Intelligent Data – Model Fusion

Application of Generalized Tikhonov Regularization To Earth Ecosystem Data-Model Fusion

P.I.: Dr. David E. Thompson / NASA Ames, Code IC

dethompson@mail.arc.nasa.gov

Co-I's: Dr. Andrew J. Meade / Rice University

meade@rice.edu

Dr. Rajkumar Thirumalainambi / SAIC, NASA Ames

rajkumar@mail.arc.nasa.gov

Mr. Roy A. Britten / QSS Group, NASA Ames

rbritten@mail.arc.nasa.gov

Ms. Alex Kanalakis / SAIC [NASA Ames] and Moss Landing Marine Lab

kanalaki@email.arc.nasa.gov & *Akanalakis@mlml.calstate.edu*

Data-Model Fusion

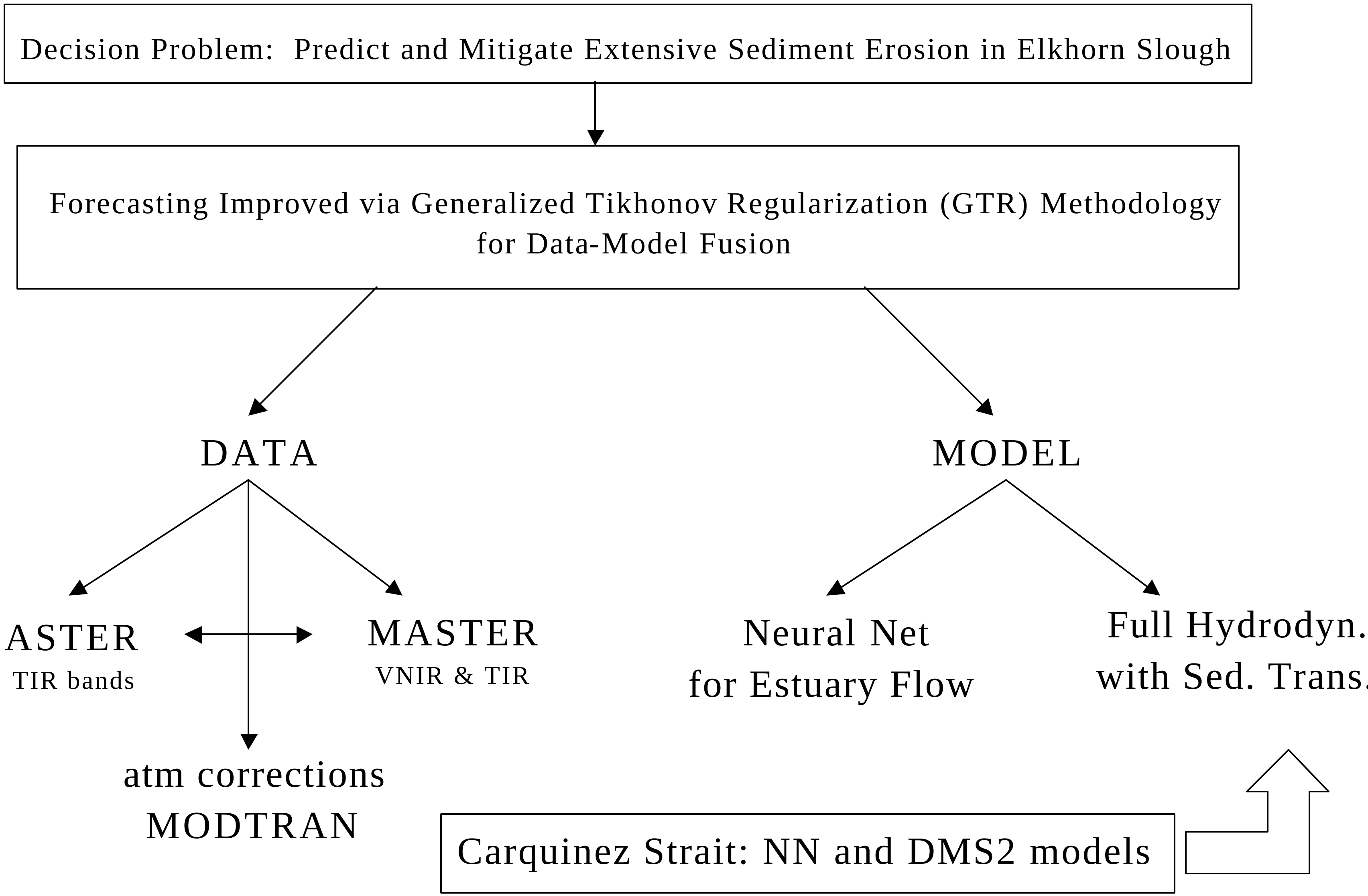
- GOALS:
  - To substantially improve the forecasting capability of Earth science process models through formal fusing of the models with diverse datasets as optimizing constraints

  - To discover / validate the appropriate fidelity of process models that can be supported by resolution of remote sensing datasets and ground-truth data

Data-Model Fusion

- OBJECTIVES:
  - Create a generalized regularization framework that fuses ecosystem process models with remote sensing datasets for data completion, model fidelity, and error analysis [**Technology Transfer from Aerospace Heritage**]
    - Necessary for this : Modify / deploy the DARWIN data-mining infrastructure to operate across ecosystem process models and remote sensing datasets, fused with COTS GIS software and custom-built neural nets

Data-Model Fusion: Project Strategy



Data-Model Fusion: Technical Problem Statement

Technical Innovation

- An innovative framework for data-model fusion is being developed, based on Generalized Tikhonov Regularization (GTR), that provides a rational means of combining theoretical models, computational data, and experimental measurements into a global representation of a domain.

- The methodology, developed for ill-posed problems, can be used for data-driven extrapolation and fine-tuning of computational analyses, physics-based interpolation of sparse data, correction of low-fidelity models based on measured observations, intelligent detection of noisy data, process-model discovery, and uncertainty management.

- The GTR foundation has been developed and validated within the Computational Fluid Dynamics domain, wherein deterministic physics-based models and data are available. This project transitions this methodology to the Earth System Sciences domain, where the development will be extended to handle the issues related to heuristic and statistical modeling, proxy data, and process-model discovery.

Data-Model Fusion: Data Sources

- ASTER TIR Data Products over Elkhorn Slough environ:
  - surface radiances [atm corrected] and sky irradiances [AST09],
  - surface emissivity [AST05],
  - surface kinetic temperature [AST08]
- MASTER Data for “same” site and time, all 50 channels, covering both Slough and the sediment plume in the Bay
- Collected *in situ* samples from Moss Landing Marine Lab’s tide gauges, sediment flux meters, bathymetry, and toxins, monitored at stations within the Elkhorn Slough
- Collected radiometer measurements for salt marshes, pickleweed, and land environment bordering the Elkhorn Slough Reserve -- these support the image classifications and provide proxy constraints on sediment bank erosion

Data-Model Fusion: Technical Approach

- Radiative Transfer and Temperature-Emissivity Separation:
  - Develop regularization framework for Thermal IR emissivities
  - Acquire ASTER scene of TIR surface radiances with atmospheric correction applied [AST09 data product] --Elkhorn Slough site
  - Acquire ASTER products of surface emissivity [AST05] and kinetic temperature [AST08] for this scene [i.e., TES algorithm already applied]
  - Acquire MASTER data product for this same site and time [ours are 1 day apart]
  - Run MODTRAN and create atmospheric correction for MASTER data, and benchmark TIR bands between ASTER and MASTER
  - Use GTR framework to create TES-type data product for MASTER data; interpolate via GTR across bad or cloudy pixels
  - Reproducing a known result initiates the data-driven evaluation of the flexibility, complexity, and error capability of the GTR method

Data-Model Fusion: Technical Approach, continued

- Run 5 “Degraded” Numerical Experiments:
  - <partial information>: re-discover a Region of Interest in the Elkhorn Slough scene using the emissivity distribution
  - <forecasting>: use time-change in ancillary climate/weather data to predict changes in the kinetic temperature distribution and compare with equivalent MASTER data [refined emissivity]
  - <proxy information>: explore how a sediment transport model and a neural net flow model changes the MASTER classification based on emissivities; experiment with different classification methods
  - <model discovery>: discover the appropriate format of emissivity representation as the regularizing functional in GTR; do variation of parameters on the regularizing parameter λ, and contour this
  - <heuristic/statistical>: identify the appropriate model fidelity for statistical-correlational models that the remote sensing community uses now, or improve on the fidelity of these

Data-Model Fusion: Technical Approach, continued

- Moss Landing Marine Lab (MLML) Collaboration:
  - The purpose of the collaboration is (1) to model the dynamics of the Elkhorn Slough, a 9-mile long estuary that is producing substantial sediment flux into Monterey Bay; and (2) to predict the sediment flux and plume distribution and erosion as functions of time, using neural nets and GTR, with ASTER and airborne MASTER data as constraints
  - *In Situ* data sets include MLML tide gauges, sediment flux meters, bathymetry, and toxins, as point datasets, that supplement the remote sensing constraints and initialize the neural net
  - Use GTR to create and refine a Neural Net model of the water and sediment flux, balancing the RS data with this model
  - Use the GTR results and NN flux model to build and adapt a fluid mechanics and full sediment flux model, tuned with raster and point data

Data-Model Fusion: Technical Approach, continued

- How our salininty-intrusion GA-Optimized neural net model at Carquinez Strait and Suisun Marsh support the overall project:
  - Build and optimize our NN model for flow / stage / salinity intrusion at Carquinez Strait, and calibrate with the Dept. of Water Resources' DSM2 Hydrodynamics Model, using their archived flow / stage data for San Francisco Bay Delta
  - Transfer this calibrated NN model to the Elkhorn Slough Estuary where the NN model acts as a constraint on the development of an Elkhorn Slough hydrodynamics and sediment transport model

    [based on the COHERENS product]
  - Fuse the NN model and the hydrodynamics model with remote sensing data, from both atm-corrected MASTER data of VNIR & TIR salt marsh/plant emissivities and from visible data of the sediment plume in Monterey Bay
  - Create a better predictive model for the mid- to long-term behavior of the Slough as a habitat

Data-Model Fusion: Schedule and Plans

- July 2002:
  - Coordinated system demo of prototype, showing interactive exploration of remote sensing and other datasets, guided by model predictions from GTR

- January 2003:
  - Demonstrated results of numerical “degraded” experiments, providing methods for analyzing robustness and flexibility of the GTR method in Earth remote sensing modelling; automated balancing of regularization parameter that selects between “belief in model vs. belief in data”

Data-Model Fusion: Schedule and Plans

- July 2003:
  - Demonstrated application of GTR methods to glacier / ice sheet data-model fusion, and flow prediction capability scaled for climate model parameterization

- July 2004:
  - Demonstrated application of GTR methods to ecosystem dynamics using percolation and/or connectivity models

Back-up Slides

Technical IT and Physics Challenges

TASK 1 OBJECTIVE

- Modify and Deploy the DARWIN Database Infrastructure to Operate across Earth Science Models and Remote Sensing Satellite Datasets

TECHNICAL CHALLENGE

- Establish split between appropriate metadata and model or sensor data
- Identify metrics, search criteria, and labeling protocol for data or schema
- Condition the datasets for appropriate feature extraction and mining
- Build active algebraic models that support either sensor or model data

TASK 2 OBJECTIVE

- Create a Generalized Regularization Framework that Fuses Earth Science Process Models with Remote Sensing Datasets, for Data Completion and Error Identification

TECHNICAL CHALLENGE

- Deploy physics-based models that interpolate across sparse point data
- Fuse low-fidelity models with high-fidelity data
- Detect data anomalies from model predictions [KDD]
- Correct and refine process models by isolating data errors

Data-Model Fusion: Technical Problem Statement

Domain of Application and Testing

• Our initial Earth Sciences models include radiative transfer algorithms for temperature-emissivity separation, to be fused (via GTR) with data from the Thermal IR bands from airborne and satellite sensors. MODIS, ASTER, and Airborne MASTER data form the basic data sets for our numerical experiments.

• We are also applying the GTR methods to validation of estuary flow and sediment loss within the Elkhorn Slough near Monterey, CA. The flow model is a neural network based model, constrained by data taken in collaboration with Moss Landing Marine Lab.

• We anticipate applying GTR to glacier and ice sheet flow models and to climate parameterization from these. Data sets include both VNIR and SAR interferometry data; these are not yet acquired.

• An ultimate goal is to apply GTR to ecosystem dynamics models for either disease or wildfire propagation based on application of percolation dynamics/connectivity methods. Success will depend on the degree of fidelity of models currently available within the Earth remote sensing community.

Data-Model Fusion: Technical Approach, continued

- Out-Year Activities:
  - Apply lessons learned from radiative transfer and sediment flux modelling efforts with GTR to glacier and ice sheet flux models, fusing interferometry data and time-differentiated VIS data over ice sheets
  - Initiate percolation dynamics / connectivity models as the GTR regularizing functionals for disease or fire propagation, coupled with VNIR, SWIR and TIR datasets
  - Experiment with degraded classification or regression methods, and couple this with a variety of basis functions in the GTR and with error-driven regularization parameters across a raster image