~~Quo Vadis~~
που πάτε, PSE Data Streams?

Michael Baldea – The University of Texas at Austin
Victor Zavala – University of Wisconsin, Madison

FIPSE -5, June 27, 2022

The Hunt for Red October (1990)

Data Insights?

Data Insights?

- Various **data sources** (sound, sonar)
- Chasing an elusive target: not clear what **actionable information** is available
- **Data analytics based on ML fails**: the computer runs “SAPS” - signal algorithmic processing system - and reports a seismic anomaly
- **Human skills** intervene in **data representation**: 10x the speed reveals a pattern
- Use sound data and map to infer trajectory of enemy submarine

Data Insights?

Most fields face the same problem today

- **Acquire:** new data types
- **Interpret:** provide actionable information
- **Act:** respond

- Leveraging **human knowledge** and skill is essential

Agenda for today: Recent and emerging data types do not fit neatly into the PSE format
– challenges and opportunities

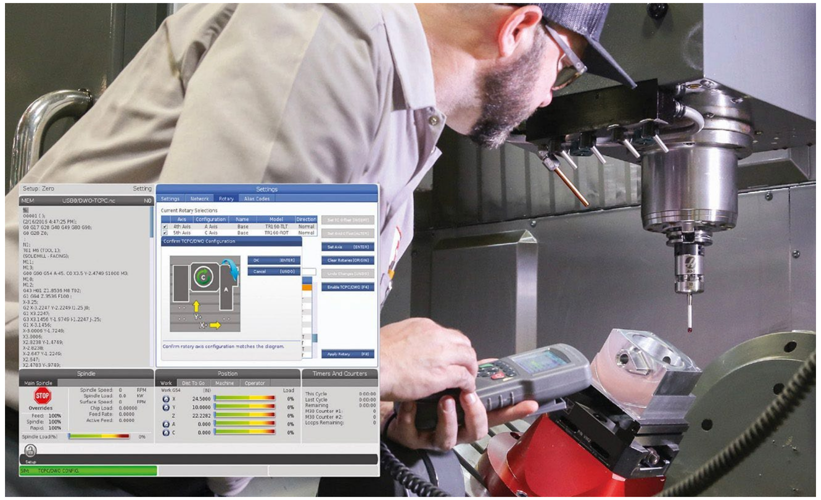
A Look Around Us…

- 1948 Ford F-series

- 2022 Ford F-150 Lightning

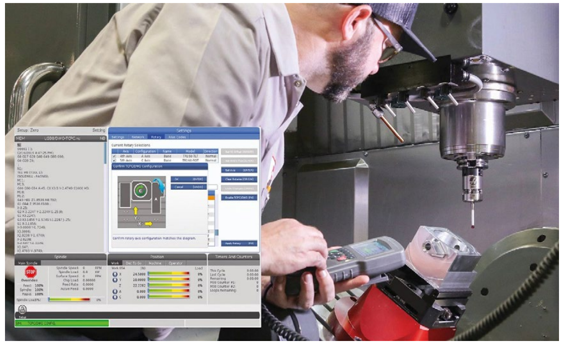
- “Lathe Workers” by F. Eichhorst (1948)

- 2022 Haas CNC 5-axis mill



Sources: Wikipedia, ford.com, haascnc.com



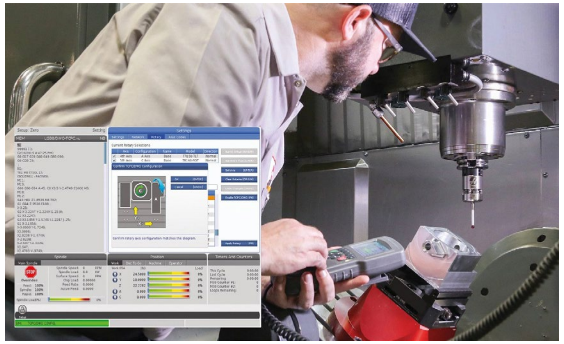



Progress since WWII

- **Developed low-cost sensors** for **many types of variables**
  - Chemical
  - Mechanical
  - Position/distance

- Integrated these sensors in **meaningful standardized systems**
  - Environmental
  - Safety
  - Communication
  - Leveraged real-time processing to implement in practice



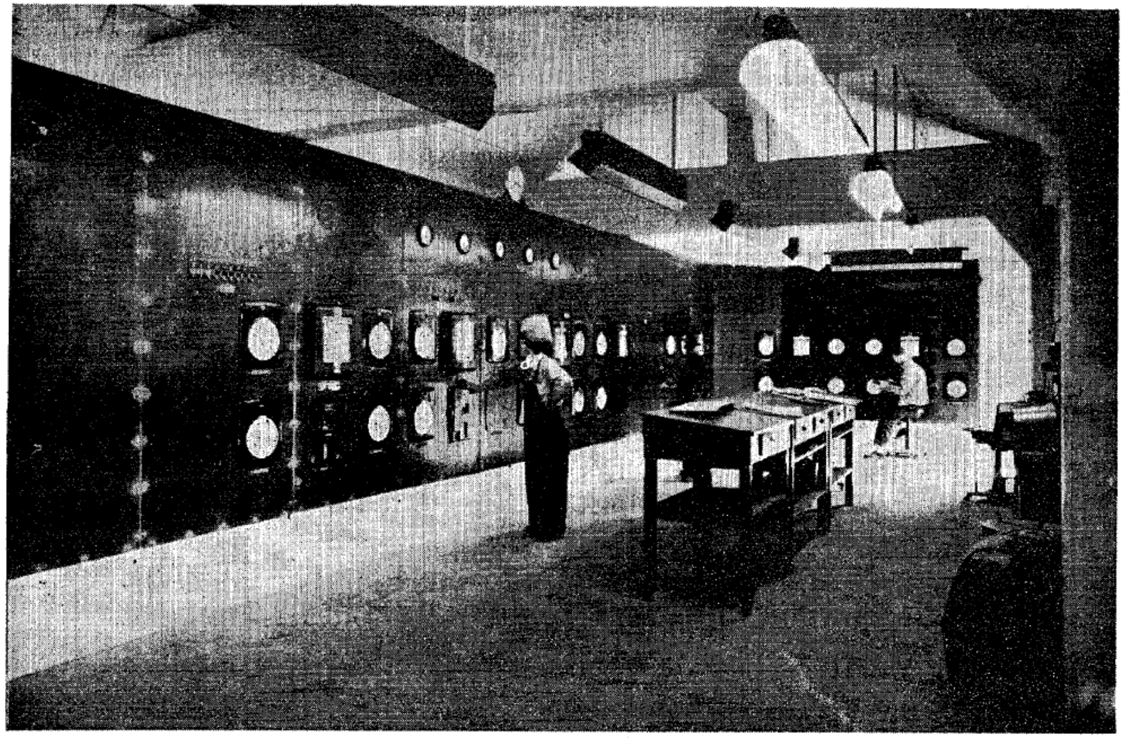



- **Massively scaled and standardized production to reduce cost**
  - A few product lines/ small number of versatile platforms
  - Profit comes from volume and margin

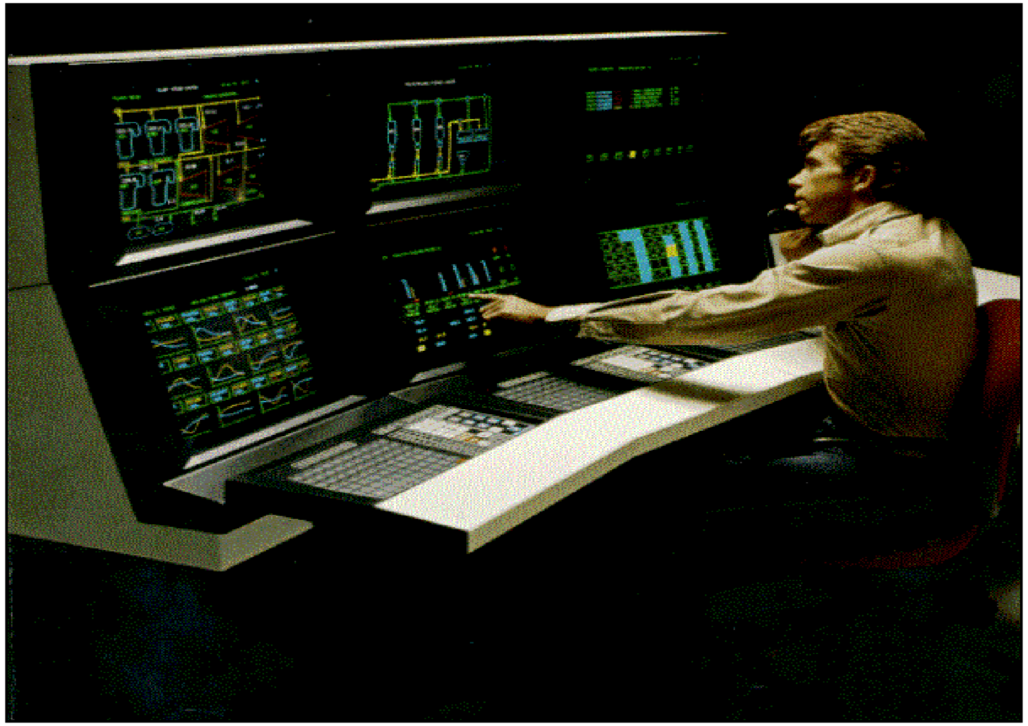
- **Accelerated the development cycle**
  - Massive R&D investment in house
  - New models marketed every 5-8 years
  - Can afford some failures/flops

Chemical Industry

Control Panel of a Thermofor Catalytic Cracking Plant



12 instruments
*F.J. Van Antwerpen, Ind. Eng. Chem.*1944,368,694-698



- Circa 1990s (?) – taken from Tom Edgar’s lecture slides

Advances in data acquisition:

- More reliable and accurate sensors, equipment
- Much higher sampling rates, wireless transmission, data storage

Stagnation:

- Largely measuring the same variables (P,T,F,x): time series of scalars
- Sensors still expensive (+installation)
- Key variables (e.g., heat rate) remain unmeasured, others (e.g., concentration/composition) are difficult to measure

Advances:
- Soft sensors
- Massive historian databases
- Improved visualization
- Data analytics

Stagnation:
- Results are not always easily interpretable (PCA)
- **GIGA (garbage in, garbage out):** still the same sensors
- **We still cannot close plant material and energy balance**

Advances (real-time):
- MPC revolution + System identification
- Real time optimization
- Fault detection and isolation

Stagnation:
- MPC is still MPC
- RTO not broadly adopted
- One-off approach for implementation

Advances in process operations (off-line):
- Algorithmic developments in planning and scheduling
- PSE community leading the way in MILP (e.g., DICOPT), NLP (IPOPT), global optimization (BARON, Antigone), and modeling environments (gPROMS, Pyomo)

Stagnation:
- Open-loop decision making
- Exploit new architectures (parallel, GPU, embedded, edge computing)

Advances in process design (off-line):
- Process simulation now common
- Data reconciliation

Stagnation:
- Still “optimizing” by hand
- Connect process and molecular design
- Open-loop decision making: implementing lessons learned **from data** in new designs?
  - Designs cycles are long, many plants are one-off

Key Strengths of the PSE Community

Fundamental:

- Deal with complex, large scale, nonlinear, distributed systems
- Real-time and off-line
- Account for uncertainty
- Create abstractions

Applications:

- Enormous scales: physical, temporal, financial
- Diversity of products
- Good safety record

Some Key Challenges

- Scalability: many plants are still one-of-a-kind
  - **DATA perspective**: reduced opportunities for transfer learning
- Failure is not an option
  - **DATA perspective**: not enough data for training ML models
- Training and thinking
  - Channeled by unit operations and input and output variables that are scalars
  - **DATA perspective**: new data streams may not fit this pattern

Potential New Data Sources

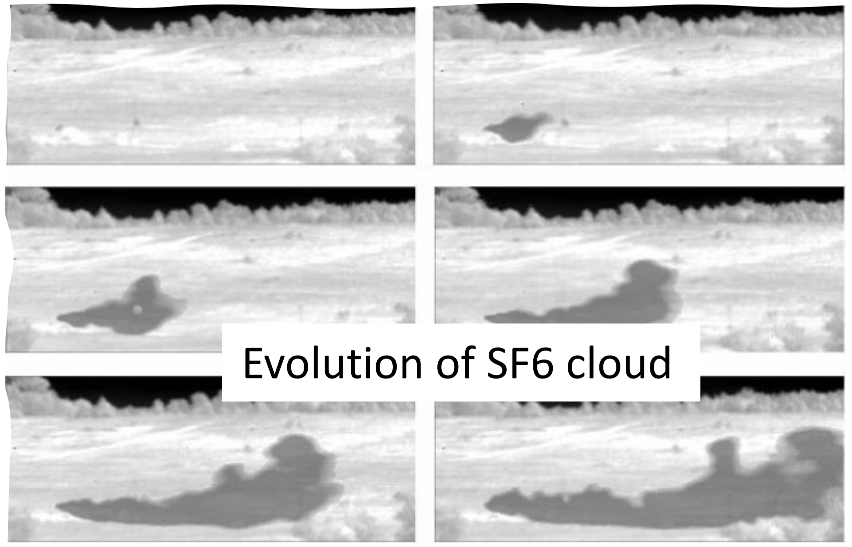
Imaging:

Hyperspectral: combine e.g., IR/Vis/UV

Chemical imaging: include spectrometry

Cover large areas

Data format: data hypercube



Evolution of SF6 cloud

Sources: Wikipedia, Chamberland et al. SPIE 2005

Potential New Data Sources

Vibration:

Acoustic imaging combines visible spectrum imaging with arrays of microphones to detect sources of sound

Data format: data hypercube

https://www.flir.com/products/si124-ld/?vertical=condition%20monitoring&segment=solutions

Potential New Data Sources

Replicate human senses:

“electronic nose” detects chemical signatures via sensor array

Data format: spectra, generally in processed format

https://www.sensigent.com/products/**cyranose**.html

https://www.norlab.com/infocenter/207

Potential New Data Sources

Disposable sensors:

The norm in medical/food applications

Multivariable sensing (T, pH, dissolved O2, …)

Use to ensure uniform processing experience for every molecule

Data format: time series for a limited time

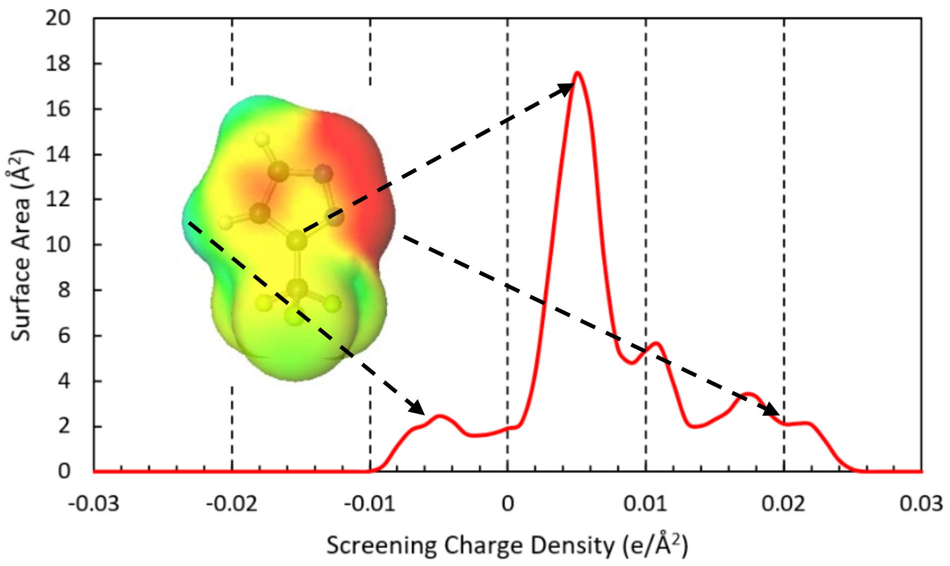
https://www.chemicalprocessing.com/articles/2011/disposable-equipment-sensors-pass-litmus-test/

Potential New Data Sources

Not all new data come from the field:
Sigma profile: probability distribution of charge density for a molecule

Generated a-priori using quantum mechanics calculations

Data format: spectrum



Mullins et al., *Ind. Eng. Chem. Res.* 2008, 47, 5, 1707–1725

Source: Brennecke group at UT

Potential New Data Sources

Stories:

Recording/videotaping experienced operators

Operator logs

Data format: audio/video recordings of natural language

Source: toppng.com

Some Trends

Data are complex mathematical structures

- Not a time series of scalar values
- Not easy to visualize or interpret

Not directly actionable

- Not a good objective function
- Condense to scalar value without loss of information

Extract Information <> Dimensionality Reduction

PCA/PLS

Convolutional neural networks: extract geometrical parameters

Recurrent neural networks: model dynamics

Dynamic mode decomposition: dynamic models from sequence of tensor data

Topological/geometric tools: projection on lower-dimensional manifolds

(geometric deep learning)

Unresolved Challenges: Interpretation

Resulting models are not generalizable

- Success in other areas (e.g., automotive) due to standardization

Not easily interpretable: (PCA – what do the components "mean")

Scalability – training requires lots of data

Unresolved Challenges: Action

Assume data condensed in meaningful way

Absorb into MPC structures

Can we get the right models? Step tests? What is a “step change” here?

Build (another!) NN to embed in the MPC - need EVEN MORE data

Switch to other control paradigm: reinforcement learning – often not an option.

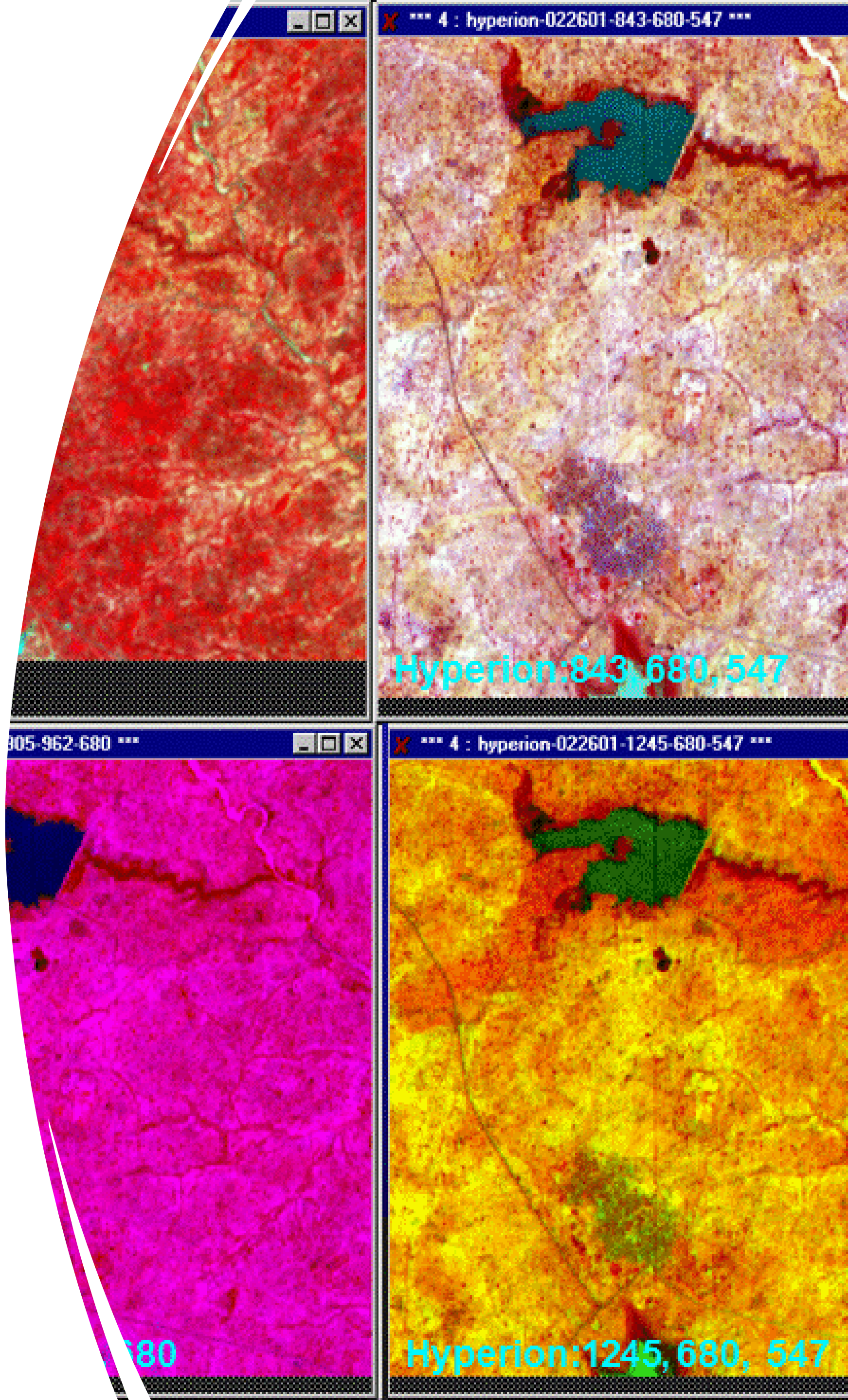
Agriculture

Problem:
- Control moisture content in soil, nutrient distribution, etc. to maximize production

Strategy:
- Predictive control of irrigation, crop dusting, fertilization

Data format:
- Hyperspectral imaging

Impact:
- Enormous, given current and future geopolitical situation



https://www.usgs.gov/media/images/earth-observing-1-eo-1-hyperion-hyperspectral-data

Agriculture

Problem:
- Production of value-added products (biogas, biofuel, fertilizers) from distributed sources (plastic waste, renewable power, food waste, wastewater)

Strategy:
- Learning-based control for autonomous modular production units (with information sharing)

Data format:
- Process data (decentralized)

Impact:
- Enormous: currently most plastic waste and livestock waste is not recycled.
- Many places around the world do not have access to fertilizer.

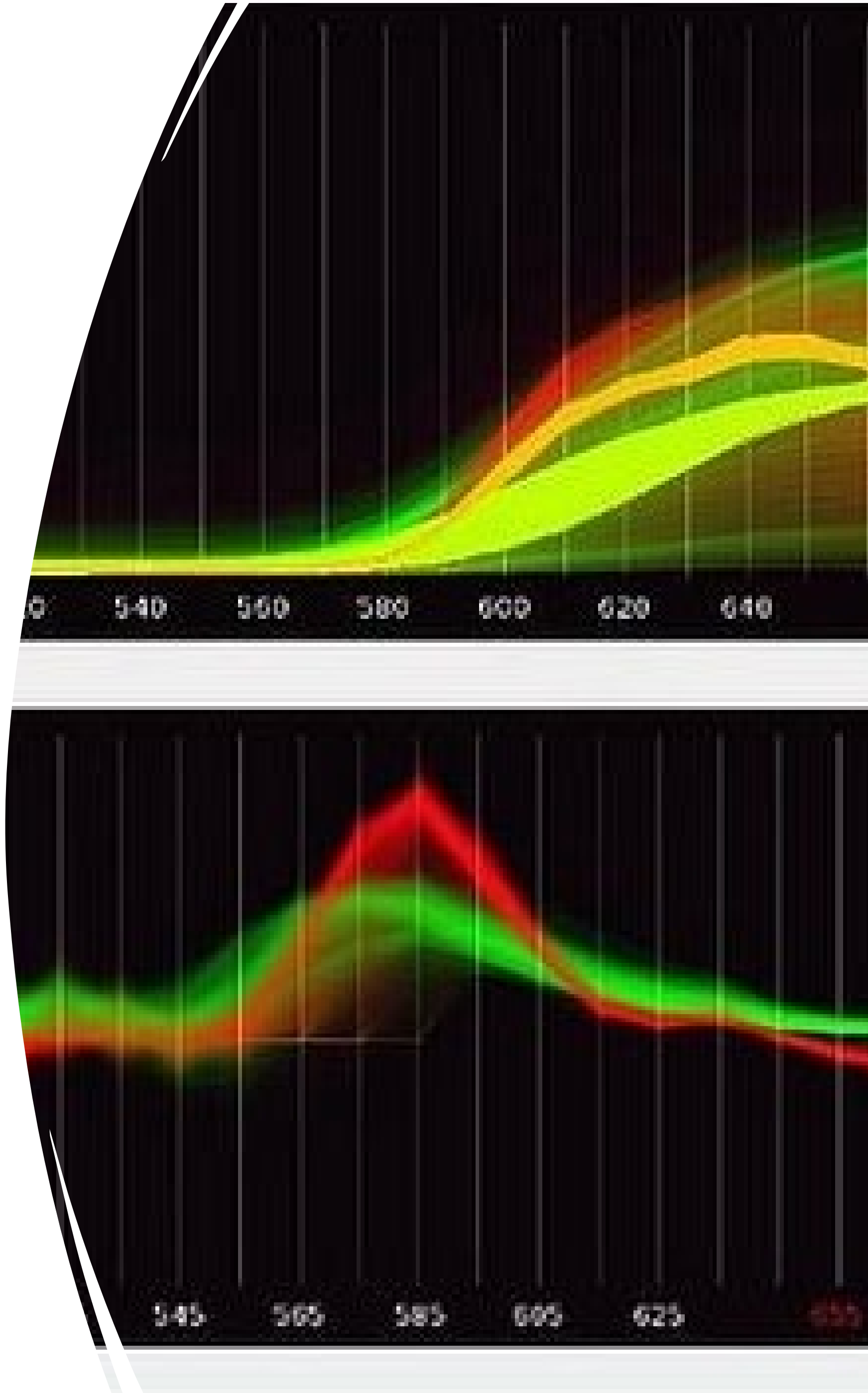
Food

Problem:
- Minimize food waste in supply chain of perishable products

Strategy:
- Feedback-based supply chain management
- Model degradation rates
- Manipulate storage/shipping temperature

Data format:
- Hyperspectral imaging, disposable sensors

Impact:
- Enormous: about 30% of food goes to waste in the supply chain while 800 million people experience food insecurity



https://www.postharvest.biz/en/news/hyperspectral-imaging-potential-to-detect-contaminants-chemometrics-bacteria-/_id:79807/

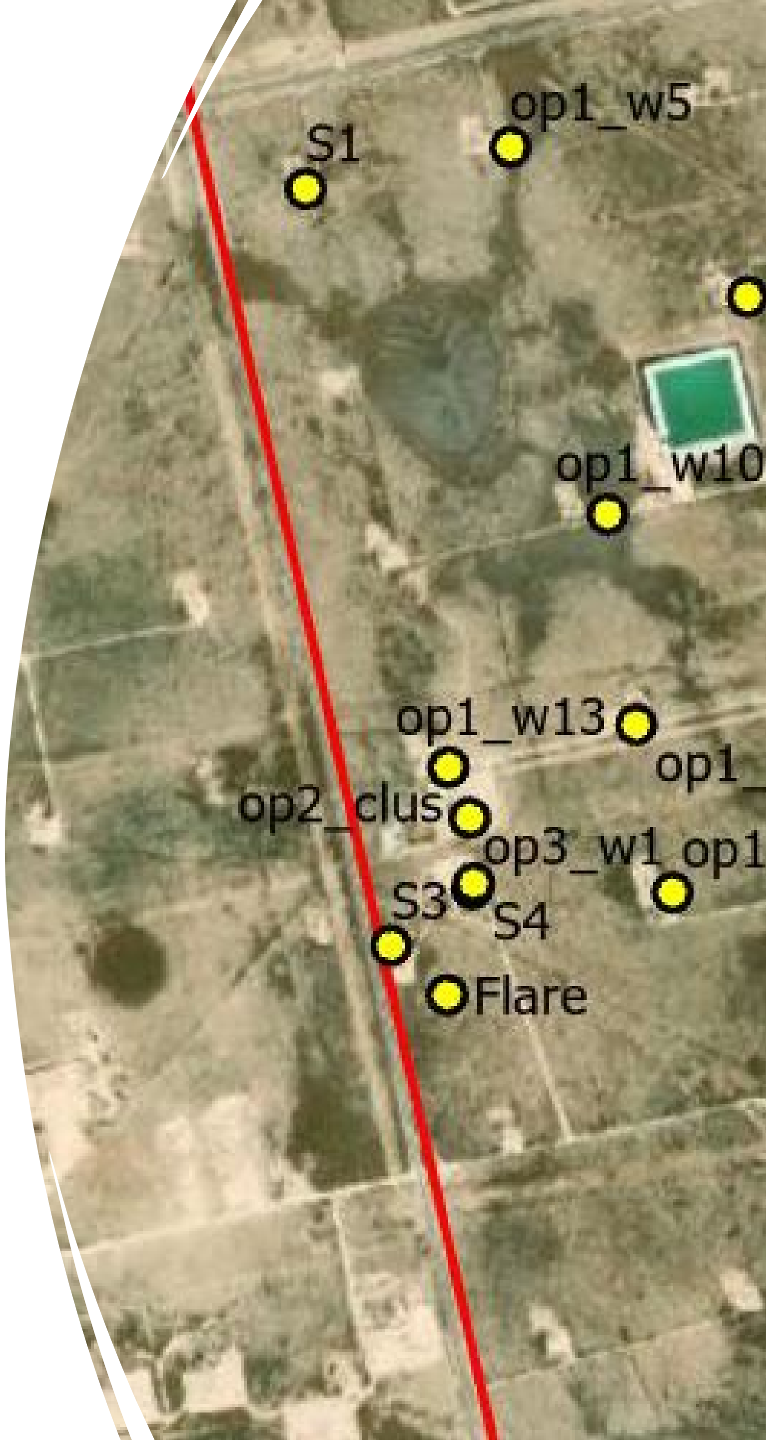
Environment

Problem:
- Minimize greenhouse gas (methane) emissions from production sites

Strategy:
- Detection at sensor location
- Trace back to source
- Dispatch service to mitigate

Data format:
- Chemical imaging
- Atmospheric dispersion models
- Weather data

Impact:
- Enormous: 200,000 sites in Permian Basin, TX alone. CH4 is 25x worse than CO2 as a GHG



Data: epa.gov. Image source: Mrinali Modi, Allen group at UT

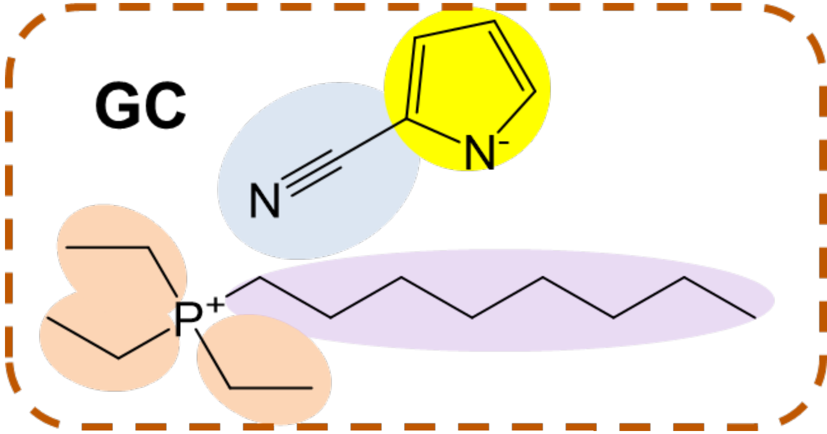
Environment

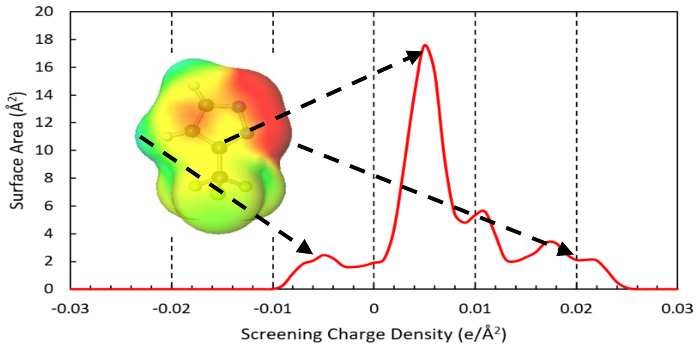
Problem:
- Simultaneous design of solvents and processes for CO2 capture

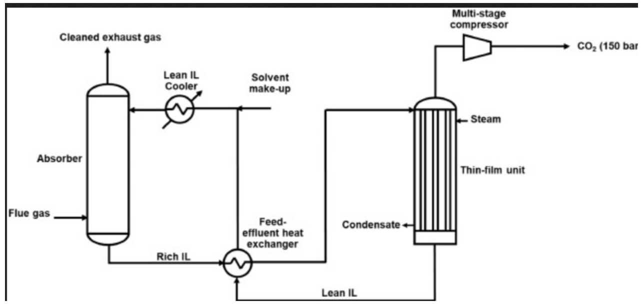
Strategy:
- Multiscale modeling, including QSPR
- Flowsheet optimization

Data format:
- Sigma profiles + experimental macro properties
- Power plant data

Impact:
- High. Power plants are not going away in the energy transition…







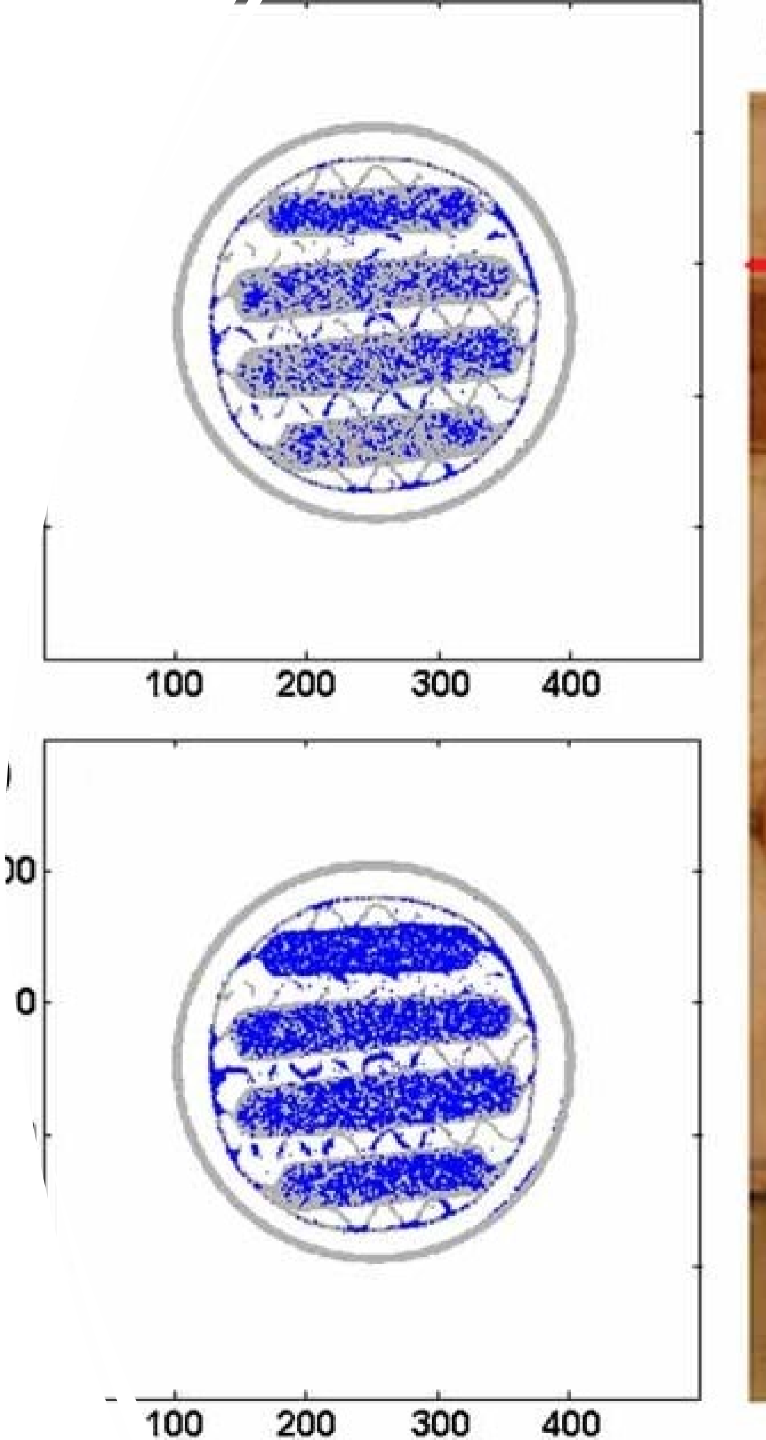
Operations

Problem:
- Electrification of process industry
- Ensure every molecule experiences same processing conditions, particularly in transient operation

Strategy:
- REAL-TIME Imaging unit operations using sound, capacitance tomography, etc.

Data format:
- Data hypercube

Impact:
- Very high – we are treading uncharted waters



Aferka et al., *Canadian J. Chem. Eng.* 88.4 (2010): 611-617

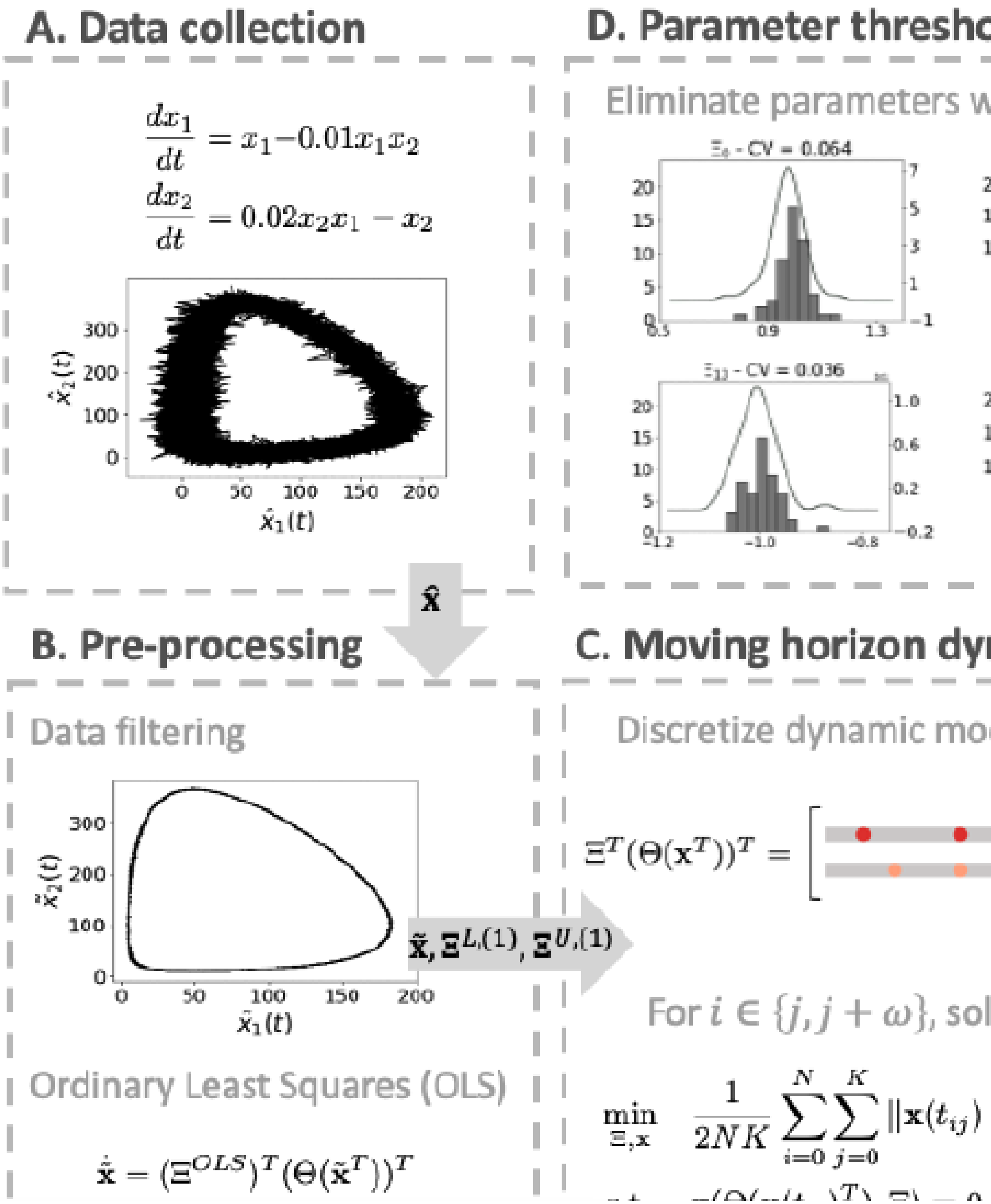
Fundamentals

Problem:
- Discover governing equations for processes/phenomena

Strategy:
- Collect experimental data
- Build models using sparse learning

Data format:
- Data hypercube

Impact:
- Very high – accelerate science



New data sources may not fit familiar data formats

Extract key information

Formulate design, control problems

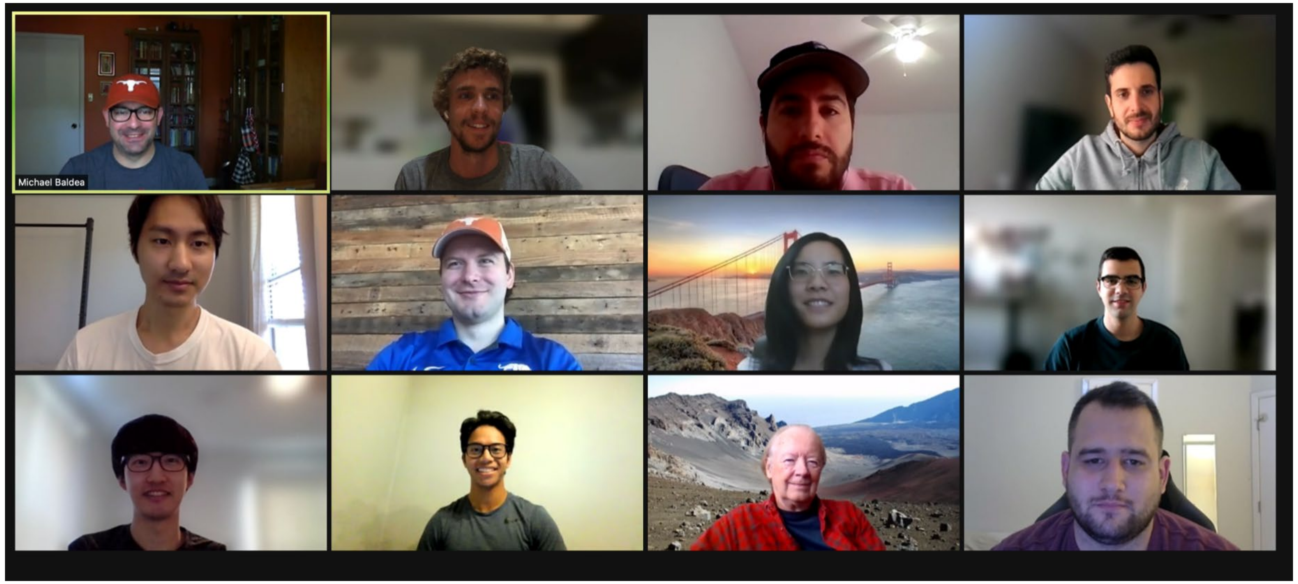
Can have enormous impact

Acknowledgments: MB

Students: R.C. Pattison, C. Tsay, F. Lejarza, K. Seo, A. Retnanto, X. Tang, J. Martorell

Collaborators: T.F. Edgar, M.A. Stadtherr, J.F. Brennecke, R. Baldick, S. Wright, G.T. Rochelle

Sponsors: NSF, DOE-NETL, ARPA-E, UT Energy Institute, ExxonMobil, Shell, ABB



Old slides

New applications

- Agriculture: control moisture content in the soil, distribution of fertilizer, etc…

[efforts in precision agriculture are underway, led by electrical engineers…]
Production of value added products from waste (biogas, biofuel, ammonia?)

- Supply chain of perishable products

Potential New Data Sources

- Vibration (in audible spectrum, but also higher/lower frequencies)
- Imaging (hyperspectral – IR/Vis/UV), three dimensional (see e.g. 3d printing)
- Analytics
  - Spectrometry (IR/UV/Mass/…)
  - Dimensional/geometry (crystal shape, length, etc.)
  - Replicate human senses (electronic nose, tongue, tasting panels…)
- Cheap disposable sensors (move with process stream?) – in the medical field all the sensors are disposable

What exactly does a controller for a chemical process have to do

- Number one mission is SAFETY (same as in automotive, btw)
  - Cannot afford a single slip-up
- Emissions: has been a concern, will become more prominent due to growing environmental concerns
- Product quality: sometimes substandard product can be “blended away” but not always
- Economics (what to make)…
- Other things: maintenance/predictive maintenance

We already do these things… what could we do to do them better, given the technology that we recounted earlier

To start: What are We NOT Good At? Vs other fields

- We cannot fail:
  - Difficult to do things in a sandbox like power systems, automotive, robotics
  - Need to try stuff on the real deal: experiments in a $10bn chemical complex are frowned upon, to say the least: cannot afford iterative learning, model updating etc.
  - Note that this $10bn facility is likely like no other, so nothing that is learned here transfers directly to somewhere else
  - We don’t have good data for failure: want these to not happen at all
  - While unit ops went a long way in adding some structure to our thought process, it is still a far cry from the level of standardization that other fields have (e.g., power systems).
  - We also have an immense diversity of feedstock and products, so standardization is difficult
  - We are process oriented (i.e., the incarnation) rather than technology oriented (i.e., the concept).

The next frontier: - theoryAutonomy?

- The Plant decides what to do (ahahahaha)
  - What could it decide: production rate, product grade, what feedstock to run and when (cannot give up on safety and environmental)
  - Basically this is a supply chain problem, so should we move control concepts into the supply chain
    - We have already… (work on MPC applied to scheduling/planning)
  - HOWEVER, it could be interesting to add more sensing to the supply chain (per the sensors we discussed earlier).
    - For example, consider the supply chain of perishable products (food, vaccines, fresh produce, meat, biologics such as transplant organs). The characteristic is that these products change (actually, unfortunately, degrade) as they move trough the supply chain, with the possibility that they decay beyond being safe to use BEFORE they reach their end-user.
    - Low-cost disposable sensors; could travel with the product and be discarded.
    - Incorporate monitoring of quality variables (freshness/degradation/rottenness) with imaging

Capturing stories? – in the wake of retiring operators

- Text processing, natural language processing
  - What I tell the computer is often not what I WANT it to do – miscommunication
  - On the other hand, a practical and pressing need is to capture the knowledge of an ageing base of experienced operators
    - Many companies are videotaping them as they relate their experiences
    - Not sure what came of this, but turning the lore into a manual for younger personnel is of paramount value.
    - Machine summarizing a text – doesn't exist yet.
    - Automated interpretation of data – can a machine look at data and provide an explanation of what is happening in the plant – this has major safety implications since it would provide operators a means to react in real time. Expert system – we still don't have them.

    *as a side note- accidents in the chemical industry happen because operators have not responded fast enough to an incident. It is not fair to expect them to do so, given that investigation boards (who are not under pressure) often take MONTHS to reach a conclusion.

Autonomous discovery: Community of automated/autonomous molecule synthesis

- Challenge is linking structure to properties in a predictive way, closing the loop on making new structures with desired properties; bypass expensive instrumentation
  - Electrical engineers cannot do this
- Natural language processing, visualization of molecules
  - Manipulating complex structures with constraints (steric…)
  - Optimization of these structures
  - Build mini chemical plants, flow chemistry
  - Cannot characterize molecules in real time, but they can identify measurable variables that they correlate to structure.
  - Scheduling robots to do work in the lab (Chris Rao)?
  - Scheduling drones for surveillance/monitoring?
- Data centers:
  - Distributed data centers linked via private network – kind of like distributed manufacturing… not sure what we can learn from them, since they have infinitely scalable capacity and the cost to shift data is much lower than the cost of shifting "Stuff"

New objective functions:

- Minimize waste through the supply chain (or in production in general)
- There could be secondary benefits in environmental metrics (emissions, toxicity – health benefits …)

- New collaboration paradigms: many supply chains are not vertically integrated (particularly food)
  - How do we communicate?
  - What does the picture of a rotten banana in a supermarket in Heraklion mean to a grower in Colombia?

  - Sensors aligned with human perception (vision, taste, touch…) – how can sensations be translated to control objectives.
    - A fruit “seems ripe” to us – what does that mean?

How do we process this information?

- Sensors aligned with human perception (vision, taste, touch…) – how can sensations be translated to control objectives.
  - A fruit “seems ripe” to us – what does that mean?
  - Classification?

  - How does the brain process this info to translate into some action?
    - Learned experiences + current sensation = action
    - “As soon as I smell beer, I start salivating”
    - Information storage and classification – we probably do that better than the current computing systems.
  - How does the brain process color/odor? We can learn from biological systems
    - The first things that brain does when it sees color is to decompose the image (not RGB, but in 3 color spaces that are orthogonal – allows brain to extract much more info than from RGB)

Leverage cloud computing

- Use mixtures of computers to process data on the edge or on the cloud
- Say we have 100 cameras collecting data – how can this be structured for processing, must do some processing locally
- Federated learning
- Broader question is how do deal with high frequency high volume data from distributed sources. Where do we process them.

- Embedded computing platforms – automotive industry is already doing this. What can we do with them?