HSPF Application to Microbial Indicator Data

Trent Schade, Katharine G. Field, Orin C. Shanks
and Matthew A. Morrison

From March 2001 to March 2003 researchers at the Tillamook Estuaries Partnership and Oregon State University collected water quality samples at regular two-week intervals at 30 sites along five rivers in the Tillamook Basin in northwestern Oregon. The unique dataset consists of *Escherichia coli* counts paired with ruminant- and human-specific 16S rRNA gene host-specific polymerase chain reaction (PCR) data (Bernard and Field 2000) that allow for the direct comparison of regulated fecal indicator bacteria (FIB) contaminant levels with fecal source tracking information. The focus of the current study was on five sampling stations in the Trask River watershed, where two USGS gauging stations provide a long-term hydrologic record. Using regular time intervals in a water quality sampling protocol produces a set of data randomized in terms of the hydrological conditions for each sampling event. With this water quality data, we produced a “dosing” type of analysis to make broad inferences about the type, source and pathways for water quality impairment. Additional analysis of the entire dataset showed that during wet periods, there was a positive correlation in some locations between flow and FIB and host-specific PCR data. The current study applied a Hydrological Simulation Program Fortran (HSPF) model of the Trask River Watershed portion of the Tillamook Basin in order to provide a more mechanistic understanding of the fate and transport of fecal pollutants. HSPF can distinguish at short time intervals (hours) over long period (decades)

Schade, T., K. Field, O. Shanks and M. Morrison. 2009. "HSPF Application to Microbial Indicator Data." *Journal of Water Management Modeling* R235-02. doi: 10.14796/JWMM.R235-02.
© CHI 2009 www.chijournal.org ISSN: 2292-6062 (Formerly in Conceptual Modeling of Urban Water Systems. ISBN: 978-0-9808853-2-3)

between flow originating as surface runoff, interflow, and groundwater flow. Applied to the microbial sampling data, the separated hydrographs may yield a stronger relationship between FIB and PCR data and hydrologic variables. Further, this work may suggest modifications for the type and placement of best management practices for both human and animal sources of microbial contamination.

2.1 Background

Figure 2.1 shows the lower portion of the Trask River watershed, which is located within the Tillamook Basin in northwestern Oregon. The headwaters of the Trask River are about 20 miles west of Portland, in the mountainous terrain of the Oregon Coast Range. The watershed is approximately 20 miles (32 km) long, rising in elevation from the Tillamook Bay to mountain peaks close to 2700 feet (823 m) MSL. The climate for the Oregon coastal region is typical of the Pacific Northwest. Rainfall is seasonal, with a very wet season October through April, and a dry season through the summer months, May through September. Mean annual rainfall is 90 in./y (2300 mm/y) though this number is variable with a standard deviation of 19 in. (480 mm). Pan evaporation rates for the Oregon coastal region are about 30 in./y (760 mm/y) (Oregon Climate Service, 2008).

From March 2001 to March 2003 researchers at the Tillamook Estuaries Partnership and Oregon State University collected water quality samples at regular two-week intervals for 30 sites along five rivers in the Tillamook Basin in northwestern Oregon (Shanks et al., 2006). The unique dataset consisted primarily of *Escherichia coli* counts (MPN/100 ml, measured using Iddex Colilert) paired with presence/absence of ruminant- and human-specific 16S rRNA gene host-specific PCR data (Bernhard and Field, 2000) that allow for the direct comparison of regulated fecal indicator bacteria (FIB) levels with fecal source tracking information. This dataset represents one of the largest efforts to implement a field study to identify fecal sources of contamination using advanced DNA analysis techniques. In addition to the microbial data (PCR and FIB), the researchers recorded temperature, pH and turbidity at each sampling event.

The design of the data set was a typical “dosing” study, using a regular sampling event cycle to produce a random set of hydrologic conditions in the river at the time of sampling. These studies can provide information about seasonal trends in water quality for the area; they can provide a “typical dose” for a single frequency of contact; and they may show

relationships between hydrology and water quality parameters. In some cases, it may be possible to estimate loading rates for the watershed, especially when the sampling frequency is very high. Often these studies are accompanied by rainfall-triggered sampling events to try to capture the water quality parameters at their peaks.

Figure 2.1 Mouth of Trask River.

A previous analysis of the entire Tillamook Basin data set showed a seasonal trend in water quality loading from ruminant sources (Shanks et al., 2006). The objective of the current study, of which this chapter is only the first part, is to develop a stronger relationship between watershed hydrology

and DNA and FIB data for the Trask River through the use of the HSPF (Hydrological Simulation Program Fortran) model. This chapter outlines the development and calibration of the HSPF model for the Trask River watershed, and highlights the utility of HSPF's hydrograph separation technique. The study further evaluated the influence of season on the predominant hydrograph response (surface runoff, interflow or groundwater flow) and showed the potential for better explanation of relationships between microbial and water quality variables. With further refinement, and possibly more localized modeling, this information may be useful for selecting agricultural nutrient management techniques, urban stormwater best management practices, and sewer overflow abatement.

2.2 Modeling Description

HSPF is one of the analysis engines available from the U.S. Environmental Protection Agency (EPA). It is a watershed model based on the Stanford Watershed Model, and is generally used for large watershed management applications including hydrology, water quality, and sediment studies. HSPF has extensive capabilities for modeling water quality, including tools for sediment loading calculations and fate and transport modeling. For this study, we configured HSPF to model the Trask River hydrology from the headwater locations to the USGS gauge number 14302480 described in the data sources section above. We utilized a supervised subcatchment delineation tool available in EPA's BASINS software package and based on the TauDEM algorithm (Tarboton, 1997). The sixteen subwatersheds above the USGS gauge site have a combined area of approximately 150 mi² (390 km²), consistent with the gauge description from the USGS (USGS, 2008).

The hydrologic model takes precipitation, temperature, and potential evaporation time series as inputs. Precipitation can occur as snow or rain depending on the temperature, and the temperature data is also used to determine the rate of snowmelt. We found that modeling snow in this system was critical. The conceptual model of HSPF hydrology is very simple. Rainfall or snowmelt enters either a pervious or impervious land segment to begin the hydrologic processes. Water from an impervious segment either evaporates or is routed to a river reach. A pervious segment is more complicated. Figure 2.2 shows the conceptual model for a pervious segment. Water can leave the land segments as evapotranspiration (ET), surface runoff (SURO), interflow (IFWO), groundwater (AGWO), or deep percolation (IGWI). If it leaves as ET, it is gone from the system. If it is

SURO, INFO, or AGWO, it is routed to a river reach section for hydraulic calculations. Deep percolation is counted as a loss from the system. Thus, HSPF can effectively separate a hydrograph into its component parts. For this study, we used HSPF to provide an accounting for each of the three types of water – surface runoff, interflow, and groundwater – that enter the Trask River system and eventually Tillamook Bay.

Figure 2.2 HSPF conceptual hydrologic model.

The hydrologic analysis includes the following steps.

1. Determine Project Location
2. Acquire land surface spatial data including:
 - Elevation Surface
 - Soils Data
 - Landuse Data
 - Stream network
 - Sampling locations
 - Raingauge locations
 - Streamgauge locations
 - Populated areas
 - Concentrated Animal Feeding Operations (CAFOs)
 - NPDES Permitted outfalls
3. Delineate Subcatchments

4. Calculate Subcatchment Properties
 - Slope
 - Landuses
 - Soil types
 - CAFOs distribution
 - NPDES outfall distributions
5. Acquire Time series for Climate Conditions
 - Precipitation Time series
 - Daily Maximum Temperature Time series
 - Daily Minimum Temperature Time series
 - Create Hourly Airtemp time series
 - Create Hourly Potential Evaporation Time series
6. Analyze Time series
 - Determine analysis periods based on available data
 - Determine raingauge weighting
 - Build time series into model environment
7. Build Model
 - Initial Parameterization (based on local estimates and watershed characteristics)
 - Sensitivity Analysis
 - Calibration/Verification
 - 10 y log plot
 - 1 y arithmetic plot
 - Single event plot
 - Flow-duration curve
 - Scatter plot (regression fit)
8. Analysis
 - Hydrograph separation
 - Simple water quality correlations
 - BMP Trials

2.3 Model Data Sources

EPA BASINS includes geospatial data that shows the location and type of most water quality loading sources as well as base data required for basin-scale hydrologic study. For the Trask River study area, we used only a small part of the data available from BASINS. The source data included with the

BASINS software is very easy to use and manipulate. We obtained additional data directly from EPA, from the National Weather Service, and USGS. Table 2.1 shows the spatial data used, the source agency, and the website.

Table 2.1 Spatial data sources.

Data	Source	Website
Elevation Surface	National Elevation Dataset	www.epa.gov/waterscience/basins/metadata/ned.htm
Soils Data	STATSGO	www.epa.gov/waterscience/basins/metadata/statsgo.htm
Landuse Data	GIRAS	www.epa.gov/waterscience/basins/metadata/giras.htm
Stream network	RF1 and RF3	Available from the BASINS base files by HUC here: www.epa.gov/waterscience/ftp/basins/gis_data/huc/
NPDES outfalls	Permit compliance System	
Populated areas	City Boundaries	
Streamgauge locations	USGS	waterdata.usgs.gov/usa/nwis/uv?site_no=14302480
Raingauge locations	National Weather Service	dipper.nws.noaa.gov/hdsb/data/archived/legacy/hlytran.html
Sampling locations	U.S. EPA	N/A
CAFOs Locations	U.S. EPA	N/A

The HSPF model requires a minimum precipitation time series as input and, since we modeled snowmelt, we also required hourly measurements of air temperature and potential evaporation time series for the period of analysis. Local raingauges provided the hourly precipitation data. There is a raingauge centered in the Trask River watershed, but because of the variability in terrain, we utilized an inverse distance weighting scheme to combine data from the Trask raingauge and six surrounding gauges.

For temperature data, BASINS has some utilities to help estimate hourly air temperature data from daily maximum and minimum temperature time

series. For the calibration runs, we used daily maximum and minimum temperature data to calculate the hourly air temperature estimates based on a sinusoidal curve. From these we could estimate hourly potential (pan) evaporation. The snowmelt model in HSPF, which relies on temperature data, also has an elevation correction based on the adiabatic lapse rate. From the spatial elevation data, we calculated an average subcatchment elevation to apply the lapse rate.

2.4 Hydrograph Processes

One of the analysis goals for HSPF is to obtain the relative weights of flows in each of the stream inflow source categories: SURO, IFWO and AGWO. The impervious land segments only produce SURO for water that does not evaporate. In the Trask River watershed there is very little impervious land surface, and much of it is below the USGS gauge station where we had observed data for the model calibration. The pervious areas implement a more complicated hydrology to estimate flows from each of the three source categories. There are some 22 parameters to calibrate each pervious land segment. Many of these parameters may be estimated, and EPA provides extensive guidance (USEPA, 2000), expert systems (USEPA, 1999), and parameter estimation tools (Doherty, 2004) on selecting and calibrating these parameters.

Table 2.2 Calibration periods.

	% Available	Frequency Available
Calibration 10/1/1961 to 9/30/1971		
Flow at USGS Gauge	100	Daily
Precipitation	100	Hourly
Air Temp (Max/Min)	100	Daily
Verification 10/1/1996 to 9/30/2003		
Flow at USGS Gauge	100	Hourly
Precipitation	100	Hourly
Air Temp (Max/Min)	100	Daily
Sampling Period 3/14/2001 to 3/5/2003		
Flow at USGS Gauge	100	Hourly
Precipitation	100	Hourly
Air Temp (Max/Min)	100	Daily

We calibrated the model by first identifying two periods during which we had a complete record of rainfall, temperature, and flow data. Table 2.2 shows the periods selected for the hydrologic calibration, verification, and the sampling period. The calibration period provided a 10 y period during which we had a complete record for all three time series. The flow time series record interval for this period is daily; this is the USGS calculated mean daily flow. For the verification period we selected a time during which we had hourly flows available and that overlapped the sampling period.

The calibration process was a trial-and-error progression of compromises. We started the calibration with parameters estimated from the BASINS data set, and gradually adjusted them until the results for flow matched the observed values. The evaluation was done by comparing plots visually at different time scales, and with the y-axis displayed in either an arithmetic or logarithmic scale. Figures 2.1-2.3 show examples of the visual clues used in calibration of an HSPF model. Figure 2.3 depicts observed versus modeled flow over ten years. By reviewing a long trend on a logarithmic scale, the user can assess general trends in system totals. This information helps to calibrate the soil types, thickness of soil layers, groundwater response (recession and deep groundwater losses), evapotranspiration (ET) parameters, and raingauge adjustments.

Figure 2.3 Observed and modeled flow for ten-years on logarithmic scale.

Figure 2.4 shows observed versus modeled flow, but now the data is displayed on an arithmetic scale for a single year (1966). This was used to correct the baseflow recession and interflow recession, and for calibrating both infiltration rates from the surface to the interflow storage and the

partitioning between the surface and the lower zones. The general trend in this model seemed to underestimate some peak flows. However, the hydrograph shapes were exceptional.

Figure 2.4 Observed and modeled flow for one-year on arithmetic scale.

Figure 2.5 Observed and modeled flow for one month on arithmetic scale.

Figure 2.5 portrays individual storm events in closer detail. The plot of a single storm, or a few days worth of rainfall, made the differences between the model and measured data very apparent. With the calibration process, adjusting parameters moves water between the different types of outflow.

For example, too much IFWO gave the model data a longer tail and less direct SURO. Adjusting this affected summertime peaks. We often found that when we corrected one response, another response was impacted. For this study, the focus was on getting the hydrograph shapes correct. For the sampling periods that we corrected for the volume of flows, we were less concerned that the total magnitude of the flows was correct. Our goal was to obtain the correct hydrograph shape, which would allow us to understand the hydrograph separation well. Similarly, the peak and low-flow conditions were not as central to the analysis as the hydrograph shape.

Figure 2.6 % chance of flow exceedence.

Figure 2.7 Recession curve measured and modeled flows.

Once the calibration parameters were acceptable, we used them on our validation dataset. Figures 2.6 and 2.7 present some summary plots to demonstrate the performance of the model. Figure 2.6 shows the % chance of flow exceedence for both the model and the observed data. In this case, the model underestimated the peak flows (less than 1% exceedence), but exhibited a very strong fit everywhere else. Figure 2.7 shows a more traditional correlation plot. Again the trend is clear that the model underpredicts the peaks for high flow events.

2.5 Hydrograph Separation

The standard output from HSPF includes time series for each hydrograph component, surface runoff, interflow and groundwater. To produce Figure 2.8 we split the observed data according to the ratios of hydrograph components (SURO/total flow, IFWO/total flow and AGWO/total flow) in the model for that period. This algorithm made it necessary to focus the calibration on matching the hydrograph shapes.

Figure 2.8 Observed time series with estimates for hydrograph separation. SURO is shown in gray, IFWO in white and AGWO in black.

Figure 2.9 shows the flow exceedence (or flow-duration curve) of the complete calibration run, as well as the seasonal trends in the runoff. The figure makes it clear how different the component statistics and processes were for the wet weather versus the dry weather. During the summer, the

Figure 2.9 Flow exceedence percentages for calibration time period.

interflow process exceeded 3 cms (cubic meters per second) a little more than 20% of the time, and about 60% of the time interflow was zero. In the winter, the interflow had a much more dominant role. Interflow exceeded 3 cms about 90% of the time and it was rarely zero.

The model demonstrated that once the soil layers become saturated in fall, they stay saturated until late spring or early summer. The cutoff appeared to be around 300 cms (1,000 cfs). Therefore, once the interflow for a single event reaches 300 cms in the fall, there will be continuous interflow for the rest of the season. A similar inference can be made about the return to summer conditions. This will occur when groundwater (probably total flow) drops below 30 cms (100 cfs). This suggests that saturated soils will transport fertilizer, pathogens and nutrients away from the root zone to the receiving water in the spring until the groundwater drops to that level.

2.6 Water Quality Analysis

A previous analysis of the water quality data used sample month to determine wet versus dry period; Shanks and colleagues (2006) defined the dry period as the six months of lowest average monthly precipitation, April through September, and October through March as the wet season. The current analysis based on modeling suggested that the wet period of the year is when interflow first exceeds 300 cms after the summer, and the dry period begins when the groundwater level drops below 30 cms. Applying these revised criteria to the modeling results allowed us to compute a shorter dry period--from the beginning of August to the end of November (four months). The wet period is December to July. However, with these criteria, the length of time for the dry period is variable each year. There was a 36 day difference in the dry period for the first year of sampling (August 11, 2001 to November 14, 2001—95 days) compared to the second year of sampling (August 2, 2002 to December 11, 2002—131 days). Table 2.3 demonstrates the benefit of a hydrology-based criteria for wet versus dry period analysis. The calendar based analysis shows more frequent hits for both ruminant sources and human sources of microbes during dry weather. Though the difference is small the result is counter to what the physical processes should produce. Using a hydrology-based criteria, the wet period has more samples with hits for both Ruminant and Human sources. Again, this matches generally accepted rainfall based loading models.

Table 2.3 Relative differences in FIB and ruminant host-specific PCR statistics.

Dry as April to September		Wet as October to March	
Samples w/ ruminant hits	84.0%	Samples w/ ruminant hits:	61.7%
Samples w/ human hits:	44.0%	Samples w/ human hits:	35.0%
Mean <i>E. coli</i> (count/100mL):	302.2	Mean <i>E. coli</i> (count/100mL):	219.7
Dry as first groundwater < 30 cms		Wet as first interflow > 300 cms	
Samples w/ ruminant hits:	70.6%	Samples w/ ruminant hits:	78.7%
Samples w/ human hits:	34.7%	Samples w/ human hits:	50.7%
Mean <i>E. coli</i> (count/100mL):	241.1	Mean <i>E. coli</i> (count/100mL):	268.4

2.7 Conclusions

The HSPF model was used for the computation of three separate sources of flow to the Trask River: surface runoff (SURO), interflow (IFWO) and groundwater flow (AGWO). The ability to discern between different sources allowed for a more sophisticated characterization of wet and dry periods in the Trask River watershed. The HSPF model analysis and generated flow data can be used to further understand the mechanisms of pollutant transport, and will improve comparison of FIB, host-specific PCR data and general water quality parameters. Future research will explore these comparisons using a duration-curve approach for the entire Trask River dataset. From a practical perspective, more accurate data for the onset of wet and dry periods in a watershed can be used to create more effective nutrient management plans and improve the performance of management practices in the Tillamook Basin.

Acknowledgments

The authors would like to thank the Tillamook Estuaries Partnership Volunteer Water Quality Monitoring Program and Jim Bonta of the Agricultural Research Service North Appalachian Experimental Watershed Station.

References

- Bernhard, A. E., and K.G. Field. 2000b. A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. *Applied Environmental Microbiology* 66:4571-4574.

- Chen, Y.D., S.C. McCutcheon, R.F. Carsel, A.S. Donigian, Jr., J.R. Cannell, and J.P. Craig. (1995) "Validation of HSPF for the Water Balance Simulation of the Upper Grande Ronde Watershed, Oregon, USA. In: Man's Influence on Freshwater Ecosystems and Water Use." G. Petts (ed). International Association of Hydrological Sciences, IAHS Pub. No. 230. pp 3-13.
- Doherty, J. Consulting. (1994) PEST: Model Independent Parameter Estimation 4th Edition. Watermark Numerical Computing, Australia.
- Oregon Climate Service. (2008) Climatological Data For Oregon Agricultural Regions." Accessed on 2008-0-14. <<http://www.ocs.oregonstate.edu/index.html>>
- Shanks, O.C. C. Nietch, M. Simonich, M. Younger, D. Reynolds, and K.G. Field. (2006) "Basin-Wide Analysis of the Dynamics of Fecal Contamination and Fecal Source Identification in Tillamook Bay, Oregon." *Applied and Environmental Microbiology*. 72(8):5537–5546.
- Tarboton, D. G., (1997), "A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models," *Water Resources Research*, 33(2): 309-319.
- U.S. Environmental Protection Agency (USEPA). (1996) Hydrological Simulation Program – Fortran: User's Manual for Release 11. Office of Water, Washington, DC.
- USEPA. (1999) BASINS Technical Note 4: Using HSPEXP with BASINS/NPSM. EPA-823-R-99-010. Office of Water, Washington, DC.
- USEPA. (2000) BASINS Technical Note 6, Estimating Hydrology and Hydraulic Parameters for HSPF. EPA 823-R-00-012. Office of Water, Washington, DC.
- USGS. (2008) "USGS Real-Time Water Data for USGS 14302480 TRASK RIVER ABOVE CEDAR CREEK, NEAR TILLAMOOK, OR." USGS Water Resources for Oregon. Retrieved on 2008-02-14. <http://waterdata.usgs.gov/or/nwis/uv/?site_no=14302480&PARAMeter_cd=00065,00060>