Model Development and Application: Data Needs and Management

Session: Adoption of CWEMF's Modeling Protocols in the California, Review and Case Studies
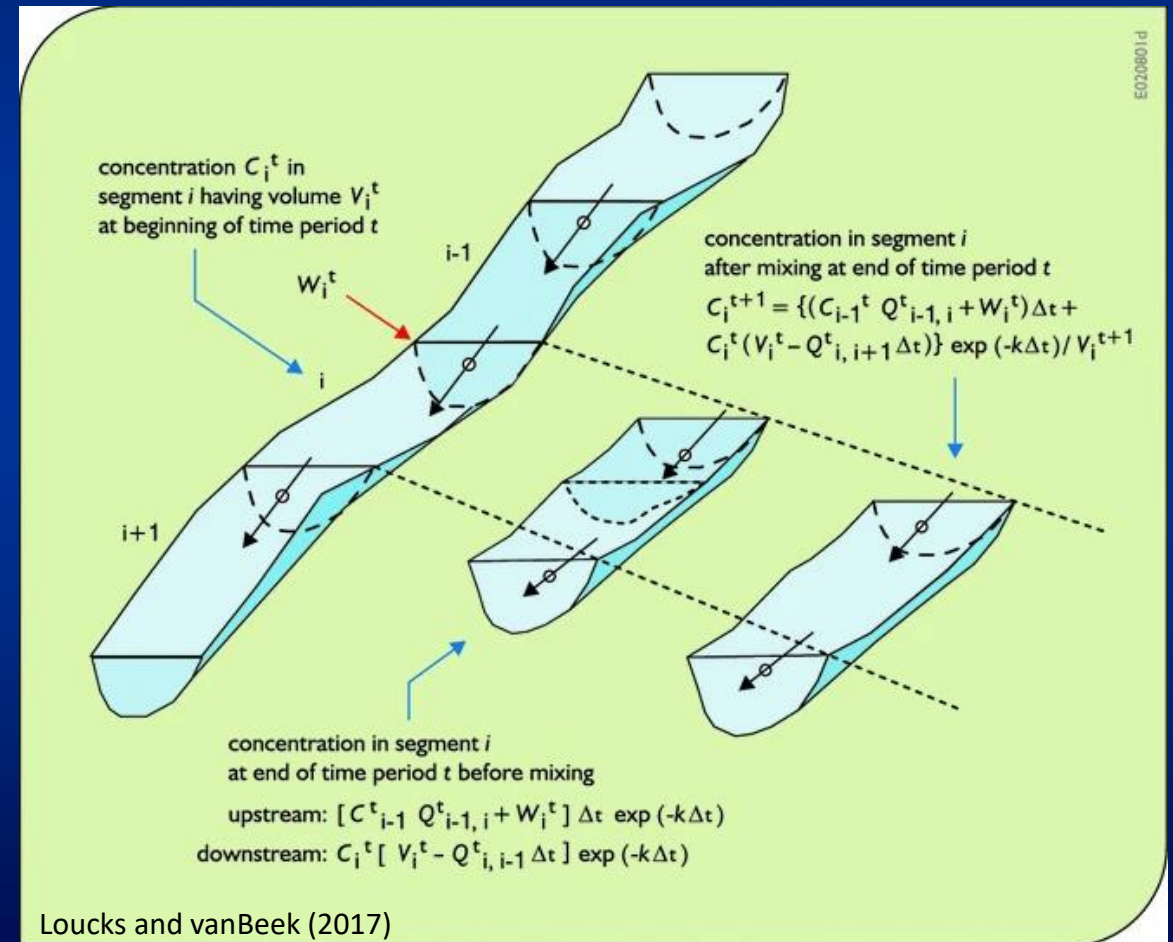
Mike Deas

Watercourse Engineering, Inc.

Davis, CA

mike.deas@watercourseinc.com

Introduction

- As part of modeling, analysts are required to
  - Identify data needs
  - Gather or acquire data
  - Assess data quality
  - Assess data completeness
- This information is used to develop data for model development, calibration, and application

concentration C_i^t in segment i having volume V_i^t at beginning of time period t

W_i^t

$i-1$

i

$i+1$

concentration in segment i after mixing at end of time period t

$C_i^{t+1} = \{(C_{i-1}^t \, Q_{i-1,i}^t + W_i^t)\Delta t + C_i^t (V_i^t - Q_{i,i+1}^t \Delta t)\} \exp(-k\Delta t)/V_i^{t+1}$

concentration in segment i at end of time period t before mixing

upstream: $[C_{i-1}^t \, Q_{i-1,i}^t + W_i^t] \, \Delta t \, \exp(-k\Delta t)$

downstream: $C_i^t [\, V_i^t - Q_{i,i-1}^t \Delta t\,] \exp(-k\Delta t)$

Loucks and vanBeek (2017)

Role of Data in Modeling

- Observed Data provide:
  - Initial Conditions
  - Boundary Conditions
  - Calibration
- Not included today...how observed data relate to
  - Geometry and installed facility data
  - Model parameterization
  - Model control

**Generalized Modeling Steps** (CWEMF 2021)

CWEMF Guidance: Data Quality

- Model results are a function of data quality (boundary conditions and calibration data)
- Some pertinent considerations
    - The need and intended use of each type of data or information (why, when, where)
    - Expected sources of these data
    - Underlying quality of the data
    - Criteria established to determine if data quality is acceptable for project (precision, accuracy, representativeness, comparability, completeness)
- Is there a need to develop a monitoring program to collect necessary data?

Model Data Types - Definition

- Observed Data
  - Initial Conditions
    - Observations (or estimations) <u>within the model spatial domain</u> to define the initial state of the system (t=0)
  - Boundary Conditions
    - Observations (or estimations) <u>at the boundaries of the model spatial domain</u> and for the duration of the modeling period at each time step
  - Calibration
    - Observations <u>within the model spatial domain</u> and for the duration of the modeling period. Sufficient to test model

Figure 67. iBCod locations at Beaver Creek, 2004 (distance and elevation in feet).
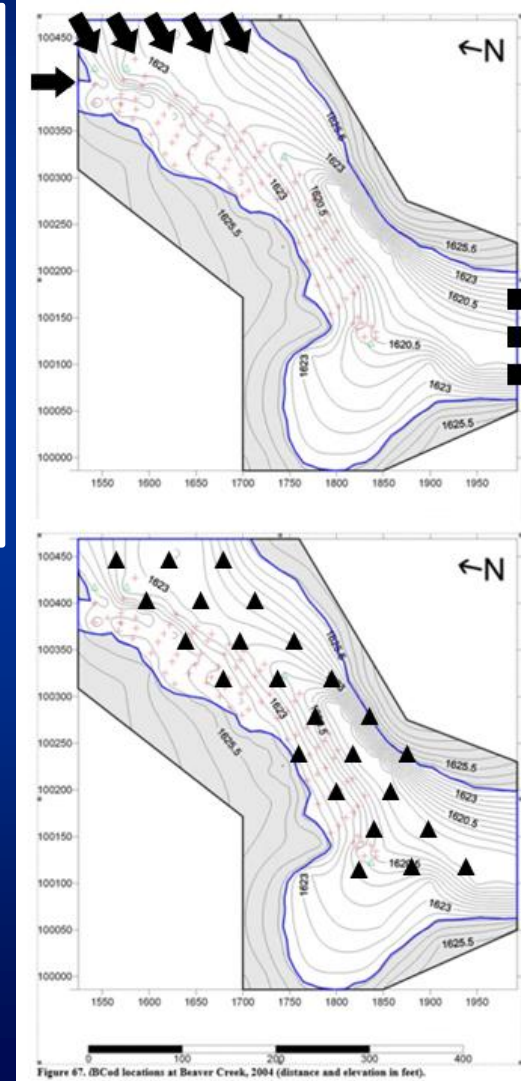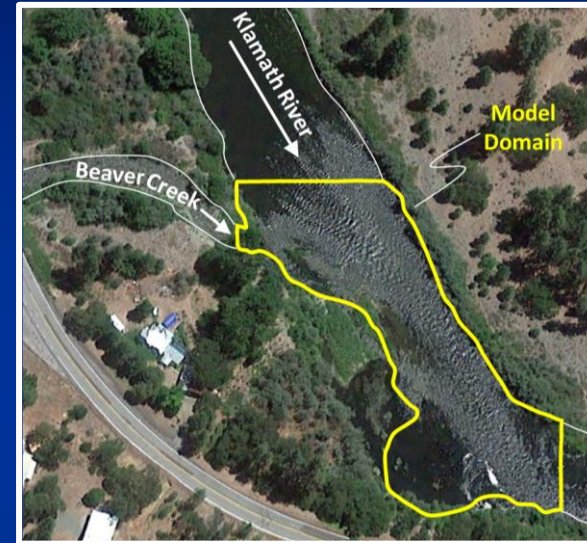
Model Data Development

- Data preparation
    - Compilation (often from different sources
    - Transformations (e.g., conversion to common units)
    - Quality control (remove known outliers, data delimiters, etc.)
- Boundary Conditions
    - Fill data gaps
- Calibration Data
    - <u>Do not</u> fill data gaps

Basic Model Data Types

- Geometry (not today)

- Model Applications
  - Boundary conditions
    - Flow/stage
    - Water quality
    - Meteorology
  - Initial Conditions
    - Stage
    - Velocity
    - Water quality

- Model Calibration
  - Flow/stage/velocity
  - Water quality

- Forecast (not today)

https://www.usgs.gov

https://geo-matching.com

https://www.nps.gov

https://www.campbellsci.com

"Model Ready" Data

- Models require "complete" data sets for boundary conditions
  - Data for each time step
  - Some models can interpolate (step, linear)
- Development of these data sets can be time consuming due to the nature of the various data types
- Data often have gaps (small or large) and at times are unavailable ("missing")
- These gaps need to be filled prior to model calibration and application

Boundary Conditions

- Data QA and gap filling are a function of:
  - Project objective
  - Type of study
  - Modeling time step
  - Steady or dynamic simulation (flow, stage, temperature, water quality)
  - Required accuracy/precision
  - Quality of data
  - Available data/information
 - Type of data
  - Project schedule and funding
  - Etc.

Boundary Conditions Data Gap (I)

- Managing daylight saving time

- Simple, short data gap
  - Last Best Value
  - Average
  - Linear interpolation
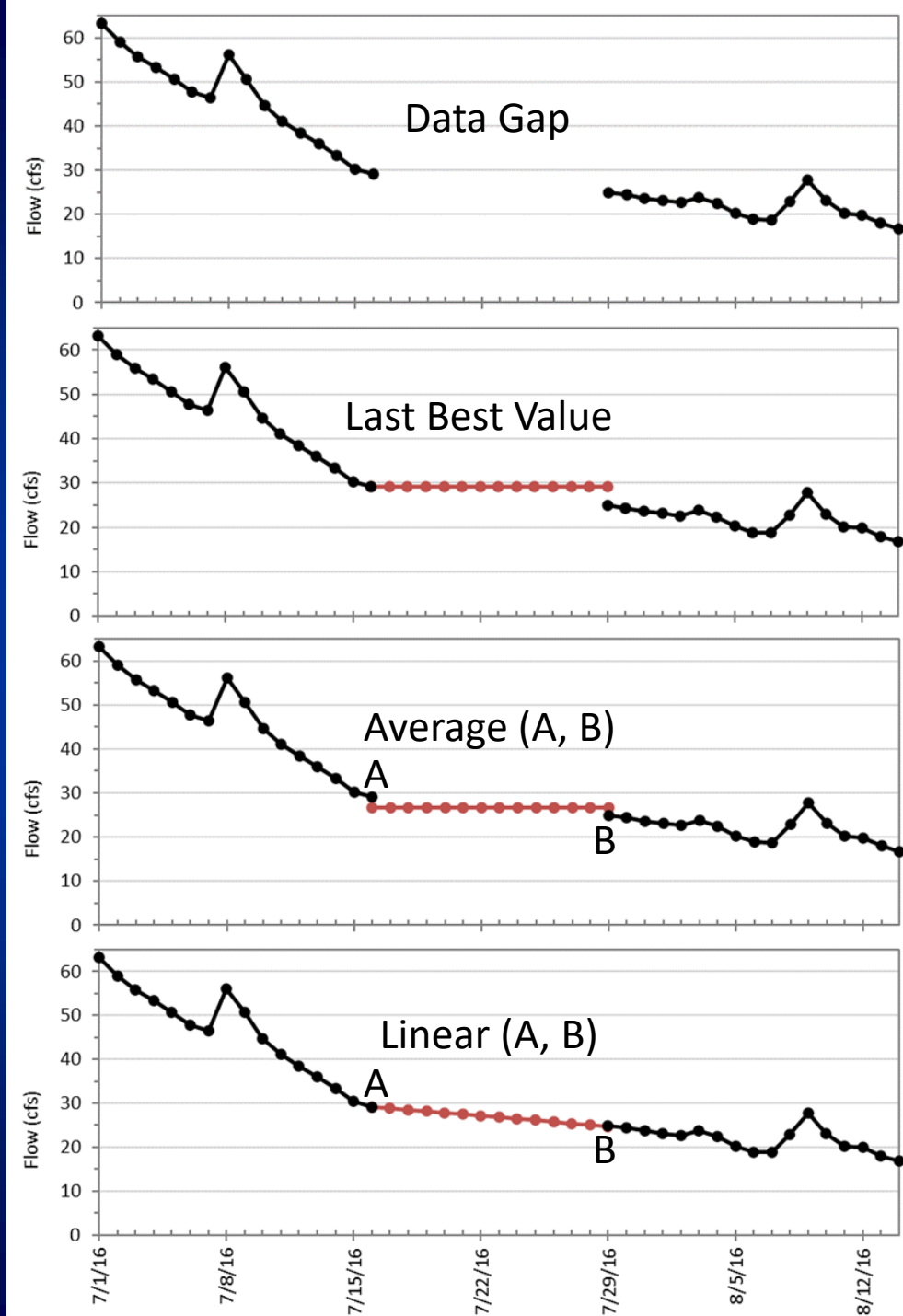
Boundary Conditions Data Gap (II)

- Dynamic response or longer data gap
- Capture system response
  - "Modeled" approaches
    - Statistical/Spatial
      - Data relationships (basin area ratios for flow)
      - Regressions
      - Statistical models/approaches
      - GIS models (proration)
      - Specific spatial and temporal models (e.g., PRISM)
    - Process or mechanistic models
      - Flow (watershed models, hydrology models)
      - Temperature (equilibrium temperature)
      - Water Quality (Dissolved oxygen (saturation))
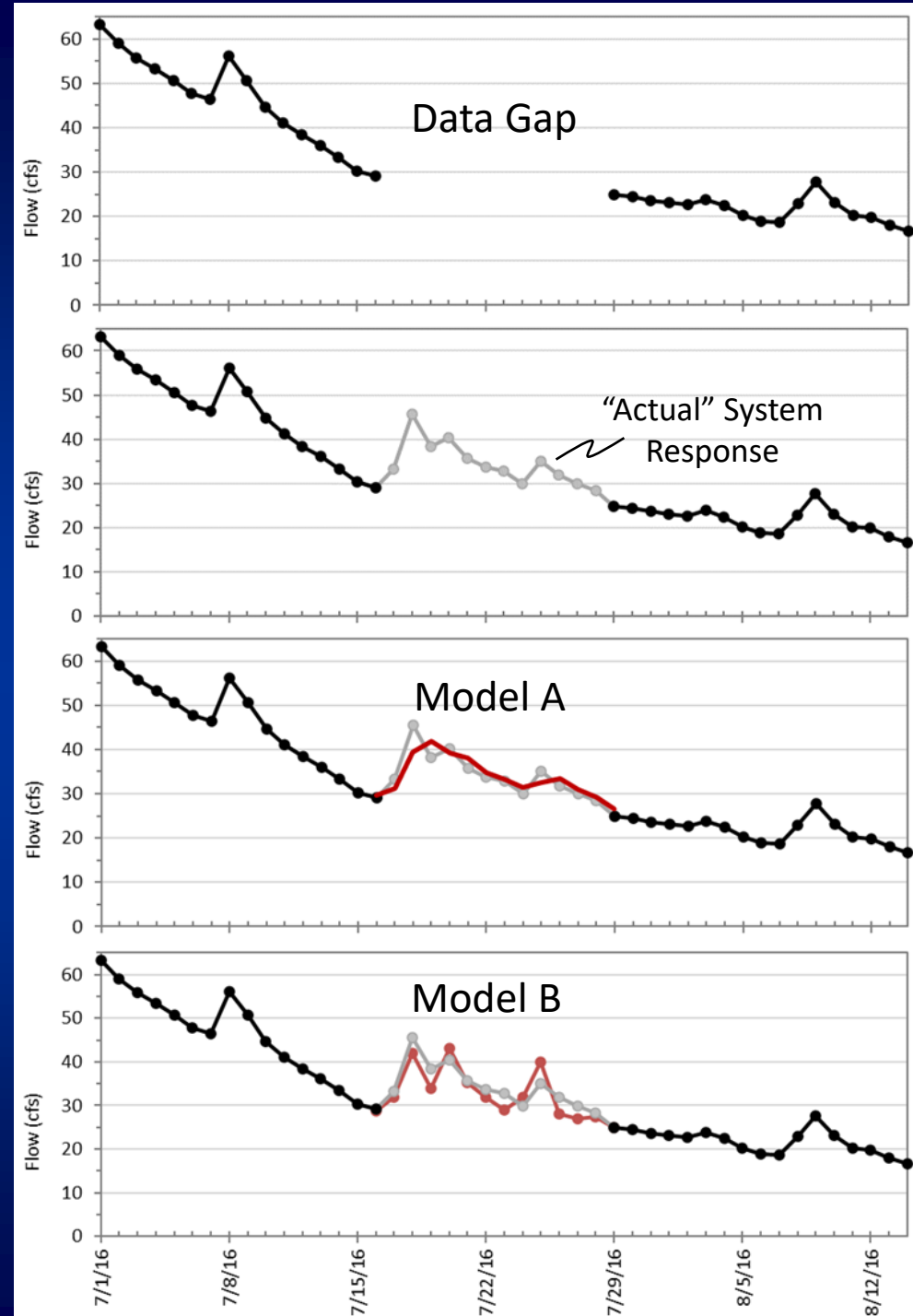
Boundary Conditions (III) Missing Data

- Definition: No available data during modeling period
  - Historic data
  - Adjacent watersheds/similar systems
- Data type considerations
  - Flow
  - Temperature
  - Water Quality (laboratory vs WQ probe, sparse)
  - Meteorology (spatial proximity/representation)
- Consider "long data gap" approaches or specific models.
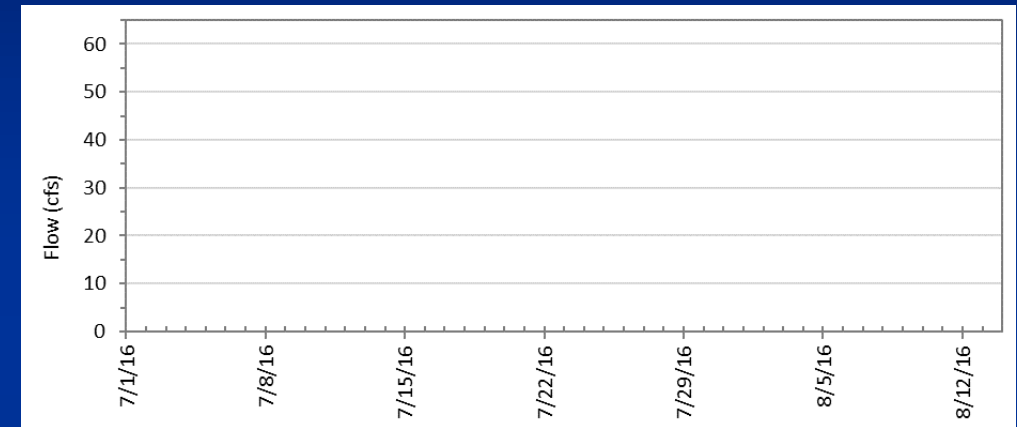
Calibration Data

- Calibration data are not typically "filled"
  - Typically, do not calibrate to estimated data

- Data quality assurance (QA)
  - Daylight savings time
  - QA includes removal of "non-data" delimiters
    - -999, dashes, flagged, or censored data

- Assess model performance based on available data (e.g., performance statistics)

Data Flag Definitions
Sorted by Data Flag

UPDATED: 07/18/2019

Data Flag	Flag Description
	No Flag
A	Precipitation accumulation
L	Waiting for observer response
N	Error in data
e	Estimated
q	New rating table
r	Revised
s	New shift started
t	Trace of precipitation
v	Out of Valid Range

https://cdec.water.ca.gov/

Site	Date	Time	NH_4^+	TKN	$NO_3^- + NO_2^-$	TP	PO_4^{3-}	BOD	Chlor_a
			(mg/l)	(mg/l)	(mg/l)	(mg/l)	(mg/l)	(mg/l)	(ug/l)
TLO	8/22/00	10:30	0.22	3.9	0.07	0.18	0.17	5.0	15j
KSDH	8/22/00	11:30	0.1	3.9	<0.05	0.17	0.17	<3.0	2j
KSD97	8/22/00	8:45	0.24	5.0	1.22	0.31	0.29	7.0	15j
KRMI	8/22/00	13:30	0.91	2.8	0.18	0.08	0.06	<3.0	6j
KRK	8/22/00	12:30	1.47	3.5	0.10	0.14	0.16	<3.0	17j
KRCR	8/22/00	14:55	<0.05	0.8	0.36	0.08	<0.05	<3.0	2j
KRIG	8/22/00	13:35	<0.05	1.3	0.24	0.13	0.17	<3.0	6j
KRSV	8/22/00	10:30	<0.05	1.6	0.11	0.12	0.14	<3.0	7j
SHR	8/22/00	12:35	<0.05	0.6	<0.05	0.14	-	<3.0	0j
SCR	8/22/00	7:55	<0.05	0.5	0.12	<0.05	<0.05	<3.0	3j

(<) less than reporting limit
(j) below reporting limit of 40 ug/l
(-) no data available

Importance of Documentation and Metadata

- Data gaps are the rule rather than the exception to develop "model ready" data
- Documentation
  - Available data (all meta data: source, method, agency/owner, other)
  - Gap filling (comprehensive)
  - Any testing (graphical, statistical, model)
  - Recommendations
- Role of a data management system that can track metadata
  - Many methods – some more robust (or fragile) than others
  - Required resources for development and maintenance = commitment (that is not always feasible)

Discussion

- Should CWEMF develop specific guidance on data gaps and management in a future version of the protocols?

- If so, what level of detail may be appropriate?

- Are there other considerations?