Hindcasting and IAMs: Theory and an Application in GCAM

VAIBHAV CHATURVEDI, STEVE SMITH, JAE EDMONDS, SONNY KIM, KATE CALVIN, MARSHALL WISE, LEON CLARKE

5th Annual Meeting of the Integrated Assessment Modeling Consortium
Utrecht, Netherlands

November 13, 2012

U.S. DEPARTMENT OF ENERGY | Office of Science
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965
UNIVERSITY OF MARYLAND

This talk is about an experiment in hindcasting in GCAM. Here's the synopsis.

- We have calibrated a version of GCAM to (1) 1990 data and (2) 2005 data.
- This version of GCAM has the U.S. energy system broken out into 50 states.
- We have run both versions of the model forward, using some exogenous input assumptions for which we have data (for population, GDP, etc.) and using rules of thumb for creating assumptions of other assumptions.
- We then evaluate two questions:
  - How well does the 1990 calibration replicate the period from 1990 through 2005?
  - How much does it matter in terms of the long-term projections.
- We are looking only at building energy consumption.

- We're doing the same thing with the AgLU model.

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Some Theory

Some Historical Perspective

At one of the first meetings devoted to modeling of energy systems:

> In the case where complex (i.e., many equation) models are used for forecasting purposes, there appears to be room for the development of better methods of measuring the confidence with which the results should be viewed, even given that one accepts the model specification and data used for estimation (Searl 1973).

This is not a new issue!

Nor have the questions asked changed dramatically. A bit more than a decade later, Hans Landsberg (1985) considered “the many questions in the energy field that are waiting for answers in the next 20 years.”:

- Will nuclear energy resume its growth? Will there be breakthroughs on the breeder and on fusion?
- Will solid-to-gas and solid-to-liquid conversion become commercially viable? For coal? For shale? If so, when?
- Will solar make inroads or continue to server only specialized markets?
- Will the price of oil resume its upward movement? If so, how soon?
- When and at what cost will we have clean-burning coal?
- Will the efficiency of the U.S. automobile climb above the government-mandated 28.5 mpg? How far and how soon?

When is it possible to "validate" any model?

- Hodges & Dewar (1992) proposed the following (widely cited) criteria:

  - Modeled variables are observable,
  - Relationships exhibit constancy of structure in time,
  - Exhibit constancy across variations in conditions not specified in the model, and
  - Permit the collection of ample and accurate data.

  Hodges & Dewar (1992)

- While models of physical systems, in principle, may satisfy some of these requirements, Oreskes et al. (1994) argue that models of natural systems can never be validated.

> *Verification and validation of numerical models of natural systems is impossible. This is because natural systems are never closed and because model results are always non-unique. Models can be confirmed by the demonstration of agreement between observation and prediction, …. The primary value of models is heuristic.*

What about "validating" IAMs?

- Modeled variables are observable,
- Relationships exhibit constancy of structure in time,
- Exhibit constancy across variations in conditions not specified in the model, and
- Permit the collection of ample and accurate data.

Hodges & Dewar (1992)

- For energy system models, Scher & Koomey (2011) state:

  Physical systems generally exhibit such structural constancy, but economic and social systems do not. Dynamic market forces, influenced by non-linear technological and behavioral changes, are highly uncertain and are subject to rapid changes. Consequently, economic modelers cannot assume a level of structural rigidity in the economy sufficient to satisfy criteria 2 and 3 above. As a result, models that attempt to describe economic systems will not yield accurate results, especially in the long term.

- This leads to a some questions:
  - What are the purposes of evaluation and hindcasting exercises?
  - How do we think about hindcasting exercises in particular?

What is the purpose of model?

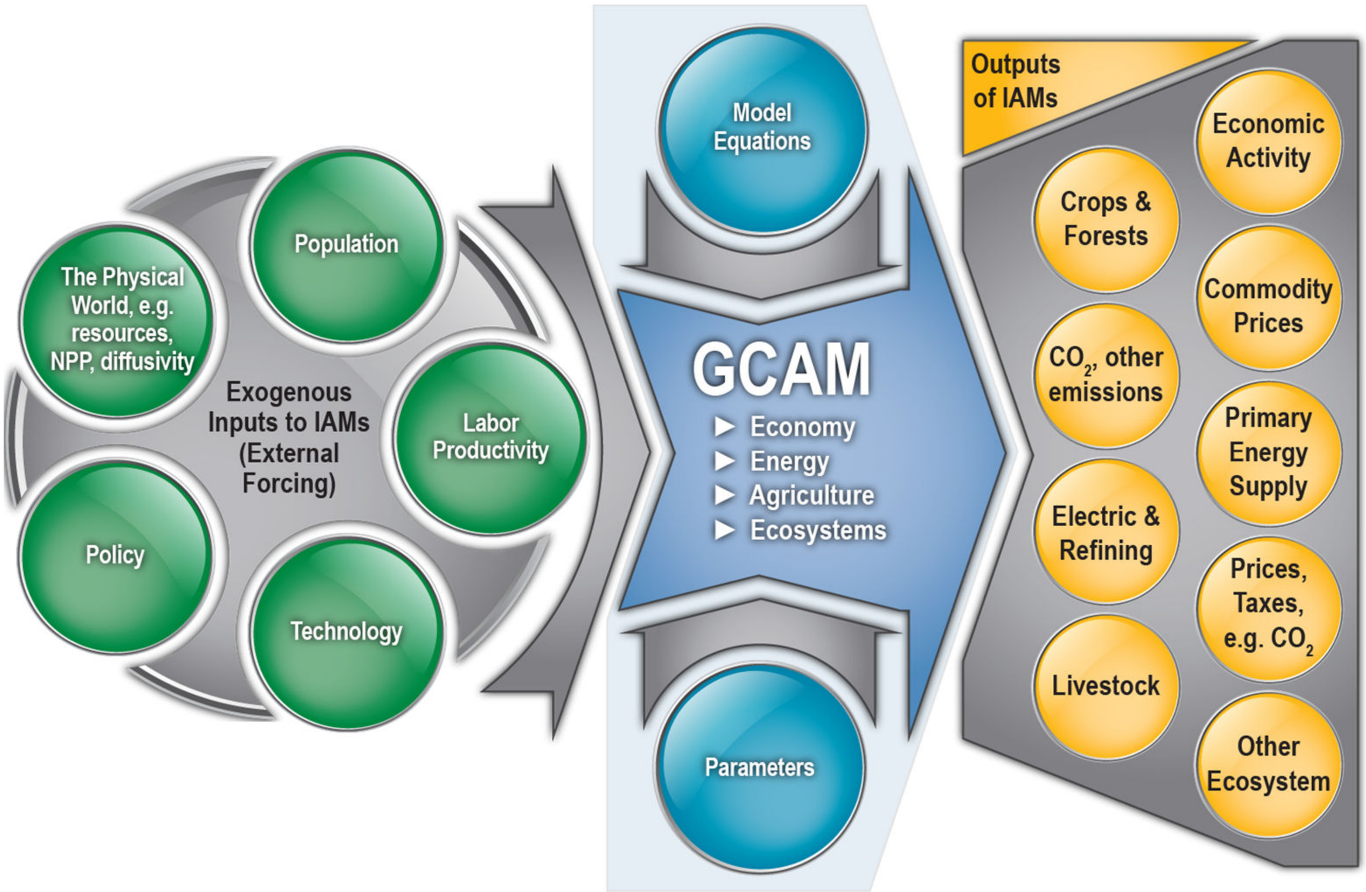
- Craig et al. (2002), expanding on Hodges & Dewar (1992), identify several uses, some of which have long been identified by the IAM community.

  1. As a bookkeeping devices,
  2. as an aid in selling an idea or achieving political ends,
  3. as a training aids,
  4. as part of an automatic management system whose efficacy is not evaluated by using the model as if it were a true representation,
  5. as aids in communication and education,
  6. to understand the bounds or limits on the range of potential outcomes, and
  7. as aids to thinking and hypothesizing.

- In light of these uses, could evaluation exercises:
  - Aid in the communication of results?
    - Projections not forecasts.
  - What, really, do our models say about future bounds or limits?
    - Diagnostics – future scenario behavior as compared to history.
  - What would further aid “thinking and hypothesizing”?
    - Important since IAMs, by necessity, go far outside of historical experience.

What is an integrated assessment model?

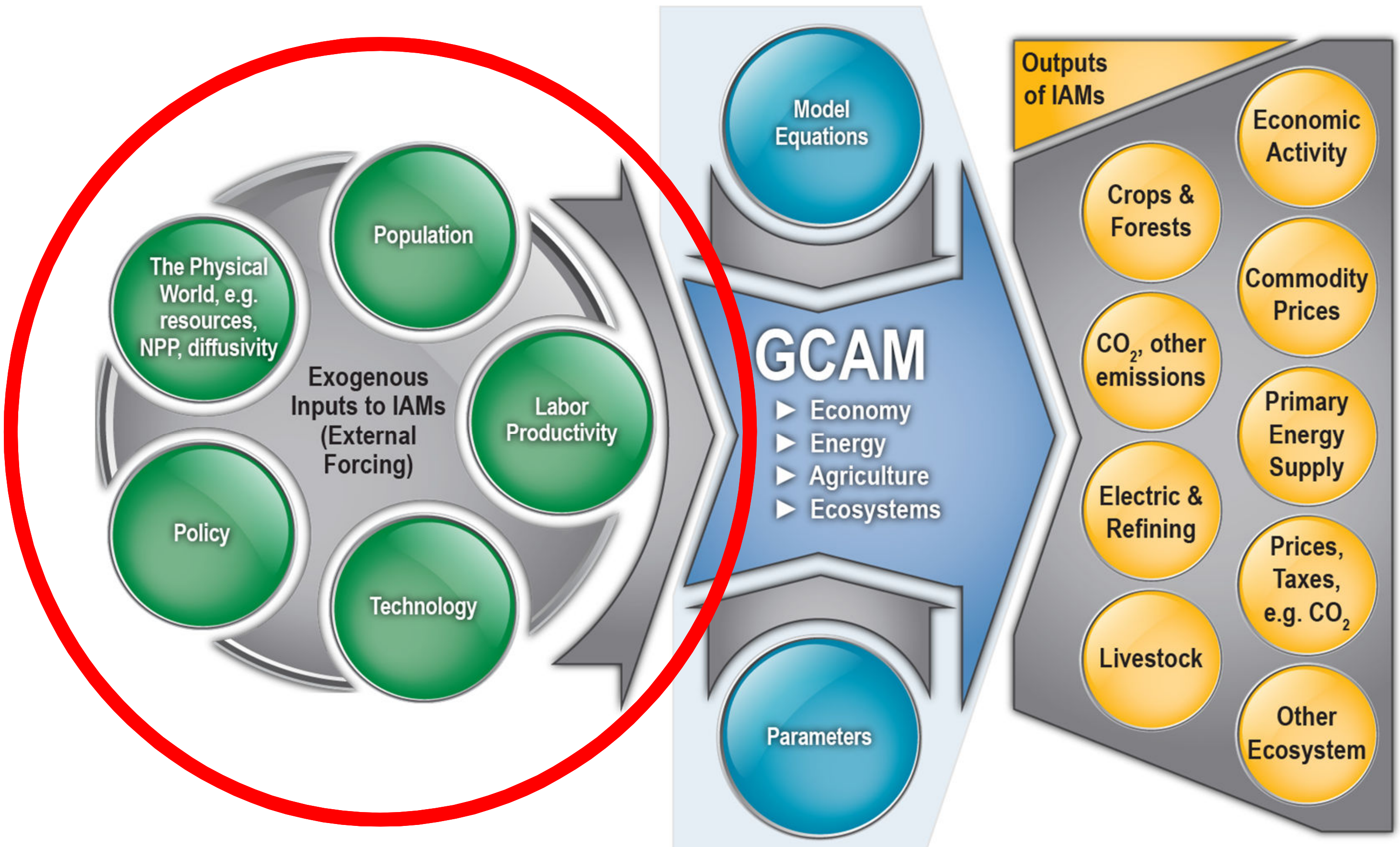
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965



What elements of an integrated assessment model do we want to evaluate?

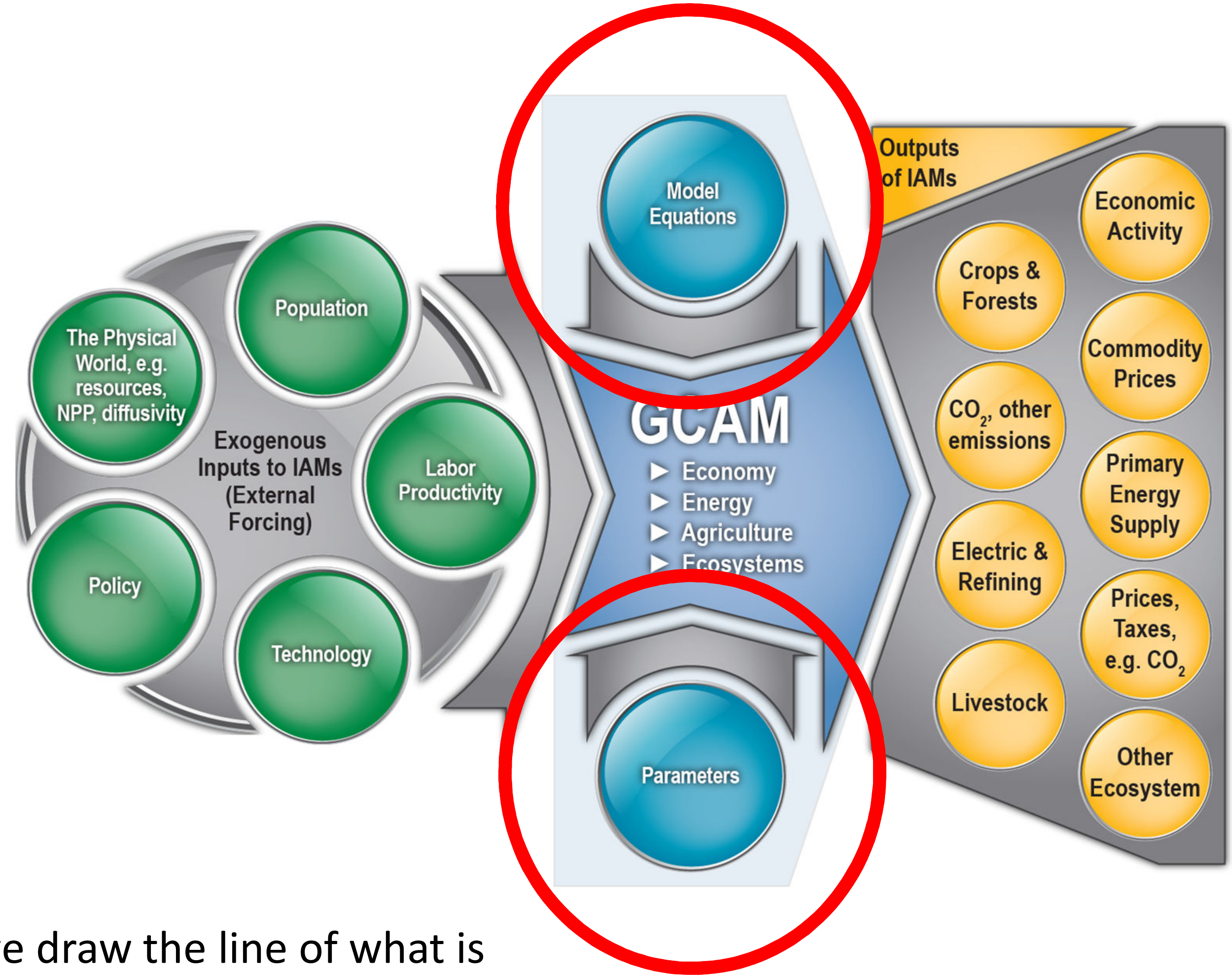
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

What's our take on evaluation of ESMs, given that they take RCP-style information as drivers?



Is assessment of the ability of modelers to create accurate projections of exogenous input assumptions part of evaluation?

Conditional on what?

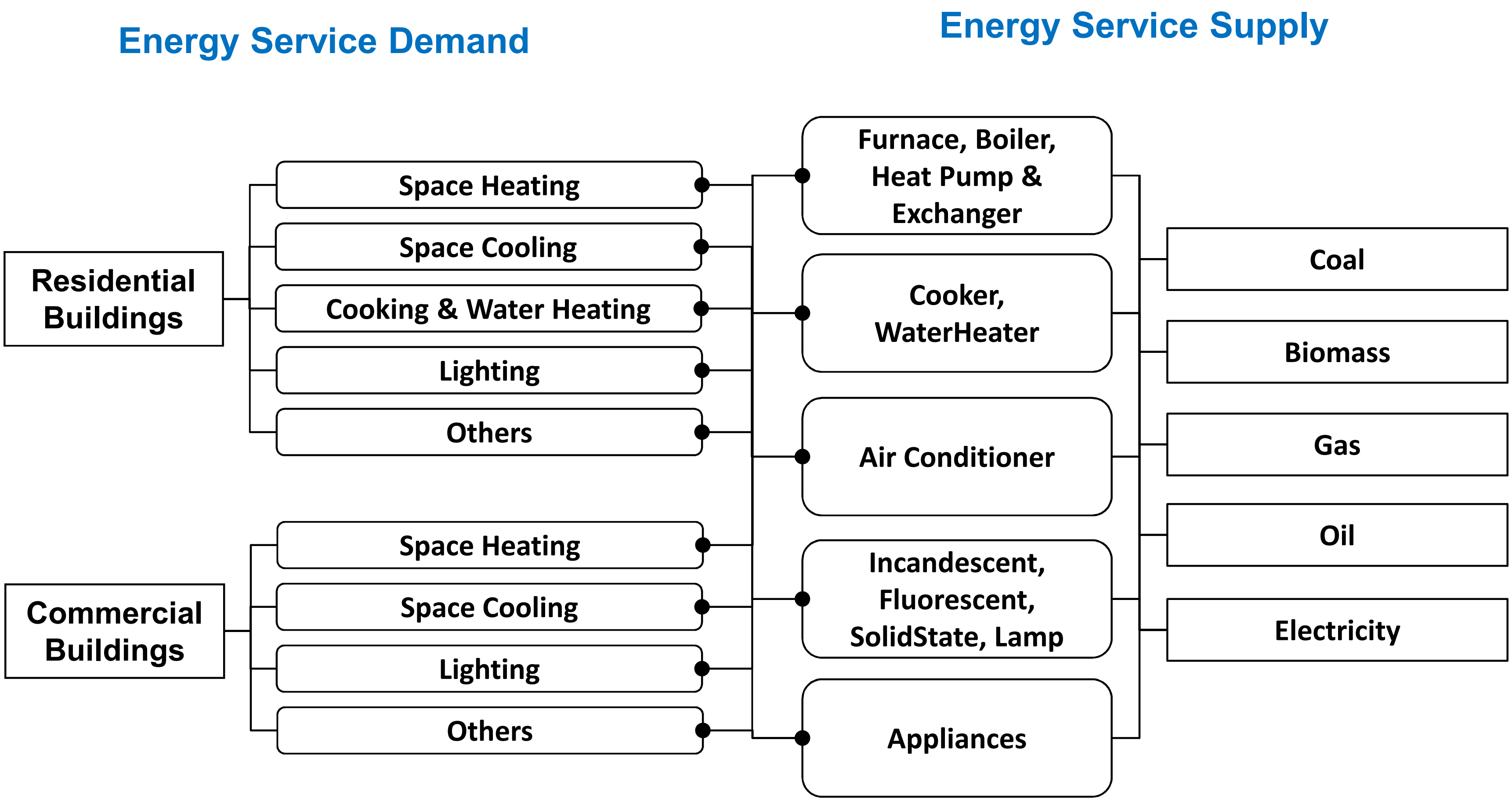


Where do we draw the line of what is model and what is assumption? What do we really want to evaluate?

A GCAM Model Evaluation Experiment

Background: the 50-state buildings model in GCAM

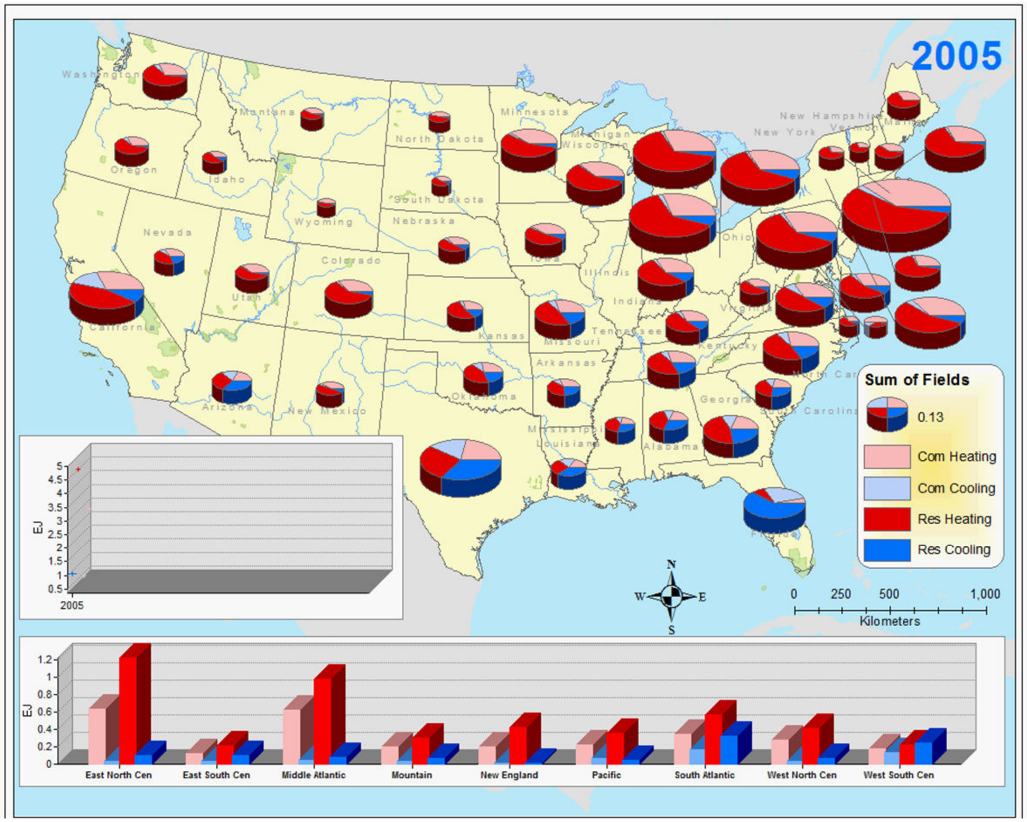
We are using our standard building modeling framework



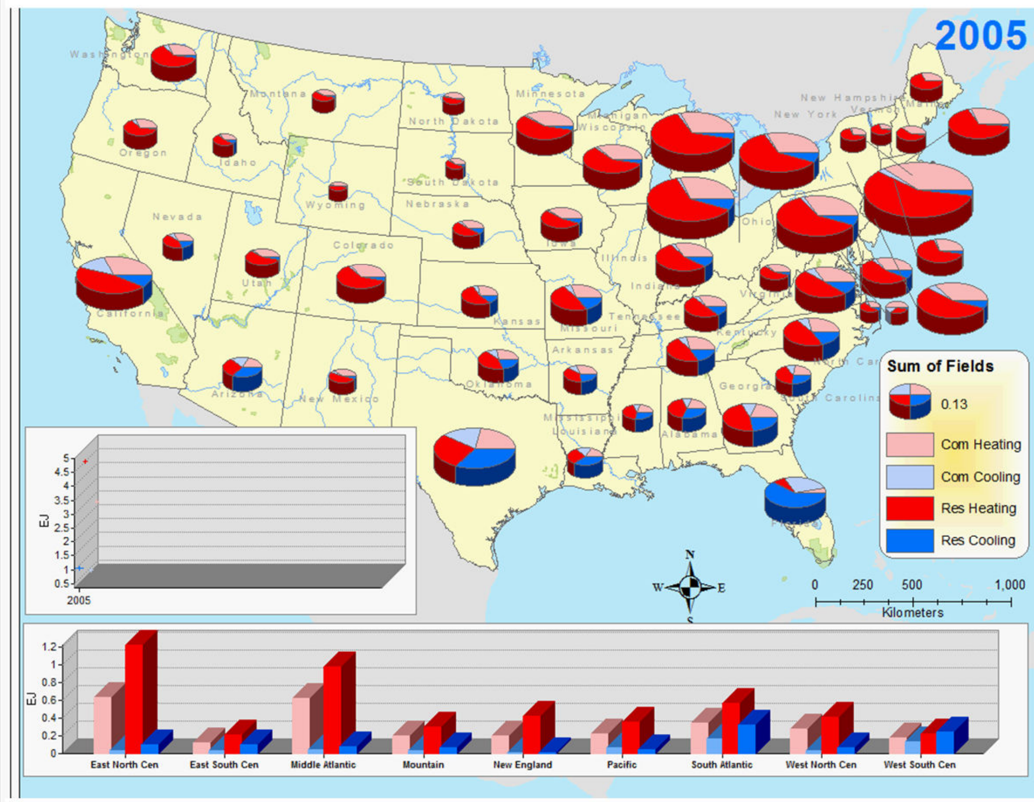
For this study, we are using a version of GCAM with the U.S. energy system split into 50-states.

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

RCP 4.5



Fixed Climate



Buildings service demands

Demand for Space Heating Service [GJ-output/m²] :

$$Q_{H,t} = k_H \cdot \left(HDD_t \cdot ShellEff_t \cdot SurfaceRatio_t - \lambda_H InternalGain_t\right) \cdot \left[1 - \exp\left(-\frac{\ln 2}{\alpha_H} \cdot \left(\frac{Y_t}{P_{H,t}}\right)\right)\right]$$

Space Heating Requirement (satiated demand)

Economic Behavior

Demand for Space Cooling Service [GJ-output/m²]

$$Q_{C,t} = k_C \cdot \left(CDD_t \cdot ShellEff_t \cdot SurfaceRatio_t + \lambda_C InternalGain_t\right) \cdot \left[1 - \exp\left(-\frac{\ln 2}{\alpha_C} \cdot \left(\frac{Y_t}{P_{C,t}}\right)\right)\right]$$

Space Cooling Requirement (satiated demand)

Economic Behavior

Hindcasting Methodology

U.S. DEPARTMENT OF ENERGY | Office of Science

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

UNIVERSITY OF MARYLAND

We did this as simply as possible

- Historical simulations from 1990-2100 and 2005-2100 in 5 year time-steps.
- These are simulation with existing data (no new data for pilot study).
  - GCAM is already calibrated for 1990 and 2005.

- No modifications to the code have been required for the pilot study.
  - GCAM is designed to be flexibly in spatial and temporal scales.
  - GCAM's behavior designed to be data driven.

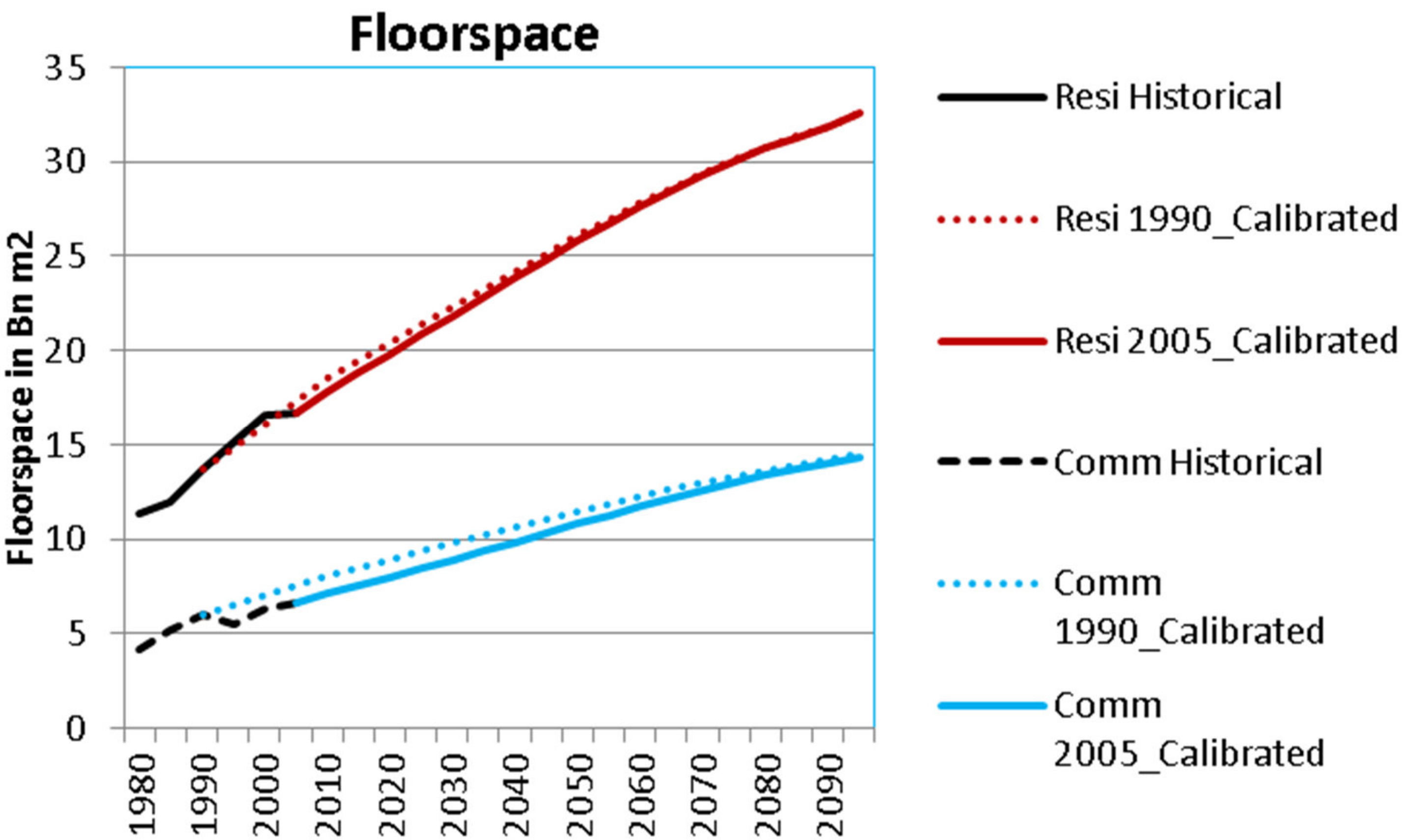
- For intermediate years, we used a range of methods to get assumptions or drivers.
  - Some are from data.
  - Some are interpolated between 1990 and 2005.
  - Some are held fixed going forward.
  - We care a lot less about assumptions outside of the buildings sector.
    - But how do we decide which assumptions will have the largest effect on the buildings sector?

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Hindcasting Results

Floorspace Results

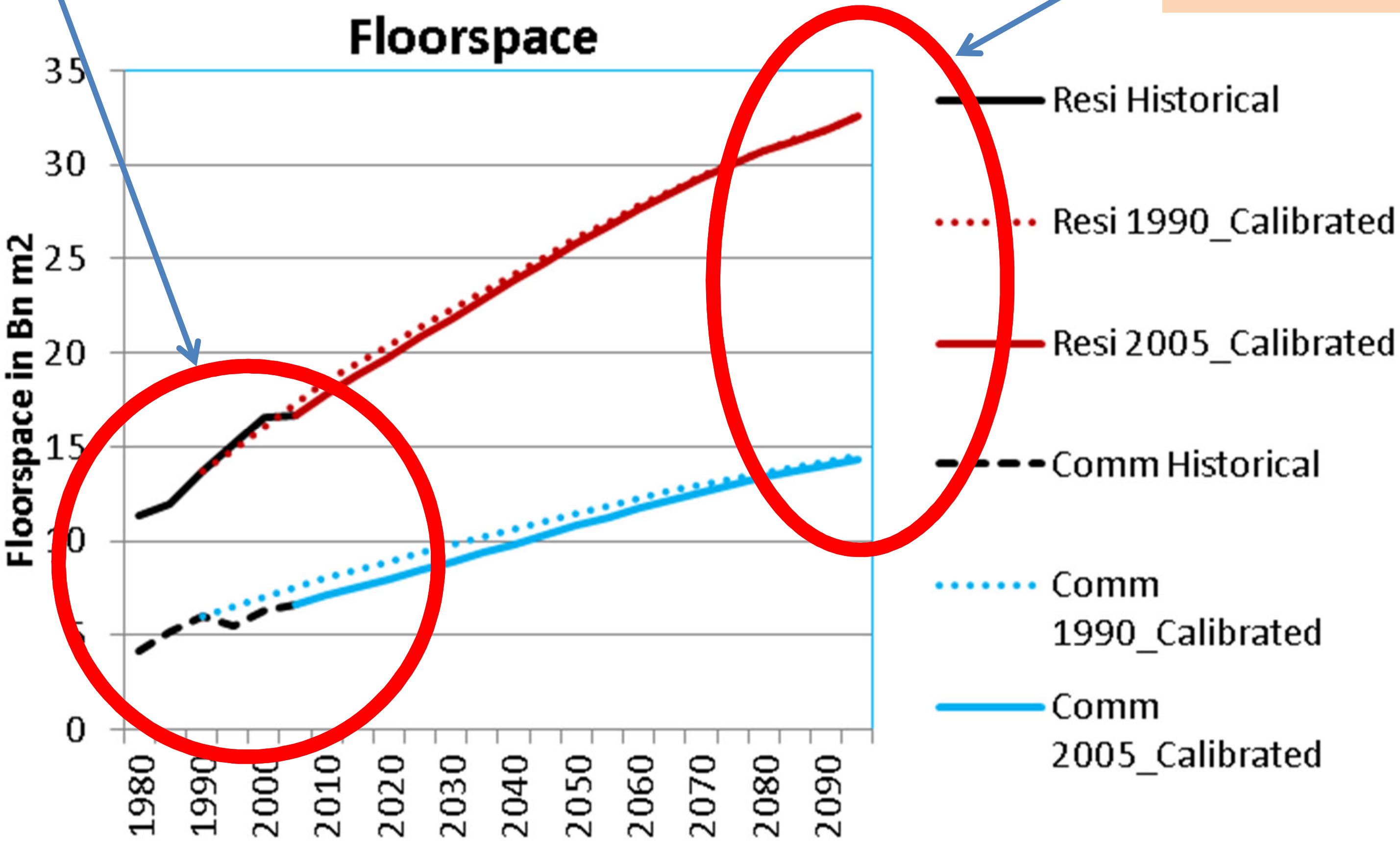


Numbers are for heated floorspace.

Floorspace Results

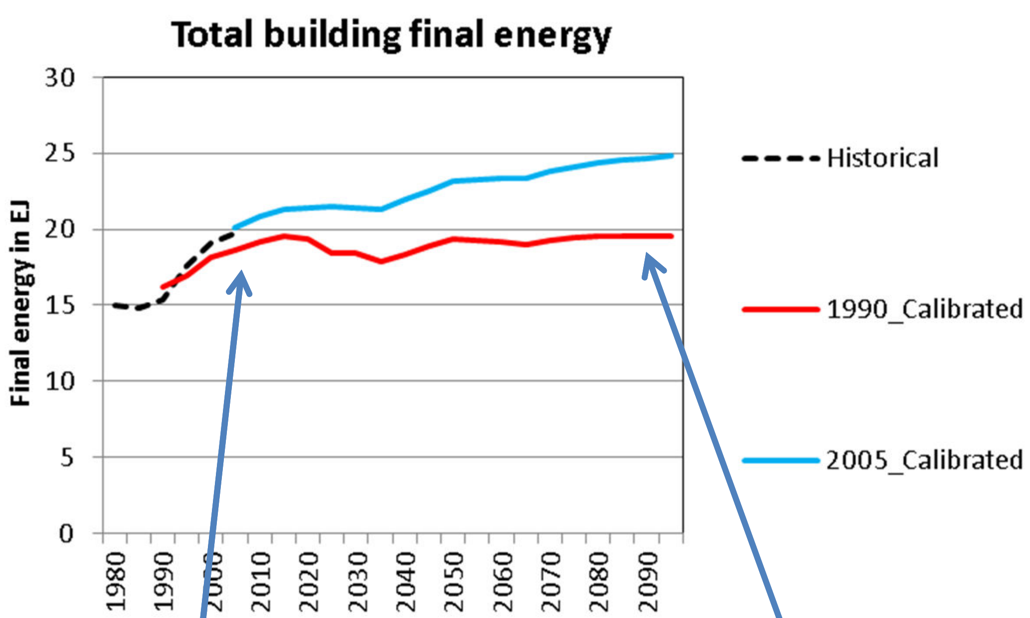
We weren't too far off in 2005.

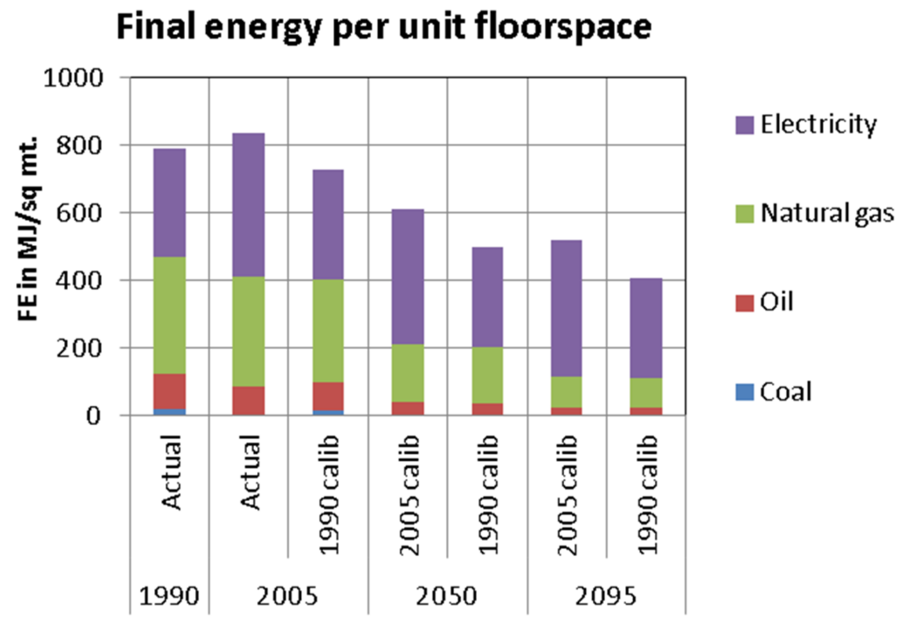
But it didn't matter much anyway in the long run.



Numbers are for heated floorspace.

Total Final Energy Results



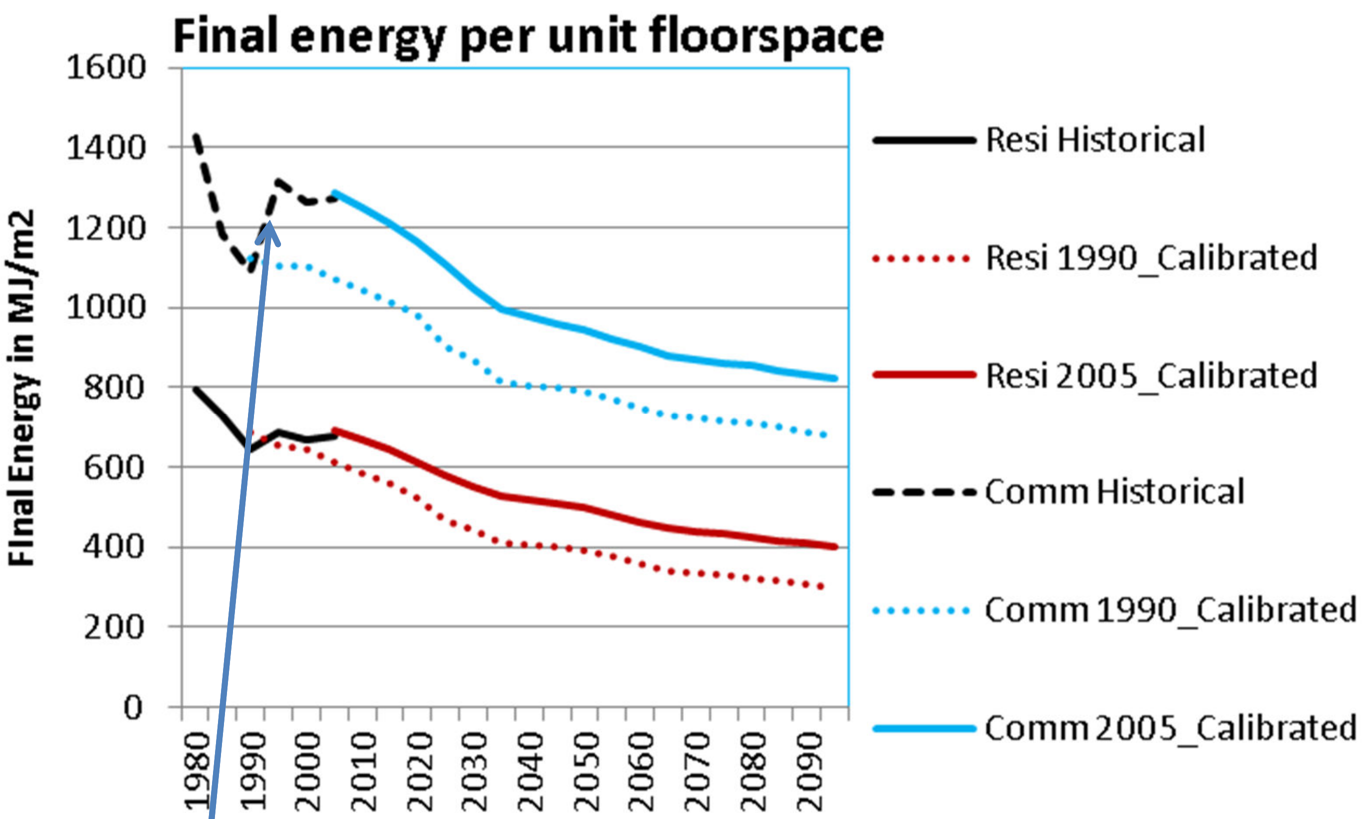


We didn’t do all that well with the 1990 calibration.

It mattered, because the effect propagated.

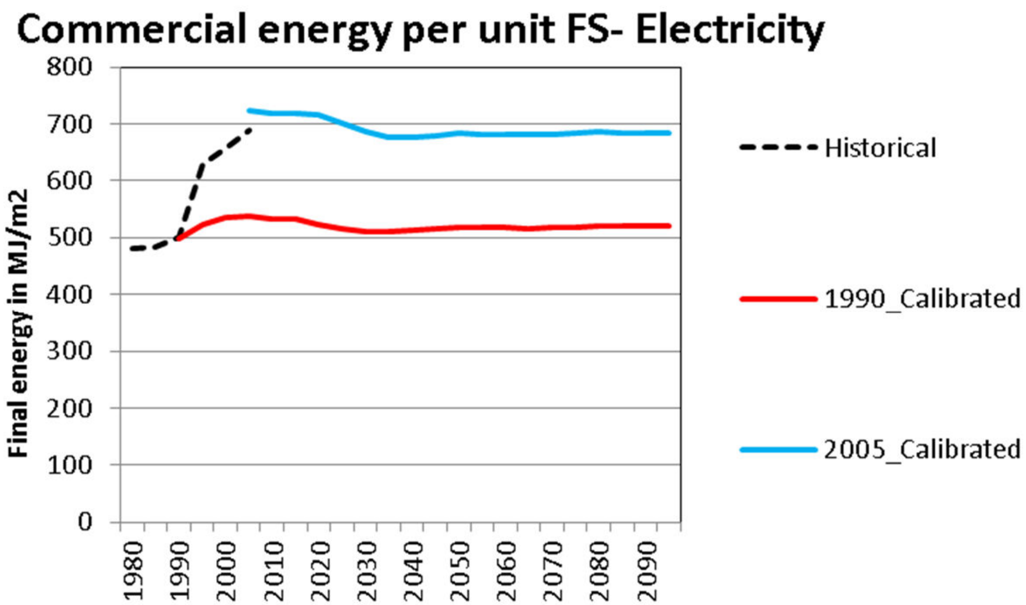
This is largely a story about electricity. We did pretty well on gas and other sources.

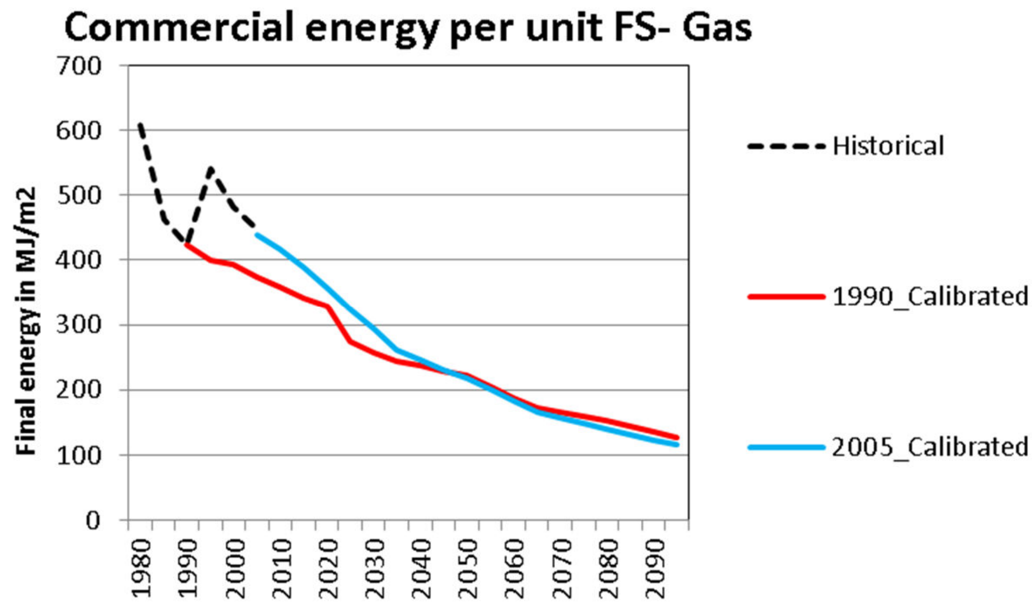
Removing the Floorspace Variation

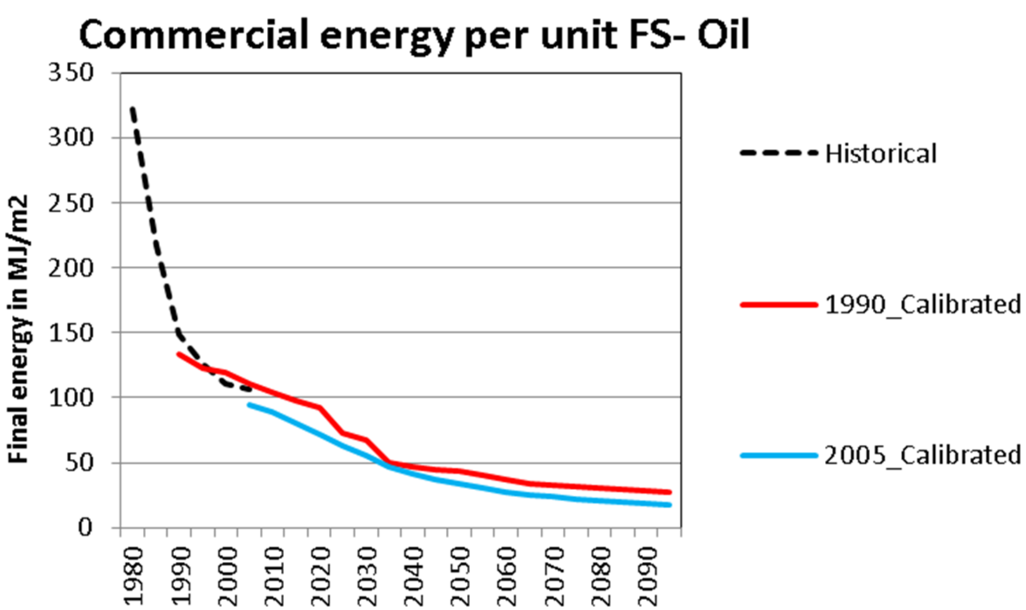


Wow!!

Fuel Consumption in Commercial Sector

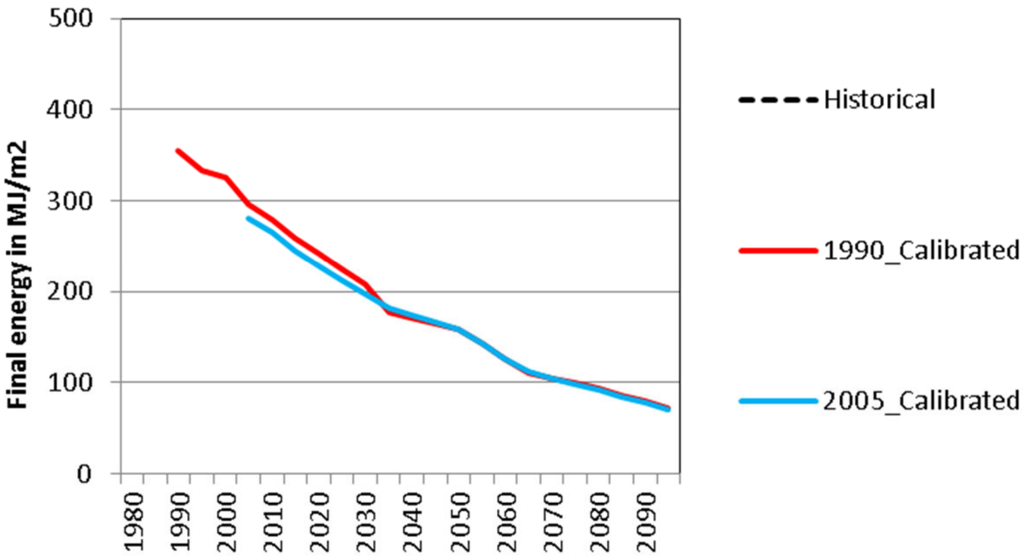




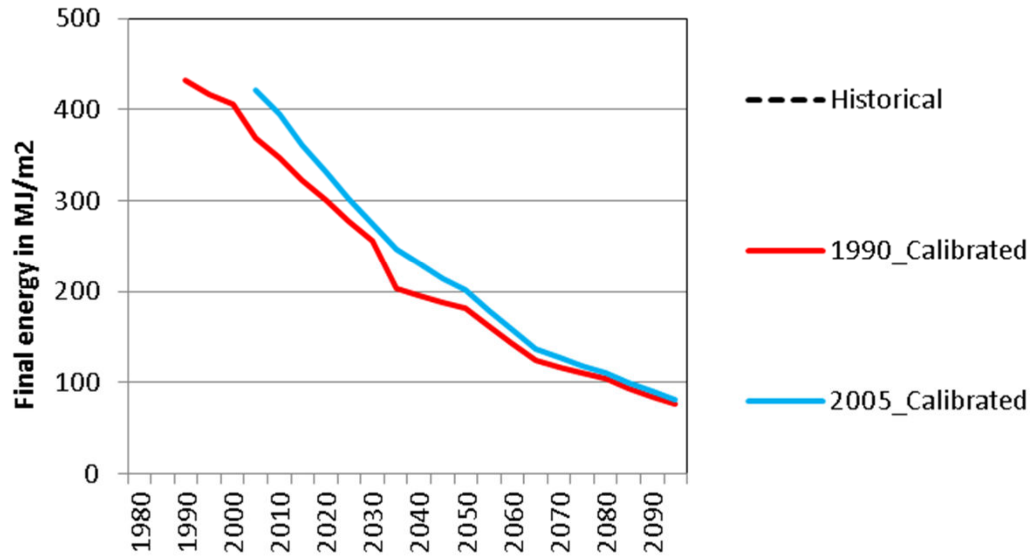


Residential cooling is off by the most in 2005 in a relative sense.

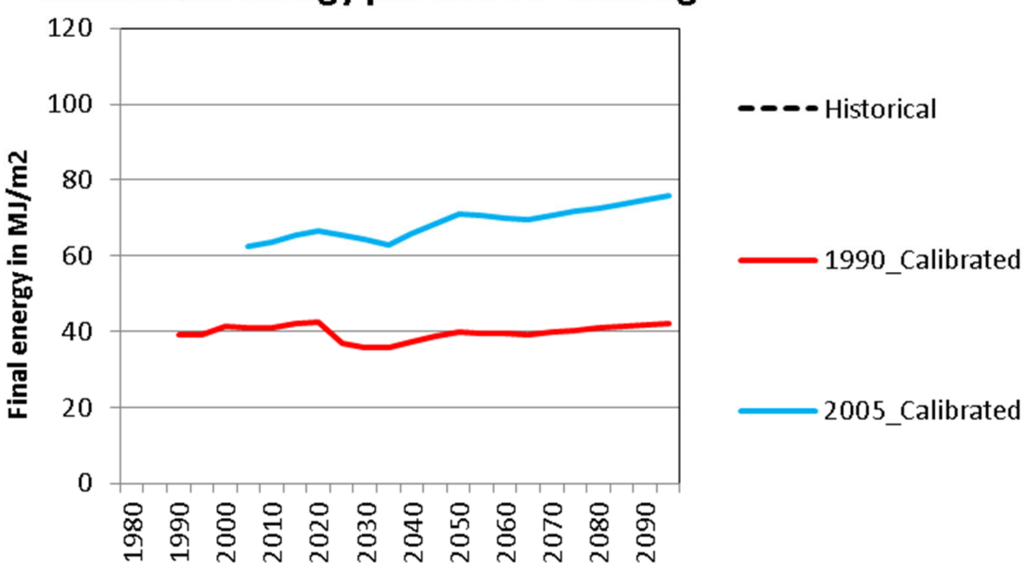
Residential energy per unit FS- Heating



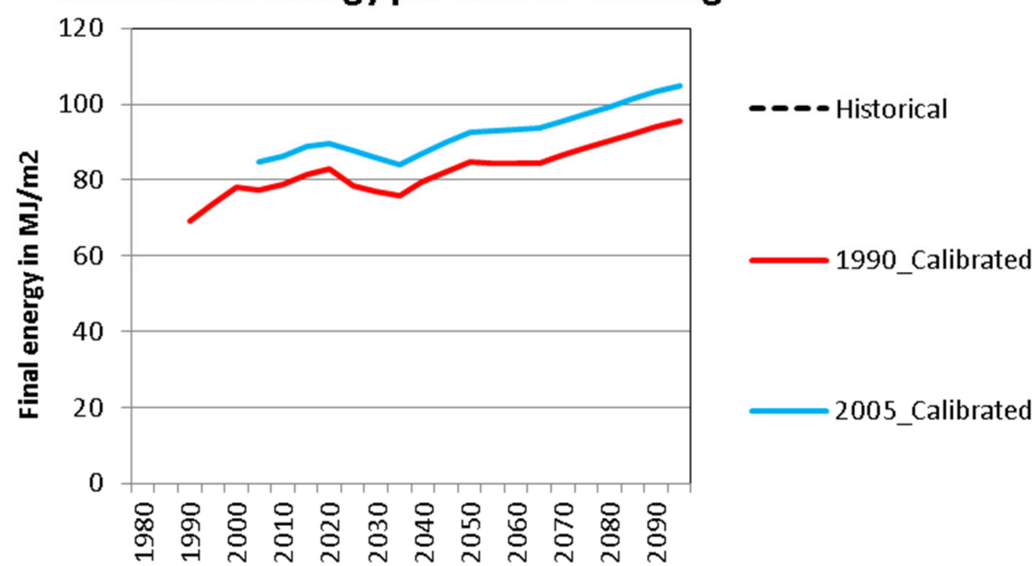
Commercial energy per unit FS- Heating



Residential energy per unit FS- Cooling

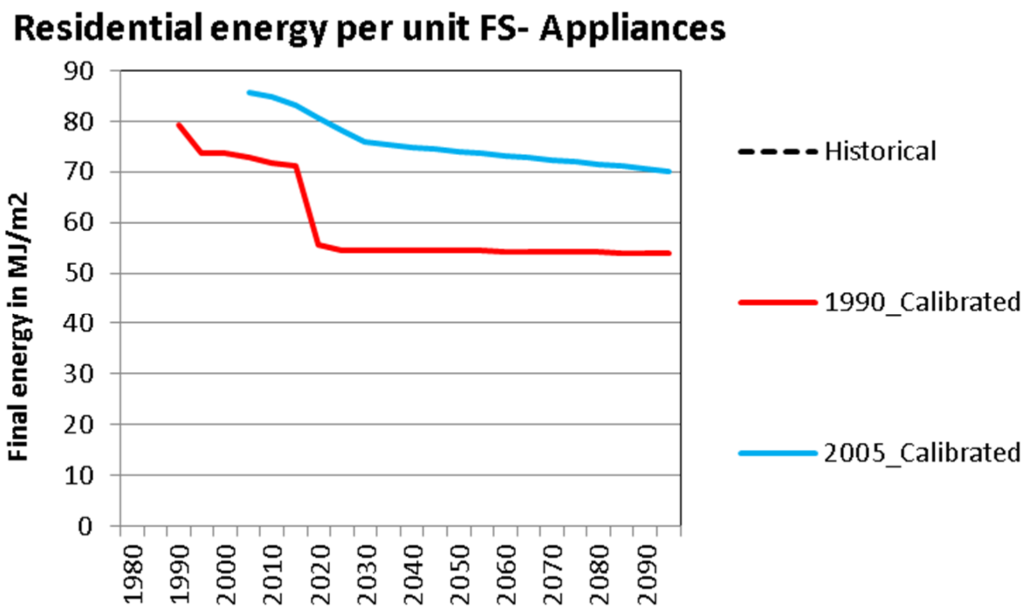


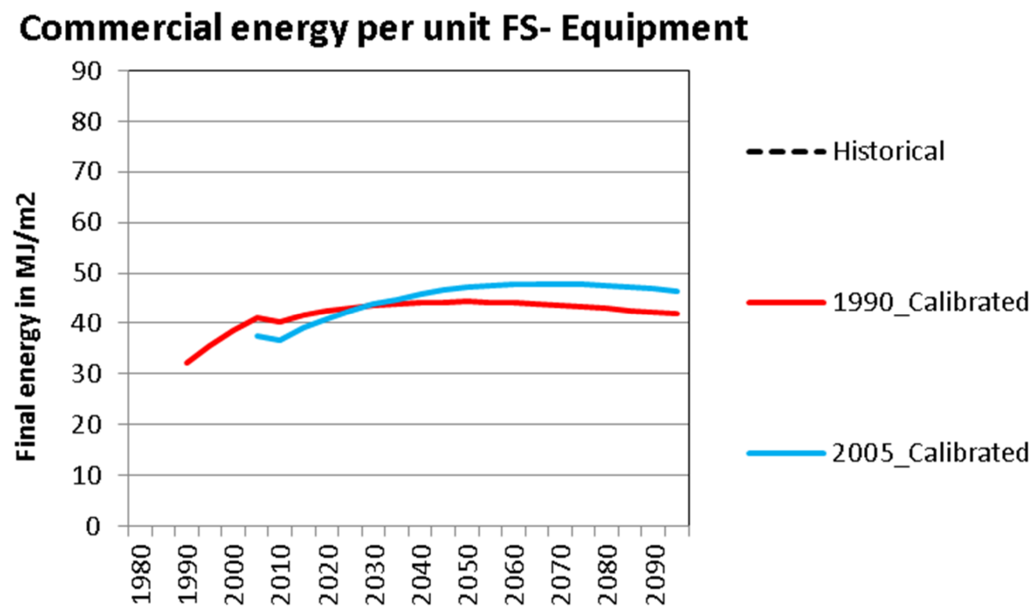
Commercial energy per unit FS- Cooling

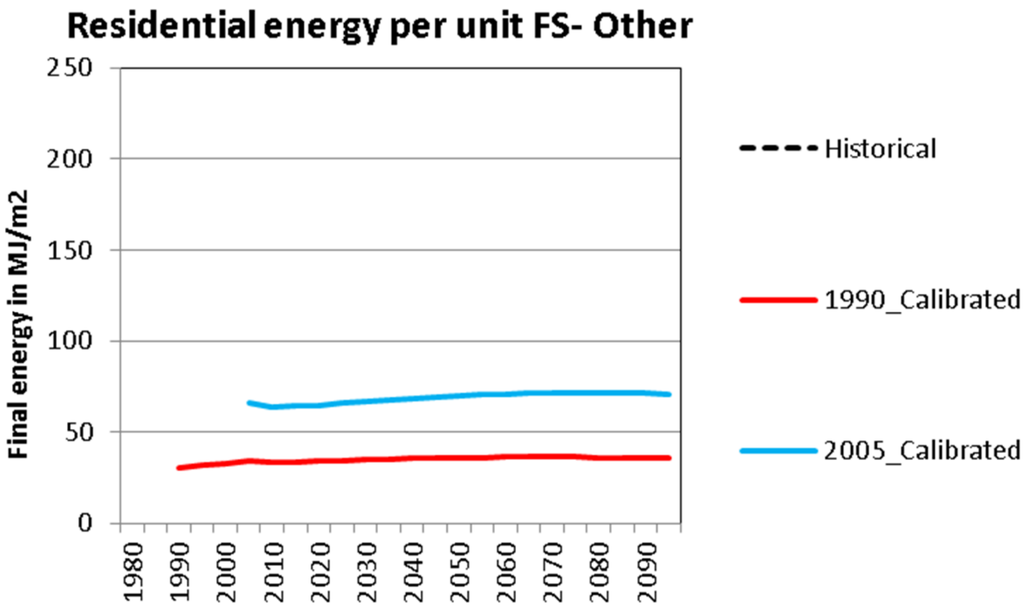


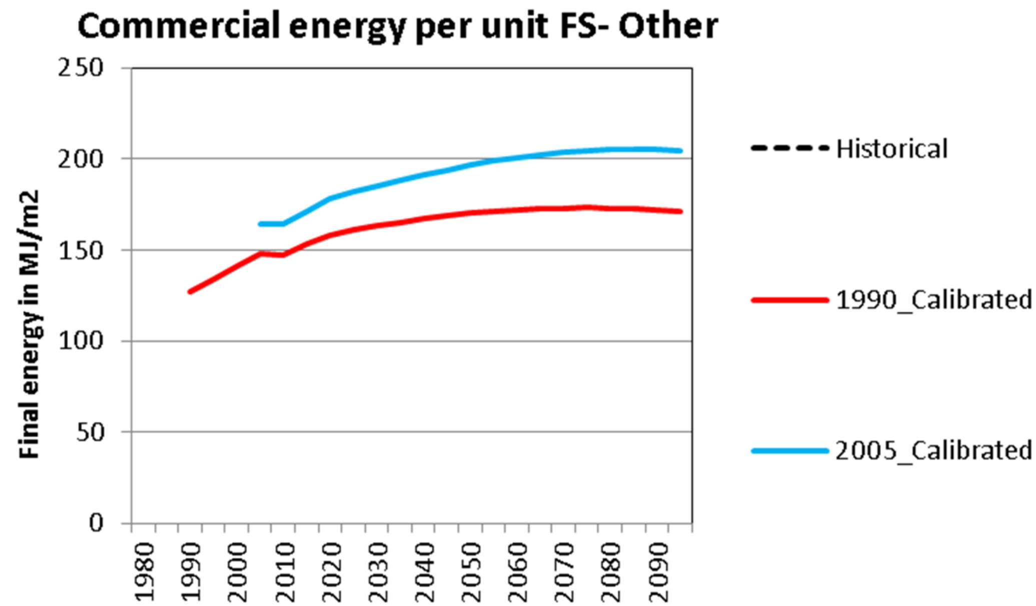
The problem is perhaps even more difficult when it comes to plug loads

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

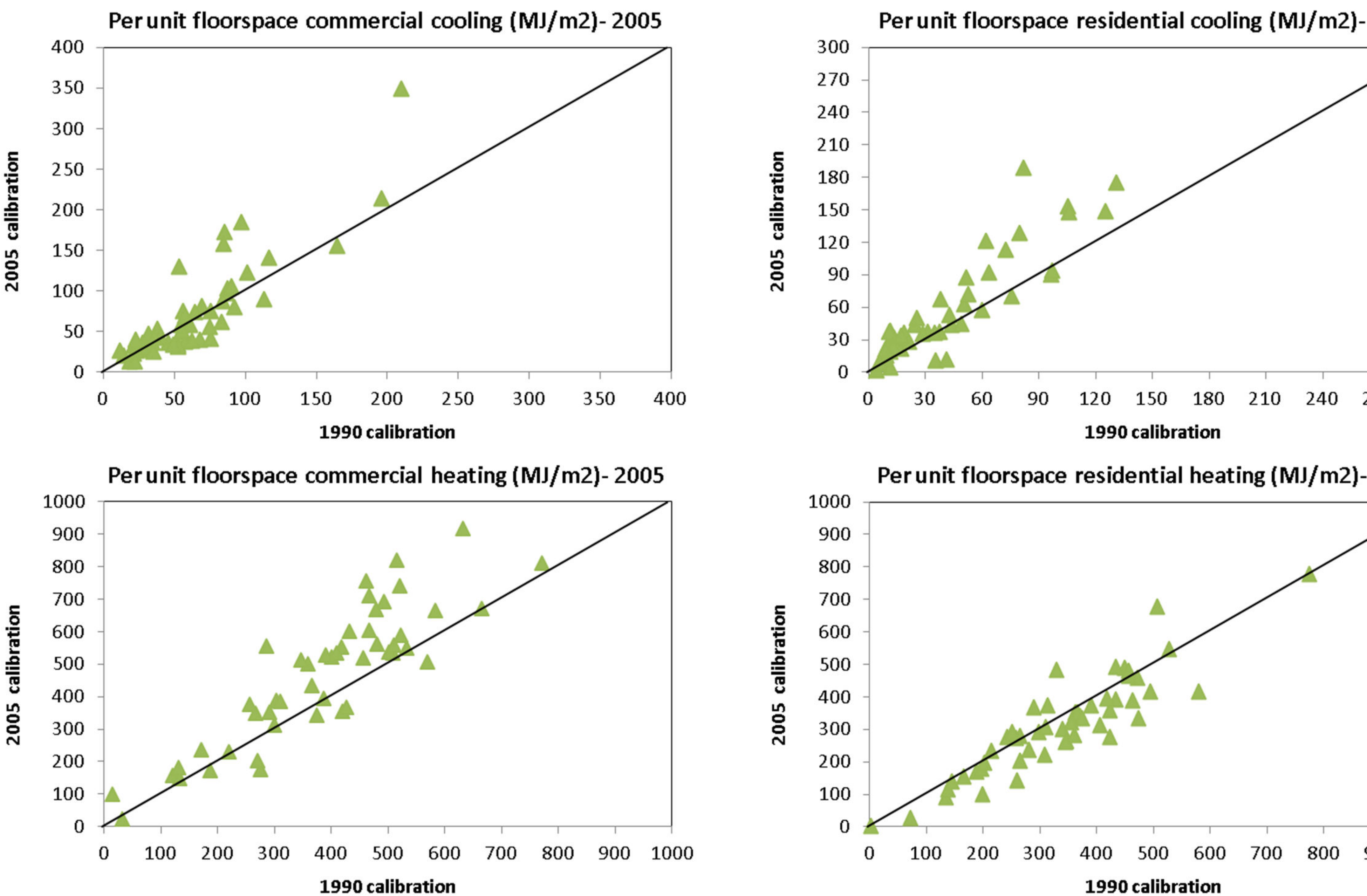








For an impacts study, the regional results will matter



Pacific Northwest NATIONAL LABORATORY

Proudly Operated by **Battelle** *Since 1965*

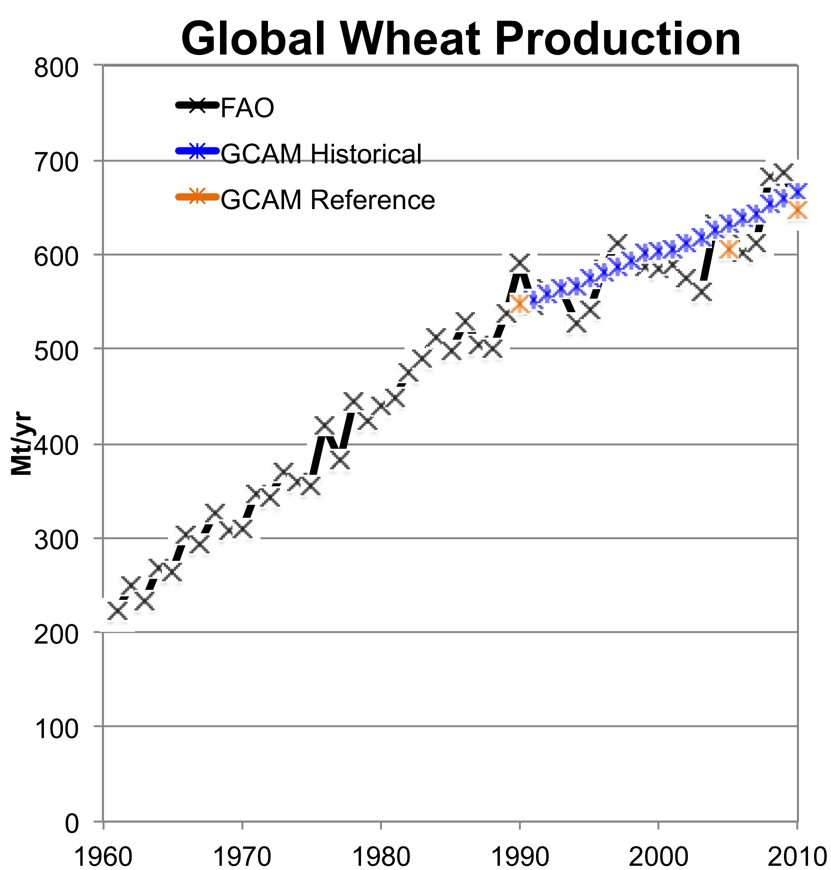
Hindcasting for the AgLU Model

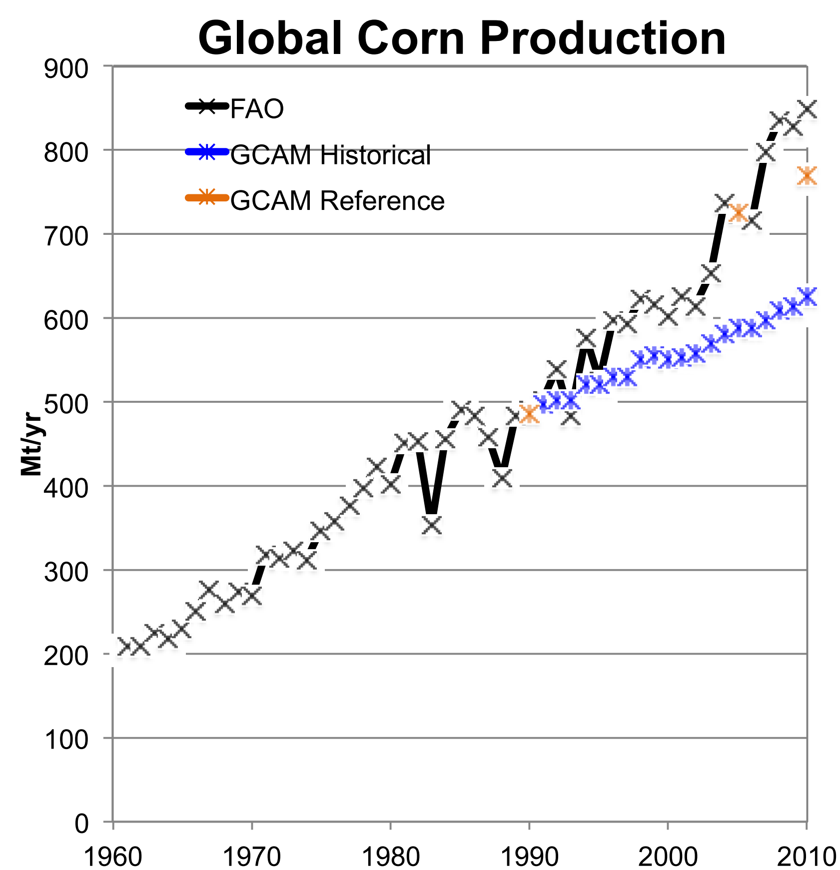
A detailed hindcast across all sectors would take a lot of data

- The assumptions in AgLU used to do the buildings hindcasting were not sufficient for the AgLU hindcasting.
- When do we know how important the interactions between other components of the model might be?
- Can we split components and treat them independently with exogenous inputs?
  - This is hard in GCAM.

Evaluating the AgLU Component of GCAM

- Approach:
  - Calibrate GCAM to 1990 and then run annually until 2010
  - We are providing the model with population, GDP, and technology characteristics (e.g., crop yield)
- Results:





Evaluating the AgLU Component of GCAM

- Approach:
  - Calibrate GCAM to 1990 and then run annually until 2010
  - We are providing the model with population, GDP, and technology characteristics (e.g., crop yield)
- Results:

- At a global level, we match production of crops used for food only pretty well. We do not get oscillations in production, but track averages.
- We miss the growth in production of crops used for biofuels (Corn, PalmFruit) because GCAM did not anticipate biofuels policies
- Next steps:
  - Examine results related to regional production, land area, LUC emissions, NPP
  - Include historic bioenergy policies

Concluding Thoughts

So what have we learned so far?

- We can learn things about our models.
- We have a better ability to “predict” some things than others. The question is how to articulate which are which and what it means to be better.
  - Some elements of projections are closer to exogenous input assumptions.
  - Others can be seen as more conditional. These are the ones we are most interested in.
- The nature of future relationships will determine whether differences in calibration matter.
- Validation could require a lot of data. Or at least you need to draw some pretty tight boundaries if you aren’t going to use a lot of data.
  - Don’t think we don’t have data problems even for the simplest issues.
- Specific to buildings, there are some other things we could potentially do hindcasting style research on:
  - How well we get the HDD/CDD effect? (but does this need to be done in our models?)
  - Effect of price changes? (but does this need to be done in our models?)
  - If we just isolate relationships for components of the model like this, does it count as hindcasting?

THE END