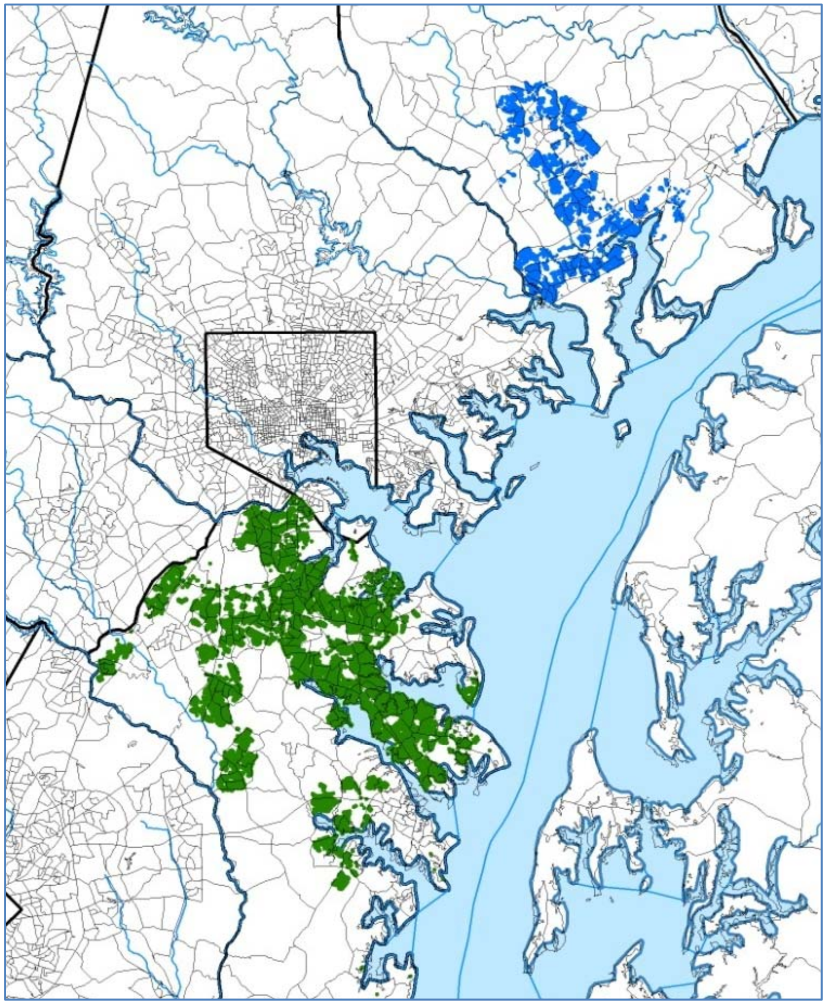
Maryland-based domestic per capita water demand regression models

Betzy Reyes

breyes@usgs.gov

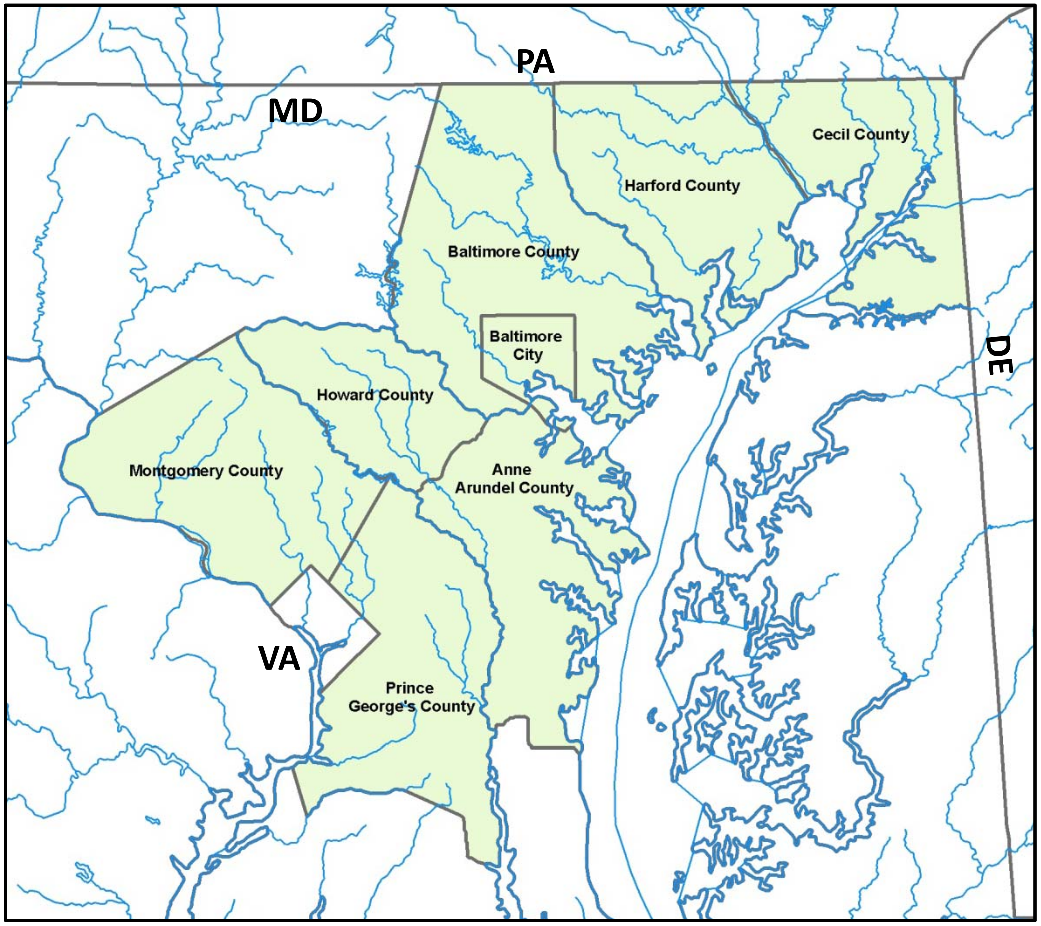
Water-Use Workshop April 19-20, 2010

Model variables –Background



- U.S. Bureau of Census block and block-group data for 2000 (road rectified data)
- Block and block-group levels data
- 2000 Maryland Census data
- Climate data
- Land use data
- Counties metered deliveries from 2005-2006

Metered Water-Use deliveries

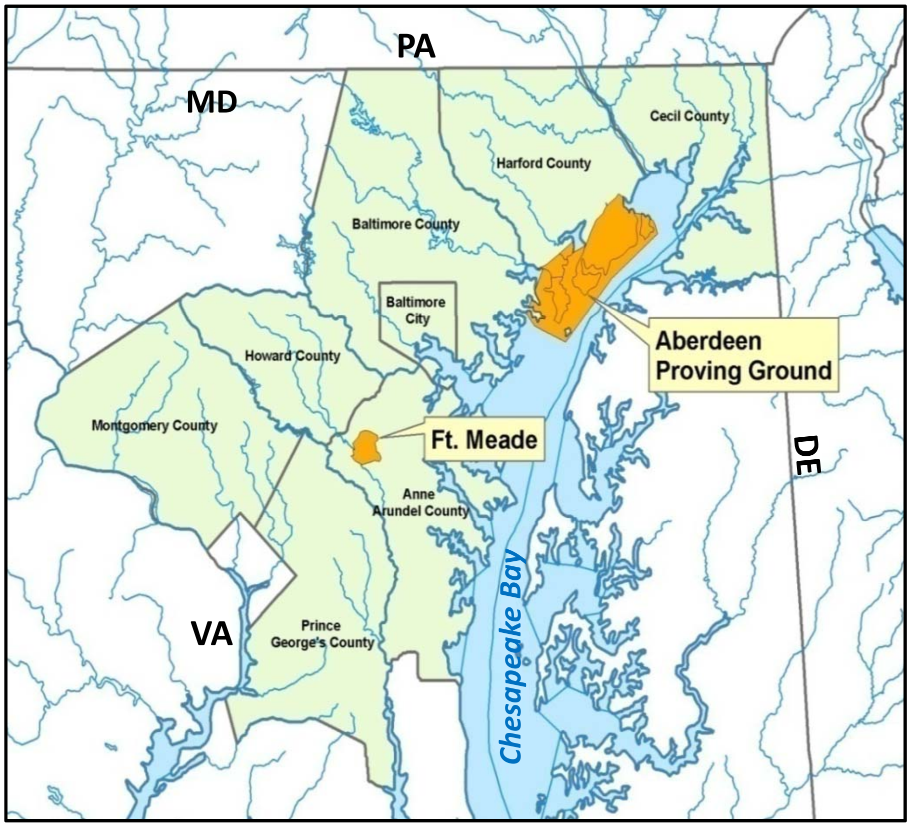


- 2005 and 2006 deliveries

- MD BRAC Study areas:

-Cecil County
-Harford County (HF)
-Baltimore County
-Howard County
-Anne Arundel County (AA)
-Montgomery County
-Prince George County
-Baltimore City

Metered Water-Use data



- Water-Use (WU) data available:
  - **-Harford County (HF)**
  - **-Anne Arundel County (AA)**

- AA and HF because:
  - -Ft. Meade and Aberdeen Proving Ground
  - -Counties were very cooperative
  - -Data sets were accessible

•Initial QA/QC for the accounts:

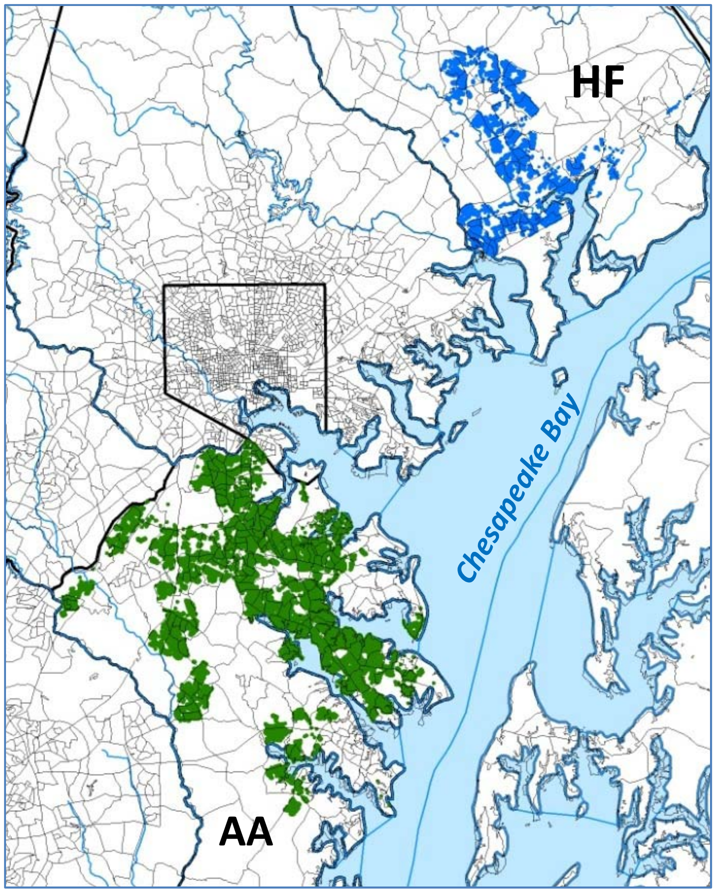
-Quarterly readings -**Domestic** WU

-**8** consecutive readings (two full years of meter readings)

-WU metered data were in thousand of gallons (Kgals.)

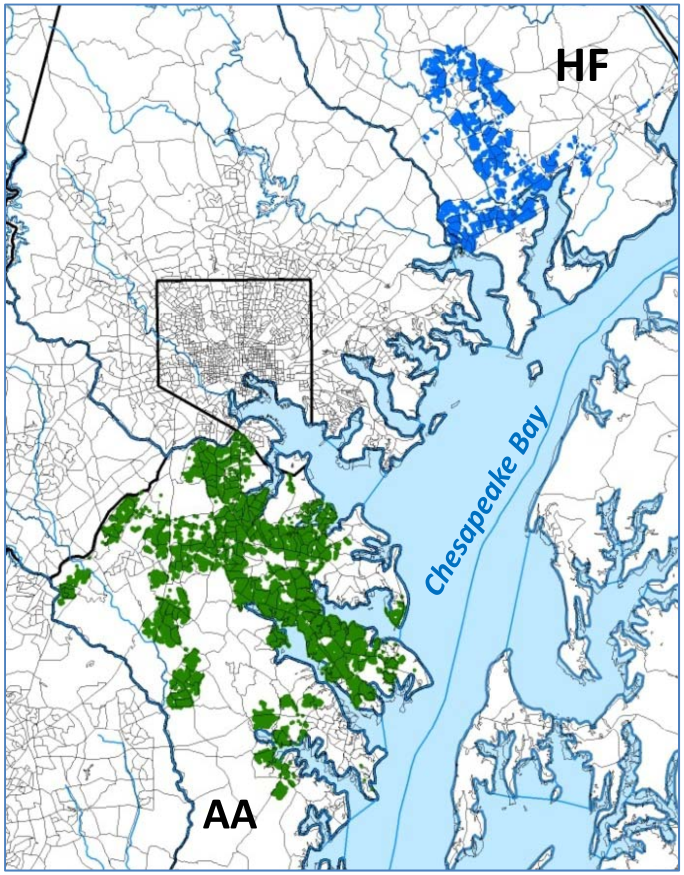
-Metered delivery data were related to U.S. Census block and block-group data designed to represent socio-economic conditions that may have an influence on water demand

Block sample selection for the models



- Compare housing densities between the blocks with and without meters data and there was no significant difference.
  - Got housing densities quartiles for the Census blocks with meter readings
  - Random sample of 350 blocks in each quartile using the census block identification number (ID)

Metered Water-use data QA process



•Population per housing unit was adjusted by the seasonal housing effect.

• Metered WU readings in the blocks selected were analyzed to see if they stayed in a difference of less than a 1 order magnitude between years for their readings.

For the models:

•2,016 blocks (1,008 blocks each year)

•113,246 number of meters

•**Ratio** of blocks/meters = 0. 0178019

Maryland-based domestic per capita water demand regression models

- Metered deliveries: 2005 and 2006 years

- Per capita water-demand models were developed to determine Annual, Summer, and Winter per capita water-demand coefficients:

1. Annual Model: Annual metered deliveries for both years

2. Summer Model: June, July, and August metered deliveries

3. Winter Model: December, January and February metered deliveries

- All census blocks used to develop the model were weighted on the basis of the number of metered accounts in the census block

Maryland-based domestic per capita water demand regression models

- Dependent variable for Models:

First step:

- averageGPD = ((total Kgals.WU from meters per block /number meters in that block) / **days) (1,000)

***days* per period due that are quartiles of the year:

- 365 days for Annual Model
- 91 days for Summer Model
- 90 days for Winter Model

Second step:

- Per capita (PC) = (averageGPD)/Population_HUB_Adj

Average Gallons per day per person per house adjusted by seasonality

Maryland-based domestic per capita water demand regression models

All statistical analyses used to build the models were done with SAS.

Significant predictor variables were identified if they had probability values (*p*-values) equal to or less than 0.05.

The natural log of the per capita water-demand coefficient was predicted in the model to account for the heteroscedastic nature of the residuals of non-transformed predictions.

The annual, summer, and winter per capita water-demand models can be expressed as:

$\ln(PC) = \beta_0 + \beta_1 (X_1) + \beta_2 (X_2) + \beta_3 (X_3) \ldots \beta_n (X_n) + E$ Where,

ln(PC)	=	natural log of the domestic per capita water-demand coefficient, census-block value for gallons per day per person;
B_0	=	intercept,
$\beta_1, \beta_2, \beta_3$	=	variable coefficients,
X	=	independent variable, and
E	=	random error

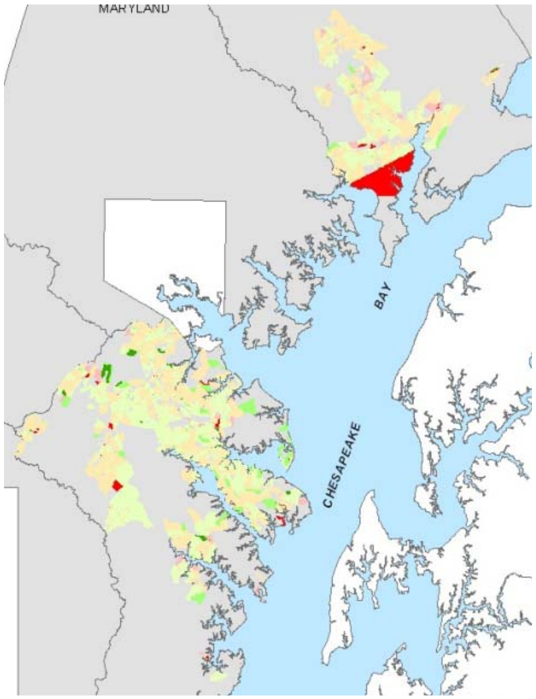
Regression model results

Note: ***p-values < 0.0001 for all coefficient predictors***				**Annual Model** $R^2 = 41.21$ RMSE = 0.18578		**Summer Model** $R^2 = 35.24$ RMSE = 0.23450		**Winter Model** $R^2 = 33.68$ RMSE = 0.18182	
Variable (Source)	**Units**	**Range**	**Median**	**Coefficient**	**Standard Error**	**Coefficient**	**Standard Error**	**Coefficient**	**Standard Error**
Intercept	Dimensionless			**4.85584**	0.03565	**5.00972**	0.04499	**4.81505**	0.03488
Median House Value Owner Occupied (BG)	Dollars / 1000	67,500 - 455,500	149,200	**0.00207**	0.00008	**0.00244**	0.00016	**0.00121**	0.00008
Housing Unit Density (B)	Units per square mile	2-30,037	1,236	**-0.00001975**	0.000002	**-0.00002637**	0.000003	**-0.00001117**	0.000002
Median Year Constructed (BG)	Median year built - 1900	39-98	72	**-0.00401**	0.00035	**-0.00614**	0.00044	**-0.00281**	0.00034
Population per housing unit (B)	people per unit	1-5	2.8	**-0.23939**	0.00925	**-0.21065**	0.01168	**-0.24884**	0.00906

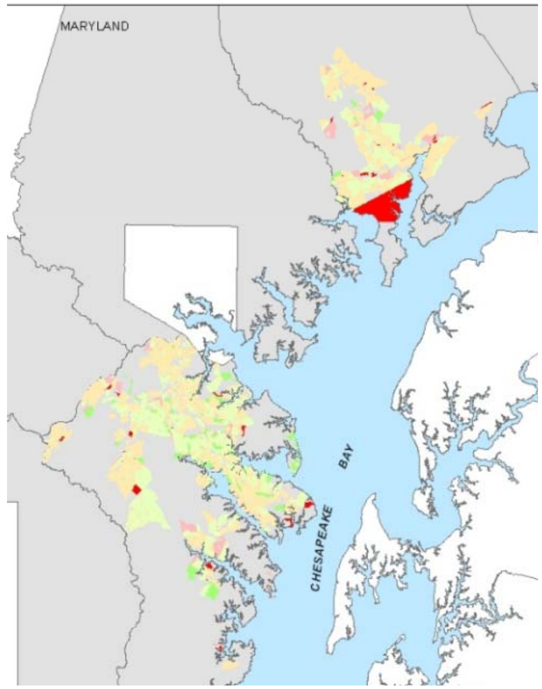
- Same significant predictors for the Annual, Summer, and Winter models

- The R^2 results indicate that the per capita water-demand models were able to account for about 40% in the variation in the metered community water systems deliveries from HF and AA counties to domestic users.

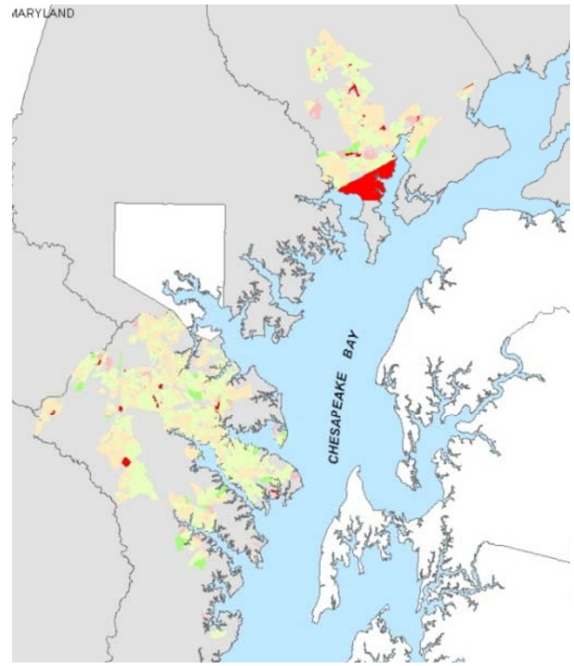
Residual Distribution



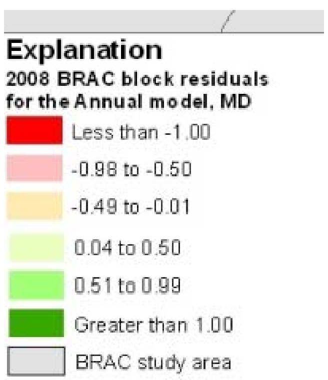
Annual Model Residuals Distribution



Summer Model Residuals Distribution

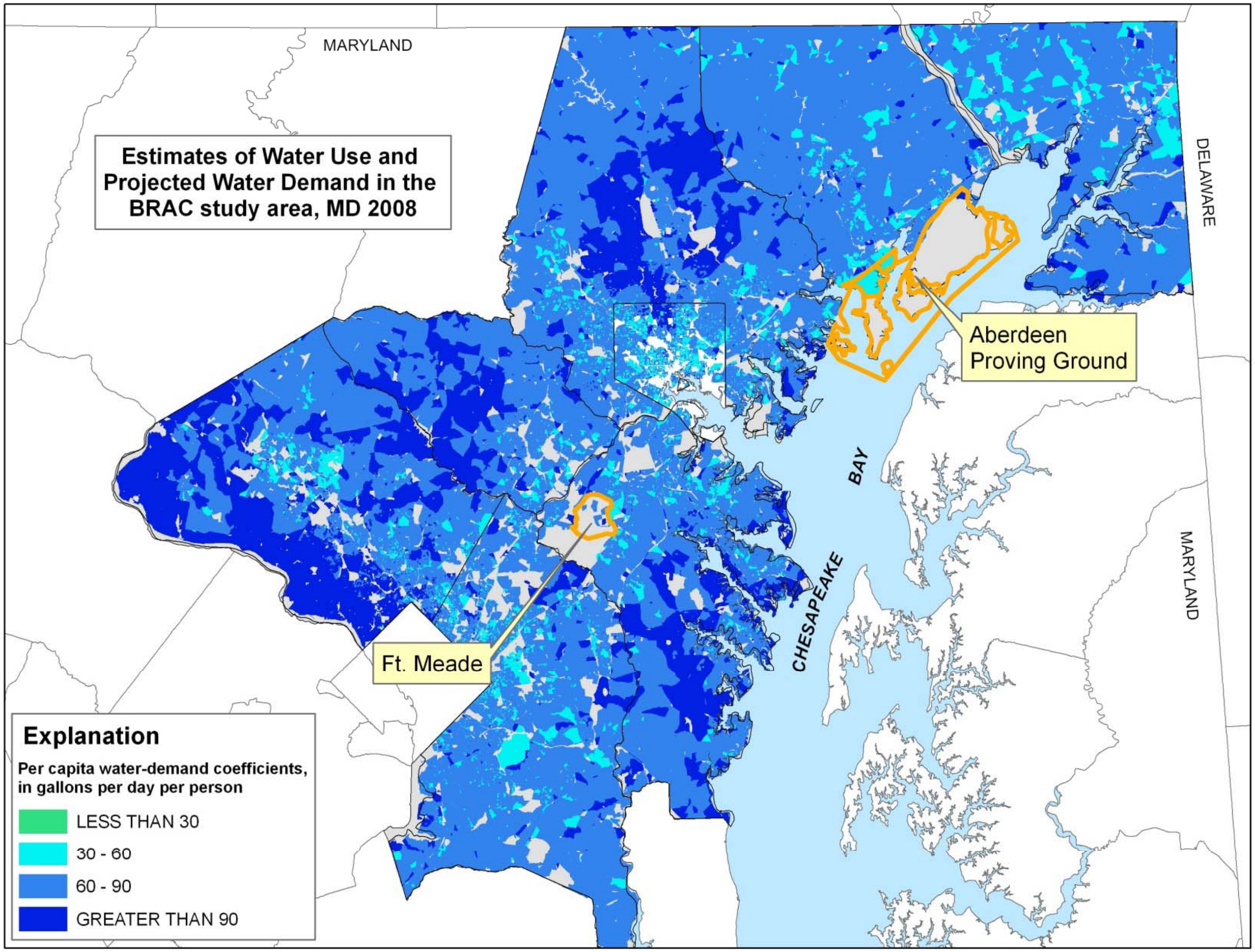


Winter Model Residuals Distribution



- Appear to be randomly distributed study area.
- Showing no spatial grouping of over- and under-predictions.

Per capita Annual water demand estimations



Model Strengths and weaknesses

Some Strengths :

(1) Assumptions defined by the Multiple Linear-Regression analysis;

(2) Residual analysis for each of the three models once the transformation of the results to the inverse of the log value, indicates that the residuals appear to be randomly distributed across the MD BRAC region with no spatial grouping of over- and under- predictions;

(3) Assumption that domestic users that depend on smaller Community Water Systems (CWS's) and private wells use water in a similar manner to those on large CWS's ;

(4) Models include the ability to provide regionally consistent characterizations of domestic per capita water demand on an annual and seasonal basis, and to provide confidence intervals associated with these assessments.

Model Strengths and weaknesses

Some limitations:
(1) Based on the data used in the modeling process (some predictor variables at block-group level which may tend to limit the range of the predicted results).

(2) R^2 values (of 41%, 35% and 34%), which indicates that additional factors are influencing variations in domestic water demand that could not be explained by the model. This may lead to an oversimplification of the social, economic and policy/political variables that influence how water is used.

(3) Needs to be acknowledged that many factors locally and regionally affect the per capita water demand, many of which are not accounted for in the per capita water-demand model or in the census data used to develop the model. Additional analysis of parameters related to climate (look more in detail), cost of water, watering restrictions, and landscape development variables, highly urbanized areas, may provide further insight as to factors influencing domestic per capita water demand.