattacharya "Distance" $d_{bc}(E_i, E_j)$

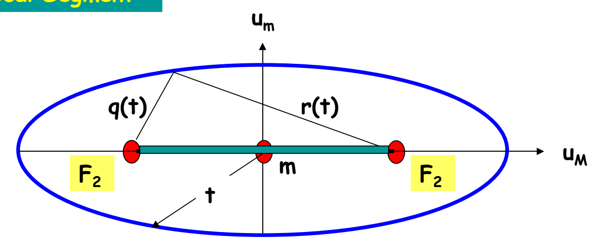
(1946) Bhattacharya distance between two p-variate Gaussians $X_1 \sim n(\mu_1, \Sigma_1)$ and $X_2 \sim n(\mu_2, \Sigma_2)$

$$d_{bc}(E_i, E_j) = e^{-\frac{1}{8}\left\| m_i - m_j \right\|^2_{\left[(A_i^{-1} + A_j^{-1})/2\right]^{-1}}} + \frac{1}{2} \ln\left[\det\left((A_i^{-1} + A_j^{-1})/2 \right) \Big/ \sqrt{\det A_i^{-1} \det A_j^{-1}} \right]$$

Focal Distance $d_{fd}(E_i, E_j)$

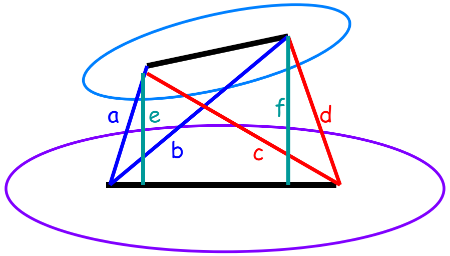
A has eigenvalues $\{\alpha_m, \alpha_M\}$ with eigenvectors $\{u_m, u_M\}$

Focal Segment



$$\text{Foci} \;:\; F_{1,2} = m \pm \frac{1}{2} \sqrt{\frac{(\alpha_M - \alpha_m)}{\alpha_M \alpha_m}}\, u_M$$

Focal Distance $d_{fd}(E_i, E_j)$



$$\delta_3 = \min\{a,\ c\} \mapsto e$$
$$\delta_4 = \min\{b,\ d\} \mapsto f$$

Compute default minimums

$$\delta_1 = \min\{a,\ b\}$$
$$\delta_2 = \min\{c,\ d\}$$
$$\delta_3 = \min\{a,\ c\}$$
$$\delta_4 = \min\{b,\ d\}$$

If OG projection of a foci *lands on opposite segment*, replace default minimum with this OG distance

$$d_{fd}(E_1, E_2) = \frac{\delta_1 + \delta_2 + \delta_3 + \delta_4}{4}$$

Focal Distance $d_{fd}(E_i, E_j)$

When p > 2

A_1 has eigenvalues $\{\alpha_1 = \alpha_m \le \ldots\ \alpha_k \le \alpha_{k+1} \ldots \le \alpha_p = \alpha_M\}$

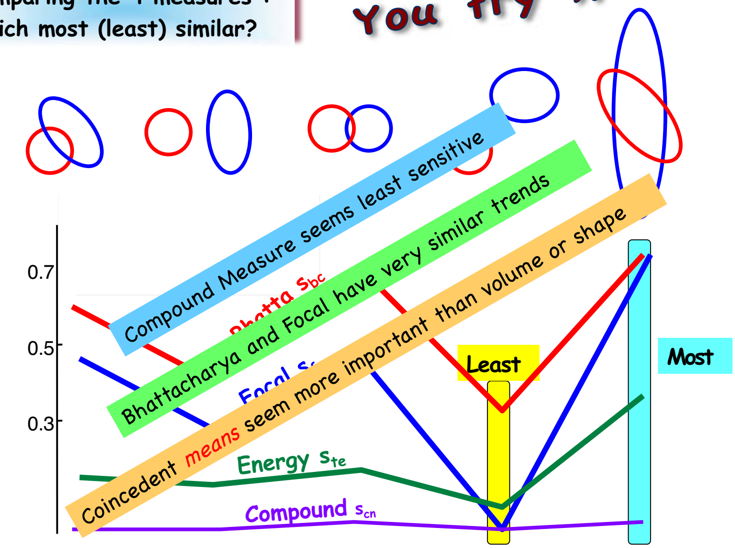
A_2 has eigenvalues $\{\beta_1 = \beta_m \le \ldots\ \beta_k \le \beta_{k+1} \ldots \le \beta_p = \beta_M\}$

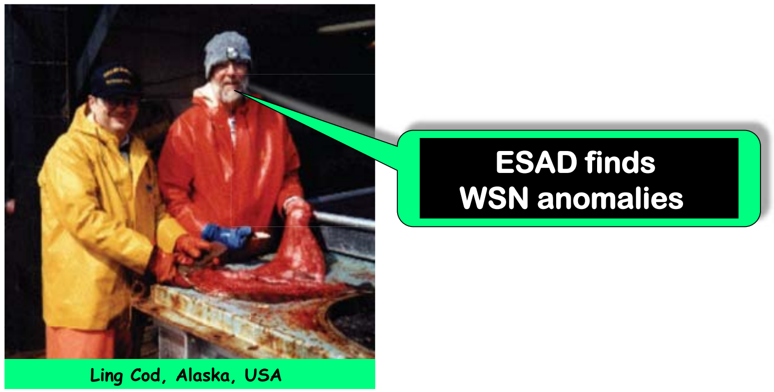
$$d_{fd,(k,k+1)}(E_1, E_2)$$

$$d_{fd}(E_1, E_2) = \frac{\sum_{k=1}^{p-1} d_{fd(k,k+1)}(E_1, E_2)}{p-1}$$

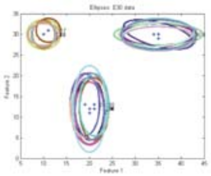
Comparing the 4 measures : Which most (least) similar?

You try it



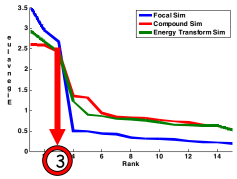


D_{focal} on Data E_{30}



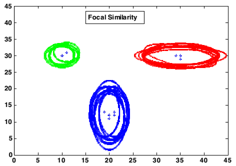
Pre-clustering estimates of c

Analytic: OEVs of D



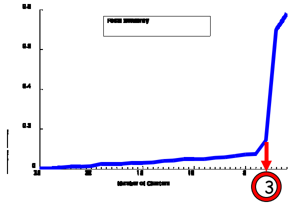
Visual: iVAT I(D'*)

SL 3-partition

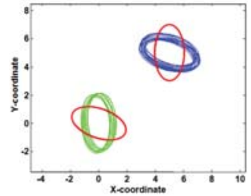


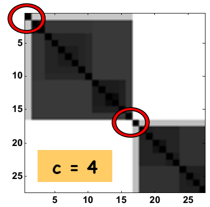
Post-clustering validation of U

Heuristic: SL merger distances

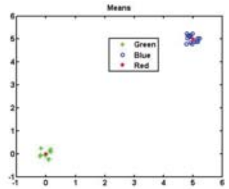


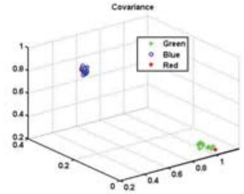
Data E_{27} = {10 green ∪1 red} ∪ {15 blue ∪1 red}





Would numerical clustering, FCM, e.g., reveal structure in E_{27} ?

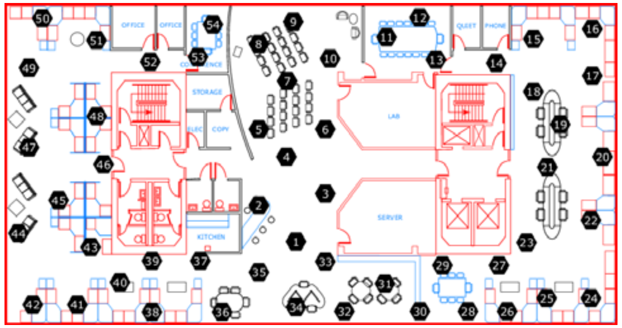




No. In these feature spaces FCM would optimize at c=2 !

Intel Berkeley Research Lab (IBRL) monitors lab conditions

55 Sensor nodes deployed at UC Berkeley, March 2004



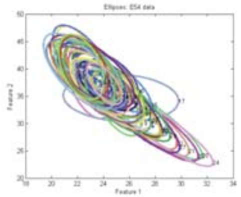
2D Data

Collection Period = 18 days, March 1-18, '09

Temperature
Humidity

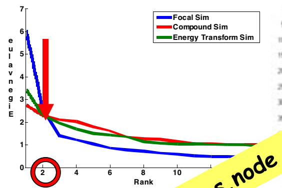
1 summary ellipse built for each IBRL node

D_{focal} on IBRL Data E_{54}

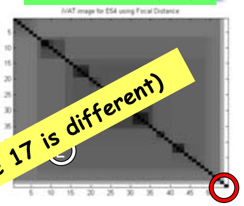


Pre-clustering estimates of c

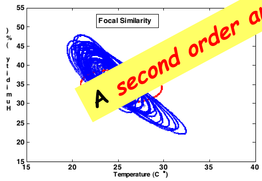
Analytic: OEVs of D



Visual: iVAT I(D'*)

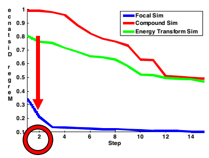


SL 2-partition



Post-clustering validation of U

Heuristic: SL merger distances



A second order anomaly (in the 54 ellipses,node 17 is different)

Great Barrier Reef Ocean Observation System (GBROOS) to monitor reef environment/sea habitat

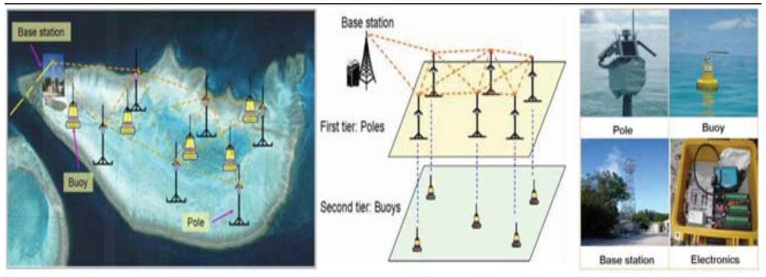


Figure1(a) Figure1(b) Figure1(c)

Land/Sea based Sensor nodes deployed at Heron Island,2007

Heron Island (GBROOS, Australia) Weather Station Data

3D Data {Temperature, Humidity, (Air) Pressure}

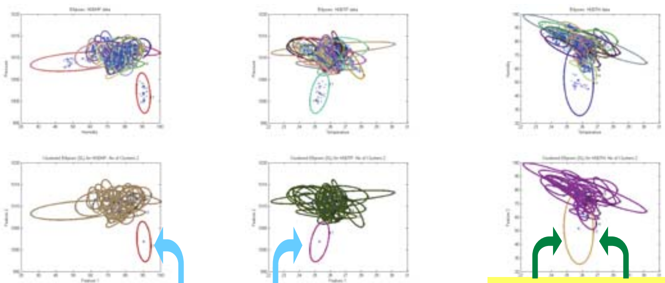
Collection Period = 30 days, Feb. 21 to March 22, 2009

Samples = 6 hrs x 6 per/hr = 37 vectors ⇒ 1 ellipsoid

Cyclone Hamish occurred on March 9, 2009, ellipse #17

A *first order anomaly* (in the 30 ellipses, one is different)

HI30(H,P) HI30(T,P) HI30(T,H)

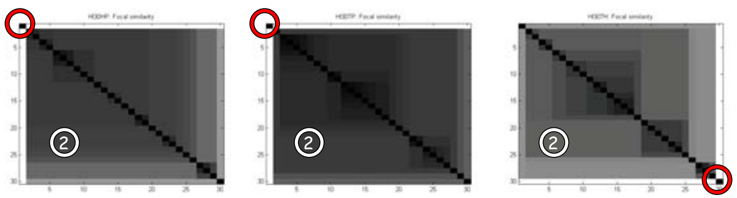


SL clusters

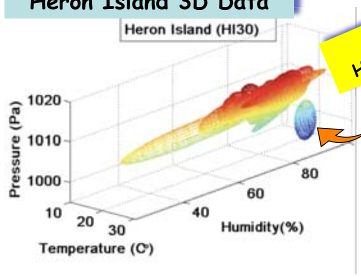
1st order Anomaly #17 (Hamish)

1st order #13 *not* Hamish (3/5, *no P*)

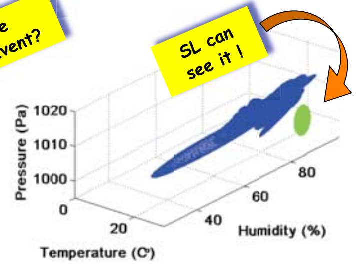
D_{focal} iVAT images



Heron Island 3D Data



See the Hamish Event?



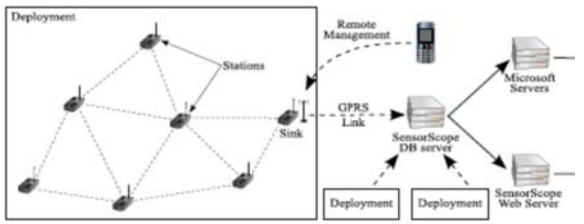
SL can see it !

D_{focal} iVAT D_{Bhatta} iVAT $D_{compound}$ iVAT D_{Energy} iVAT

Focal and Bhatta Can see it !

Compound and Energy Cannot !

Grand St. Bernard GSB) WSN



WSN configuration

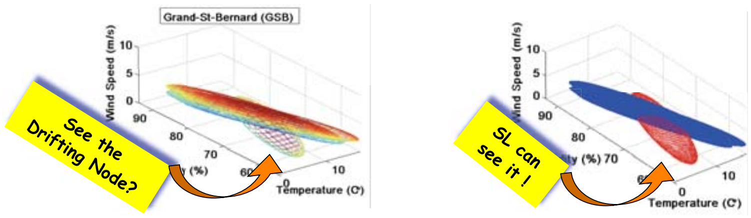
Sensor

17 stations in a mountain pass at Wannengrat, Switzerland.

3D Data

Temperature
Humidity
Windspeed

The GSB Data has a *2nd Order* Anomaly : (Drifting Node)

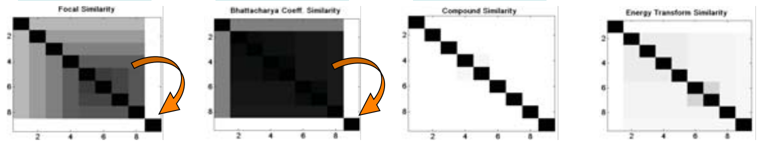


D_{focal} iVAT

D_{Bhatta} iVAT

$D_{compound}$ iVAT

D_{Energy} iVAT



Focal and Bhatta can see it !

Compound and Energy cannot !

Conclusions

Three types of WSN anomalies

1st order (1/E) = internal node faults

2nd order = whole node faults

Higher order = subtree of nodes faults

Two models for detection that use data-based ellipsoids

DCAD model uses level sets of ellipsoids to capture data

ESAD model uses visual assessment of similar ellipsoids

Four measures of (dis)similarity for ellipsoids

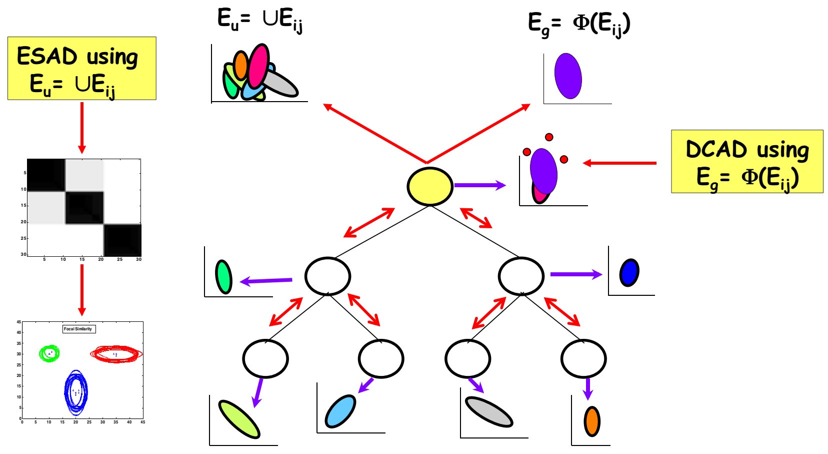
Compound normalized

Transformation energy

Bhattacharya Distance

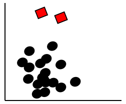
Focal Distance

Comparing the DCAD and ESAD Models



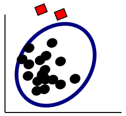
Local ellipses $\{E_{ij}\}$

Comparing the DCAD and ESAD Models



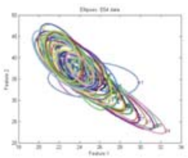
DCAD : Detection in *data space*

Cardinality | # sigmas | χ^2

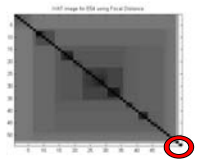


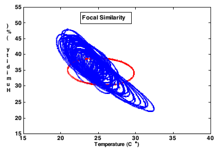
ESAD : Detection in *ellipsoid space*

Ellipsoidal summaries → Similarity measures → iVAT(D'*) estimates c → SL clustering for detection



Bhatta
Focal





Wakker worden

Het is voorbij !