Numerical Models of Streamflow Depletion by Wells

Difficulties in applying analytical approaches to streamflow-depletion problems in real-world settings are apparent in the diagram of a portion of a stream-aquifer system shown in figure 42. Analytical solutions assume a single straight stream, yet the system shown includes a stream and two tributaries, each with irregular geometry. Similarly, analytical solutions would not be able to account for effects of the irregular edges of the aquifer. When faced with these and other complexities, a numerical-modeling approach is needed for analysis of streamflow depletion. Numerical groundwater models are the most powerful tools for understanding streamflow depletion from groundwater pumping. Some of the more common complexities of real systems that require a numerical-modeling approach include

- Irregular geometry of lateral and vertical boundaries of aquifers.
- Irregular geometry of streams, rivers, and other surface-water features.
- Non-uniform (heterogeneous) aquifer properties.
- Complex, time-varying pumping schedules at multiple wells or well fields pumping within a basin.
- Nonlinearities, such as boundary conditions and aquifer properties that change with changes in groundwater levels.

Many of the examples in this report that illustrate various aspects of streamflow depletion are derived from groundwater models of actual stream-aquifer systems. Investigators in those studies chose a numerical-modeling approach, in part because of the complexities listed above.

Groundwater-flow models simulate movement of water from areas of recharge, through an aquifer or an aquifer system, to streams and other features where groundwater discharges. Any groundwater-model program can be used to

---

Figure 42. A, Part of a hypothetical stream-aquifer system. B, Representation of that system with a finite-difference model grid consisting of 26 rows, 22 columns, and 2 layers of rectangular finite-difference blocks.
simulate depletion, as long as the program carries out rigorous calculations of system water-balance components, including inflow to the aquifer, change in storage within the aquifer, and outflow from the aquifer. The discussion that follows will reference the USGS finite-difference groundwater-model program, MODFLOW (McDonald and Harbaugh, 1988; Harbaugh, 2005), which is used worldwide to simulate many aspects of groundwater flow, including streamflow depletion. This type of model uses a grid of rectangular or square blocks to represent an aquifer (fig. 42B). In this example, a portion of a valley-fill aquifer is represented using a finite-difference grid consisting of 26 rows and 22 columns of equally spaced model cells. Aquifer properties are represented as being constant in each grid cell, and locations of boundaries occur either at the center of the cell or along the edges of cells. Use of a regular grid of finite-difference cells leads to approximations of locations of features such as the edges of the aquifer, streams, and wells (fig. 42B); however, use of a finer finite-difference grid will allow more accurate representation of locations of these features. In this example, two layers of grid cells were used to represent the aquifer in the vertical dimension.

Steady-State Flow Models

Steady-state groundwater models solve for head (groundwater levels) and flow components for the condition in which inflow rates balance outflow rates, and the rate of storage change in the aquifer is zero. As shown in the example in figure 42, inflow components might include recharge to the aquifer surface (not shown), lateral inflow from the rocks surrounding the aquifer, and flow from some stream segments to the aquifer. Outflow components would include groundwater underflow out of the model domain, flow from the aquifer to stream segments, and discharge by wells. Ultimate effects of pumping on streams (including tributaries) can be computed in three steps as follows:

**Step 1.** Run the model without pumping by a well or wells of interest and record model-computed flow rates to and from stream segments.

**Step 2.** Run the model with pumping by a well or wells of interest and record model-computed flow rates to and from stream segments.

**Step 3.** Subtract model-computed flow rates in step 2 from corresponding flow rates in step 1 to get net change in flow between the aquifer and streams.

If the pumping cannot increase recharge to the aquifer, or increase lateral inflow, or decrease underflow out, then the total change in flow to and from stream segments will equal the total pumping rate. This type of steady-state analysis cannot address the timing of depletion but is useful in understanding which features would ultimately be affected by the pumping (fig. 43).

**Figure 43.** Possible ultimate rate of depletion of different surface-water features by pumping well A at a rate of 100 gallons per minute until steady-state conditions are reached.
Streamflow Depletion by Wells—Understanding and Managing the Effects of Groundwater Pumping on Streamflow

Transient Flow Models

Transient groundwater models solve for head and flow components at discrete intervals of time, called “time steps.” In these models, head may change with time and the rate of change in aquifer storage is a component in model water budgets. For the example in figure 42, inflow components to the aquifer would be recharge to the aquifer surface, lateral inflow from surrounding rocks, flow from some stream segments to the aquifer, and the rate that water is released from aquifer storage (the condition that happens when aquifer head declines). Outflow components would include groundwater underflow out of the model domain, flow from the aquifer to stream segments, discharge by wells, and the rate that water is going into aquifer storage (the condition that occurs when aquifer head rises). The latter condition of water going into storage would not occur as a result of pumping, but it is a possible condition in part of the model domain if other water-budget components are varying through time.

The procedure for computing depletion in transient models uses the same three steps outlined above for steady-state models except that these steps must be carried out for each time step for which depletion is to be calculated. For example, if a transient model uses 10 time steps to simulate 1 year of pumping, depletion at a pumping time of 1 year can be calculated by recording flow components at time step 10 in model runs with and without pumping, and computing differences in corresponding components.

Simulated Features that can be Affected by Groundwater Pumping

Although the focus of this report is streamflow depletion, many models simulate additional features including rivers, lakes, springs, wetlands, and evapotranspiration areas. Evaluation of total effects of pumping involves calculating pumping-induced changes in inflow to and outflow from the aquifer from all relevant features. As opposed to the term “streamflow depletion,” total change in pumping-induced inflow to and outflow from the aquifer is referred to here as “capture.” Table 2 lists select MODFLOW packages that can be used to simulate features from which capture may occur.

In the Upper San Pedro groundwater model (fig. 13), outflow to streams, springs, and riparian vegetation is simulated, respectively, with the Stream, Drain, and Evapotranspiration Packages. For any given pumping location, total capture may include reduced outflow to a combination of these features. For example, at the location of well C in figure 13, total capture consists mostly of streamflow depletion with some evapotranspiration capture and no capture of spring discharge (fig. 44). Numerical models, such as presented in this example, are the only approach to compute capture from different features in a real-world setting.
Table 2. Select MODFLOW packages for representing boundary conditions in which pumping may increase inflow to the aquifer or decrease outflow from the aquifer.

<table>
<thead>
<tr>
<th>MODFLOW package</th>
<th>Common uses</th>
<th>Possible responses to pumping</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specified head (CHD)</td>
<td>Well-connected surface-water features</td>
<td>Increased inflow to aquifer, decreased outflow from aquifer</td>
<td>The package sets head in aquifer equal to head in connected surface-water feature</td>
</tr>
<tr>
<td>General-head flow (GHB)</td>
<td>Streams, rivers, other surface-water features</td>
<td>Increased inflow to aquifer, decreased outflow from aquifer</td>
<td>A linear boundary condition in which flow between boundary and aquifer is proportional to difference between boundary head and aquifer head</td>
</tr>
<tr>
<td>Stream (STR) or Streamflow Routing (SFR)</td>
<td>Streams, rivers</td>
<td>Increased inflow to aquifer, decreased outflow from aquifer</td>
<td>Can calculate stream stage, keeps track of flow in streams, streams may go dry</td>
</tr>
<tr>
<td>River (RIV)</td>
<td>Rivers, streams that do not go dry</td>
<td>Increased inflow to aquifer, decreased outflow from aquifer</td>
<td>River stage is specified at each location of cell with a river, seepage rate to aquifer becomes steady if groundwater level drops below bottom of streambed sediments</td>
</tr>
<tr>
<td>Drain (DRN)</td>
<td>Agricultural drains, springs</td>
<td>Decreased outflow from aquifer</td>
<td>Discharge to a simulated drain ceases if groundwater level drops below drain altitude</td>
</tr>
<tr>
<td>Lake (LAK)</td>
<td>Lakes</td>
<td>Increased inflow to aquifer, decreased outflow from aquifer</td>
<td>Can calculate lake stage, maintains mass balances of lakes, lakes may go dry</td>
</tr>
<tr>
<td>Evapotranspiration (EVT)</td>
<td>Groundwater evapotranspiration</td>
<td>Decreased outflow from aquifer</td>
<td>Evapotranspiration ceases if groundwater level drops below a specified level; evapotranspiration is constant with groundwater levels above another specified level</td>
</tr>
</tbody>
</table>

Figure 44. Model-computed streamflow depletion, evapotranspiration capture, and total capture for location of hypothetical well C (see figure 13) in the Upper San Pedro Basin, Arizona.
In addition to depleting streamflow, groundwater pumping can capture groundwater that otherwise would be used by plants (phreatophytes). Riparian trees, shown here, use shallow groundwater along the channel of the Mojave River in California.
Superposition Models

In the suite of methods available for computing depletion, superposition models are an intermediate approach between simple analytical solutions and complex calibrated groundwater-flow models. Unlike flow models, superposition groundwater models do not simulate natural movement of water through an aquifer. Instead of computing head and flow, these models directly compute change in head and change in flow from an added stress such as pumping. To compute streamflow depletion, the initial pre-pumping state of the superposition model is to have no flow between the stream and the aquifer. After addition of a pumping stress, computed flow from a boundary representing a stream is a direct calculation of total streamflow depletion. Because the natural flow system is not simulated, superposition models cannot determine if the depletion represents reduced groundwater discharge to the stream, increased flow of water from the stream to the aquifer (that is, induced infiltration), or a combination of these two components. Regardless, the streamflow depletion computed by a superposition model is a direct calculation of the reduced availability of surface water in the stream.

Application of the principle of superposition strictly applies to groundwater systems that respond linearly to stresses such as groundwater pumping (Reilly and others, 1987). Linearity of response means that changes from the added stress do not change the aquifer properties or configuration or function of the boundary conditions. Some examples of nonlinear responses include (1) drawdown that causes substantial changes in aquifer saturated thickness and corresponding changes in transmissivity, (2) drawing aquifer water levels below the base of a streambed so that the stream is no longer in direct hydraulic connection with the aquifer, (3) drawing water levels down below the evapotranspiration extinction depth so that evapotranspiration ceases, and (4) drying up a spring or reach of a stream. Many aquifer systems respond linearly to some range of lower stresses, and superposition can be applied in many mildly nonlinear systems (Reilly and others, 1987).

Leake and others (2005) used a superposition modeling approach to compute streamflow depletion from proposed pumping in the C aquifer in northern Arizona (figs. 17–19). In that model, both confined and unconfined areas of the aquifer and complex variations in aquifer thickness were represented. In contrast, Leake, Greer, and others (2008) computed possible depletion of the lower Colorado River using superposition models that were representative of aquifer material of uniform thickness and aquifer properties. In that application, vertical geometry and aquifer properties are treated simplistically as they would be in an analytical solution, yet all complexities of horizontal aquifer and river geometry are represented in greater detail than would be possible by an analytical solution. These types of superposition models can be constructed faster and at less expense than more complex numerical flow models and are useful in gaining an initial understanding of the possible timing of depletion. For details on how to set up a groundwater model to compute changes using superposition, see Reilly and others (1987). Durbin and others (2008) present methods of representing nonlinear boundaries in superposition models.

Simulating the Effects of Other Boundary Conditions on Streamflow Depletion

In addition to boundary conditions representing surface-water features and evapotranspiration (table 2), models can simulate the effects of no-flow or specified-flow boundaries at appropriate locations. For example, the area outside of the aquifer depicted in figure 42 may be crystalline rocks of low permeability. If interchange of water between these rocks and the aquifer is insignificant, the lateral edges of the aquifer shown in the figure could be represented as a no-flow boundary. Alternately, if some mountain-block recharge to the aquifer occurs through these rocks, the interface could be represented as a specified-flow boundary. Whether this boundary is represented as no flow or specified flow, the presence of impermeable or low-permeability rocks tends to speed up the timing of streamflow depletion because drawdown and storage change from pumping cannot extend beyond the boundary.

Ideally, all model boundaries should represent physical features such as the edge of the aquifer or a surface-water boundary. In some cases, it is impractical to construct a model that extends to all physical boundaries. In the example shown in figure 42, the area of interest may be around wells A and B, but the aquifer may extend a great distance down the valley from this area. Using the model domain shown in figure 42, an “artificial” boundary must be implemented to represent flow out of the model domain along model row 26, columns 2–17 in layer 1 and columns 4–8 in layer 2. Options for representing artificial boundaries at this location include (1) specified flow—that is, estimated downvalley flow is input for each boundary cell and the model will compute head at these cells, (2) specified head—that is, head is set to the estimated water level for each boundary cell, and the model will compute flow into or out of the model domain at each of these cells, and (3) head-dependent flow—a boundary head and “conductance” value are specified at each boundary cell so that computed flow into or out of the model varies with changes in head in the connected model cells. No matter which boundary type is selected, proximity of artificial boundaries to pumping wells is a potential problem in calculations of depletion. In figure 42, an artificial boundary along row 26 is distant from well A. Furthermore, well A is surrounded by surface-water boundaries and the natural boundary of the edge of the aquifer. Placement of an artificial boundary in model row 26 is not likely to affect calculations of depletion by pumping well A. In contrast, well B is as close to the artificial boundary as it is to the surface-water boundary. A constant-head artificial boundary along row 26 likely will result in an underestimation of depletion by well B for any given time. In contrast, a specified-flow (including no-flow) artificial boundary at that location would result in an overestimation of the progression...
of depletion by well B. To calculate depletion for well B, the model should be extended enough distance downstream so that the drawdown from this well does not reach the artificial downstream boundary.

In some aquifers, groundwater divides that approximately underlie watershed boundaries define the extent of a subunit of the aquifer beneath the watershed. If interest is in modeling groundwater processes in the particular watershed, a common practice is to represent the bounding groundwater divides as no-flow boundaries. Under flow conditions that are steady, the groundwater divides are in fact no-flow boundaries because there is no movement of groundwater across the divides. A possible result of added pumping in the watershed, however, is that groundwater divides will be moved outward into adjacent watersheds. Divides that are represented as no-flow boundaries that are fixed in space may result in computed rates of streamflow depletion that occur faster than would be computed using a representation of divides that can move in response to pumping. If drawdown from pumping can propagate to groundwater divides, the best approach is to make the domain of the model large enough so that model boundaries are not on the groundwater divides adjacent to the pumping locations. In the example shown in figure 45, pumping locations A and B are both in watershed 1. Pumping location B is close to the stream segments in watershed 1, and drawdown from pumping at this location probably would not reach the boundaries of the watershed. In this case, a model that includes only the portion of the aquifer underlying watershed 1 may be a reasonable approach to simulating depletion from pumping at location B. In contrast, location A is closer to the watershed boundary than it is to stream segments in watershed 1. Pumping at location A likely would deplete surface water in stream segments in watersheds 1, 2, 3, and 4. Use of a model that includes only watershed 1 for this pumping location would force some of the drawdown, storage change, and depletion that should occur in adjacent watersheds to occur only in watershed 1. The result is an overestimation of depletion in watershed 1 and an underestimation of depletion in adjacent watersheds 2, 3, and 4. To effectively simulate depletion from pumping at location A, the model must include the part of the aquifer underlying watersheds 1, 2, 3, and 4.

Response Functions and Capture Maps

Two important uses of analytical and numerical models are to generate streamflow-depletion response functions and capture maps (which are a type of response function). Response functions characterize the unique functional relation between pumping at a particular location in an aquifer and the resulting depletion in a nearby stream and provide hydrologists and water-resource managers with insight into how a particular stream or stream reach will respond to pumping at a particular well. Although response functions have been defined and used in different ways (and referred to by different names), all response functions have the common characteristic that they represent a change in streamflow that results from a change in pumping rate at a single well, independently of other pumping or recharge stresses that may be occurring simultaneously within the aquifer. As demonstrated by the many examples provided in this report, the response function for a particular well and streamflow-location pair reflects the combined effects of several factors, including the distance of the well from the stream, the geometry of the aquifer system and stream network, the hydraulic properties of the aquifer and streambed materials, and the vertical depth of pumping from the aquifer.

Theoretically, response functions could be determined by monitoring changes in streamflow that result from pumping at a particular well, but this approach is often not technically feasible because of difficulty in separating depletion changes from streamflow responses to other changes, such as those driven by climate. In practice, response functions are determined by using analytical or numerical models. Model-simulated response functions are shown as either the rate or volume of streamflow depletion that occurs in response to pumping at a particular rate or, alternatively, as dimensionless fractions of the pumping rate or total volume of withdrawal at a well, as described in Box B. Reporting response functions as dimensionless quantities is particularly useful when streamflow depletion responds linearly to pumping, because the dimensionless quantities are constants whose values are independent of the particular pumping rate used for their calculation. For example, if the dimensionless response function were 0.5 for a time and location of interest, the rate of streamflow depletion would be 0.5 Mgal/d for a pumping rate of 1.0 Mgal/d, and 2.0 Mgal/d for a pumping rate of 4.0 Mgal/d. As described previously, a stream-aquifer system is linear if (1) the transmissivity of the aquifer does not change as the pumping rates of the wells change and (2) the rate of flow at other pumping or recharge stresses that may be occurring simultaneously within the aquifer.

Some examples of the application of response functions to stream-aquifer systems include those described by Maddock (1974), Morel-Seytoux and Daly (1975), Morel-Seytoux (1975), Illangasekare and Morel-Seytoux (1982), Danskin and Gorelick (1985), Maddock and Lacher (1991), Reichard (1995), Male and Mueller (1992), Mueller and Male (1993), Fredericks and others (1998), Barlow and others (2003), Cosgrove and Johnson (2004, 2005), and Ahlfeld and Hoque (2008). Although this report focuses on streamflow-depletion response functions, it should be noted that response functions also can be generated for other types of variables that describe the state of a groundwater system, such as groundwater-level declines, groundwater velocities, and aquifer-storage changes (see, for example, Maddock and Lacher, 1991; Gorelick and others, 1993; Ahlfeld and Mulligan, 2000; and Ahlfeld and others, 2005 and 2011).
Figure 45. Five adjacent watersheds in north-central Michigan overlying a groundwater system. Pumping locations A and B are both within watershed 1, but construction of a model to compute depletion for a well at location A will require inclusion of some adjacent watersheds in the model domain (modified from Reeves and others, 2009).

the stream-aquifer boundary is a linear function of the groundwater level near the stream.

Response functions that characterize total depletion of all streams (and sometimes other features) within a basin are referred to here as “global response functions.” Conversely, response functions that characterize depletion in a particular stream or segment of a stream are referred to as “local response functions.” Furthermore, “transient response functions” characterize depletion through time until some maximum time interval and “steady-state response functions” characterize ultimate depletion without regard to the time required to reach that state. Some key points relating to these types of response functions are as follows:

1. Transient response functions for each pumping location are defined by a number of values through time.

2. Global transient response functions expressed as a fraction of pumping rate will start at zero at the onset of pumping and will trend toward a maximum value of 1.0, as shown by the curve in figure B–1B in Box B.
3. Local transient response functions may trend toward a value less than 1.0 if the pumping causes depletion in locations in addition to the stream or segment of interest.

4. Steady-state response functions are a single value for each pumping location.

5. Global steady-state response functions are equal to 1.0, assuming that streams are the ultimate source of pumped water.

6. The sum of local steady-state response functions for all stream segments affected by a pumped well is equal to 1.0.

Concepts relating to global and local, transient and steady-state response functions are further illustrated by the two maps in figure 46. Dimensionless response-function values are shown in figure 46A for three wells in the watershed after 10 years of pumping. The stream location for which the response coefficients were determined is the outflow point from the basin. For this hypothetical aquifer, the response-function value for well A is largest because the well is closer to the stream network than the other two wells; the value for well C is smallest because it is furthest from the stream network. Figure 46B illustrates steady-state local response-function values for one of the tributaries to the main stem (stream segment 1). In this example, the system has reached steady-state conditions, and streamflow depletion is the only source of water to the wells. Each response-function value shown in figure 46B represents the change in streamflow at the point just upstream from the confluence of the tributary with the main stem in response to pumping at each of the three wells. The response-function value is largest for well A because it is adjacent to the tributary, whereas the response-function value for well B is lowest because it captures most of its discharge from the main stem (stream segments 3 and 5) and very little discharge from the tributary denoted as stream segment 1.

An alternative approach to calculating response functions for only a few locations is to show maps of the spatial distribution of values of response functions for large regions of an aquifer. Response-function maps are particularly useful for illustrating the effects of pumping location on streamflow depletion within a large set of possible pumping locations within an aquifer (Leake and others, 2010). One approach has been to show values of the global transient response function for a particular pumping time, such as 10 years (for examples, see Leake, Pool, and Leenhouts, 2008; and Leake and Pool, 2010). Such maps, referred to as “capture maps,” provide water-resource managers with a visual tool that can be used to determine the effects of pumping at specific locations on total streamflow depletion. Using values from local transient or steady-state response functions, capture maps also can be created to illustrate effects of pumping location on specific streams or stream segments (Cosgrove and Johnson, 2005; Leake and others, 2010). The goal of any of these types of capture maps is to help convey an understanding of the effects of well placement on depletion in areas of interest and to provide a possible tool for use in siting new wells or recharge facilities.

The procedure for making response-function or capture maps requires use of a well-constructed groundwater model. The model must include streams and other appropriate features as head-dependent boundaries, and any boundaries that do not represent actual physical features must be at distances such that they do not affect calculated depletion. For details on constructing these maps, see Leake and others (2010).

Example capture maps showing global transient response functions for the Upper San Pedro Basin (Leake, Pool, and Leenhouts, 2008) are shown in figure 47. The mapped area is the extent of the lower basin-fill aquifer, represented as layer 4 of the groundwater model by Pool and Dickinson (2007). Global response in this case is mostly from changes to streamflow in the San Pedro and Babocomari Rivers, but also includes minor components of reductions in groundwater evapotranspiration and in springflow (that is, groundwater discharge to springs). For the times shown, 10 years (fig. 47A) and 50 years (fig. 47B), pumping in the area shaded in the darkest blue indicates that depletion would be between 0 and 10 percent (a fraction from 0 to 0.1) of the pumping rate for that time. Similarly, depletion from pumping in the darkest red area on each map indicates depletion would be between 90 and 100 percent (a fraction from 0.9 to 1.0) of the pumping rate for that time. As would be expected, the general pattern is that depletion from pumping nearer the rivers is greater than from pumping at more distant locations for either time shown; however, amounts of depletion vary along the streams. Leake, Pool, and Leenhouts (2008) attribute the complexities in the patterns shown to spatial variations in aquifer geometry and aquifer properties. A low-permeability clay layer that exists between some pumping locations and connected streams may...
contribute to complexity in the patterns shown. Comparison of the 10-year and 50-year capture maps indicates the progression of depletion through time, with substantially more areas of yellow, orange, and red colors in the 50-year map than in the 10-year map. Because these maps show global response, maps for increased pumping time would be more red, and if pumping time was such that a new steady-state condition would be reached for any pumping location, the map would be solid red.

Other examples of maps to understand depletion as a function of pumping location include mapping of stream-depletion factors by Jenkins and Taylor (1974) and Burns (1983). COHYST Technical Committee (2004), Peterson and others (2008), and Stanton and others (2010) used numerical models to map lines of equal depletion as a fraction of volume pumped at specific times for locations in Nebraska. Some authors have used response-function maps to group wells (or regions of an aquifer) having similar effects on specific stream reaches into aquifer response zones. Examples of this approach are provided for the Eastern Snake River Plain aquifer in Idaho by Hubbell and others (1997) and Cosgrove and Johnson (2004 and 2005).

In addition to mapping responses for a specific time, it is also possible to construct maps showing the time it would take to reach a particular depletion level of interest. For example, depletion-dominated supply of pumped water (fig. 9) occurs when depletion exceeds half of the pumping rate. Figure 48 shows the time it would take to reach depletion-dominated supply of pumping from the lower basin-fill aquifer in the Upper San Pedro Valley. For most areas adjacent to the Babocomari and San Pedro Rivers, depletion-dominated supply is reached within 20 years of pumping, but in the southern extent of the aquifer and in places along the east and west margins of the aquifer, depletion-dominated supply would not be reached within 100 years (fig. 48).

---

**Figure 46.** A, Diagram of transient response functions for the outflow point of the basin after 10 years of pumping. B, Diagram of steady-state response functions for a tributary stream to the main stem (modified from Leake and others, 2010).
Figure 47. Computed capture of streamflow, riparian evapotranspiration, and springflow that would result for withdrawal of water at a constant rate for, A, 10 years and, B, 50 years from the lower basin-fill aquifer in the upper San Pedro Basin, Arizona. The color at any location represents the fraction of the withdrawal rate by a well at that location that can be accounted for as changes in outflow from and (or) inflow to the aquifer for model boundaries representing streams, riparian vegetation, and springs (from Leake, Pool, and Leenhouts, 2008).
**Approaches for Monitoring, Understanding, and Managing Streamflow Depletion by Wells**

*B. 50 years*

**Figure 47.** Continued.
**Figure 48.** Computed time to reach a depletion-dominated supply of pumped water for the lower basin-fill aquifer in the Upper San Pedro Basin, Arizona, when streamflow depletion exceeds half of the pumping rate.
Management of Streamflow Depletion

Managing the effects of streamflow depletion by wells is one of the most common and often one of the most challenging aspects of conjunctively managing groundwater and surface-water systems. The effect of a groundwater withdrawal on the timing, rates (or volumes), and locations of streamflow depletions is substantially different from those caused by a surface-water withdrawal, which has an immediate effect on the rate of streamflow at the point of withdrawal. As demonstrated throughout this report, there can be a significant delay between when a well begins to pump and when the impacts of that pumping are realized in nearby streams. These delays can range from days to decades, and in some cases the full impact of pumping may not be realized within a period of time that is meaningful for practical management of a water-supply system. Moreover, unless the pumping site is located very close to the stream, streamflow will not recover immediately after pumping stops because of the residual pumping effects on streamflow depletion. As a result, in many hydrogeologic settings management of pumping rates in response to short-term fluctuations in streamflow conditions such as might be desired during periods of low streamflow or drought is unlikely to have an immediate impact on streamflow (Jenkins, 1968a; Bredehoeft, 2011a).

Other factors, such as determining the locations of streamflow depletions, also complicate management strategies. Streamflow reductions caused by pumping occur both upstream and downstream from the point of withdrawal, and may be distributed among more than one stream; the exact locations of these reductions may not be easily defined without extensive field investigations or modeling studies. Also, many aquifers are tapped by large numbers of wells, and it may not be possible to accurately determine the history of groundwater pumping at each well. It is the sum total of streamflow effects caused by pumping from many wells that need to be managed. A related issue is that an individual well may not produce depletion that is measurable. This is particularly true for large rivers. Finally, aquifers are hidden from view, and even extensive field programs may not be able to define the hydrogeology of a groundwater system in sufficient detail to accurately define the timing of streamflow depletion from an individual well.

In spite of these challenges, water-resource managers often want to understand how pumping rates and pumping schedules might be managed to control the effects of pumping on streamflow depletion. Doing so requires both a long-term perspective (Bredehoeft, 2011a) and an understanding of how streamflow responds to pumping at each well individually and at all wells simultaneously. Several examples of the types of analyses that can be done to determine long-term impacts have been illustrated in this report, such as the generation of response functions and capture maps by use of numerical models. Simulations of specific time-varying and cyclic pumping schedules at individual wells also are useful to determine how aquifer properties and well distance may affect the timing and variability of streamflow depletion, such as demonstrated for three irrigation wells pumping at various distances from a stream (fig. 21).

An example of some of the issues involved in managing streamflow depletion is illustrated for a typical water-resource management problem, which is to determine pumping schedules that meet water-supply demands while simultaneously meeting minimum streamflow requirements at specific stream locations and for specific periods of time. For this example, an evaluation is made of a single, hypothetical stream that is in hydraulic connection with an aquifer that is pumped from June through August to supply water for irrigation. In the absence of pumping, the annual pattern of streamflow for the hypothetical system ranges from a maximum of 55.0 ft³/s in early spring (March 31) to a minimum of 40.5 ft³/s in early fall (September 30) (fig. 49). Water managers have determined that a minimum streamflow requirement of 35 ft³/s is to be maintained throughout the year to meet instream flow needs. Irrigators want to pump 6 Mgal/d (9.3 ft³/s) from the aquifer from two possible well sites to meet their irrigation requirements. The management problem is to determine whether or not pumping rates can be determined for the two wells to simultaneously meet the irrigation demands and instream flow requirements.

Because of the simplicity of the physical system, the Glover analytical model is used to determine streamflow depletion caused by different combinations of pumping rates at each well (fig. 49). The first well (A) is located 300 ft from the stream and the second well (B), 1,000 ft from the stream. Three of the many possible combinations of pumping rates at the two wells to meet the irrigation demand are shown in figure 49. When the well closest to the stream is pumped at the full 6 Mgal/d, the minimum streamflow requirement is not met for a short period of time at the end of each pumping cycle (late August into early September). However, when the pumping rate at this well is reduced to 3 Mgal/d and the remaining 3 Mgal/d of the demand is supplied by pumping at the well furthest from the stream, the maximum rate of depletion is reduced and the minimum streamflow requirement is met. The maximum rate of depletion is further reduced as the proportion of pumping from well B increases, with the smallest effect occurring for the case in which all of the withdrawal is from well B. The results shown for this simple stream-aquifer system reflect differences in the underlying streamflow response functions for each well, which in this case result from differences in the distance of each well from the stream.

This simple example demonstrates how pumping rates might be managed to control the timing of streamflow depletion by taking advantage of the variability in streamflow responses to pumping at different wells. For a water-supply system with just a few wells and a single stream location of interest, alternative pumping rates can be tested relatively easily to determine if pumping schedules can be found that simultaneously meet water-supply demands and minimum instreamflow requirements. A trial-and-error testing approach such as this becomes impractical however for a typical hydrogeologic
Figure 49. Streamflow for a hypothetical stream-aquifer system for different pumping conditions. Hydraulic diffusivity of aquifer is 10,000 feet squared per day. Wells are located 300 feet (well A) and 1,000 feet (well B) from the stream. (Rates of streamflow depletion were calculated by using a computer program described in Reeves (2008), which includes the Glover analytical model. The calculated depletion rates were then subtracted from the streamflow hydrograph without pumping (top curve on the figure) to determine the resulting decreased rates of streamflow. Mgal/d, million gallons per day)

setting in which there are multiple pumping wells and multiple streams for which minimum streamflow requirements have been established. For complex settings such as these, a technique called simulation-optimization modeling might be used. In this approach, a numerical simulation model (or, less often, an analytical model) is combined with a mathematical optimization technique to identify pumping schedules that best meet management objectives and constraints. The simulation model accounts for the physical behavior of the stream-aquifer system, whereas the optimization model accounts for the management aspects of the problem. Examples of the use of simulation-optimization modeling for management of streamflow depletion by wells include those described by Young and Bredehoeft (1972), Maddock (1974), Morel-Seytoux and Daly (1975), Morel-Seytoux (1975), Illangasekare and Morel-Seytoux (1982), Bredehoeft and Young (1983), Peralta and others (1988), Matsukawa and others (1992), Male and Mueller (1992), Mueller and Male (1993), Basagaoglu and Marino (1999), Barlow and others (2003), Ashfeld and Hoque (2008), and Stanton and others (2010). The technique is described in detail by Gorelick and others (1993) and Ahlfeld and Mulligan (2000) and has been implemented for use with some of the widely available groundwater models (for example, the Groundwater-Management Process developed for MODFLOW by Ahlfeld and others, 2005).

An example of the use of simulation-optimization modeling to determine long-term average pumping schedules that meet groundwater-development goals and minimum streamflow requirements is provided by the results of a study for the Big River Basin of Rhode Island by Granato and Barlow (2005). Minimum streamflow requirements that are protective of aquatic habitats are often not well defined, and, as a result, water-resource and environmental agencies commonly evaluate the effects of alternative streamflow standards on groundwater-development options before implementing a particular regulatory standard. This was the case for the Big River Basin when, at the time of the study, State water-resource and environmental-management agencies were considering more than a dozen alternative minimum streamflow standards for implementation.

A numerical model developed to simulate groundwater flow and groundwater/surface-water interactions within the basin was linked with an optimization model that represented management goals and constraints. The management object was to determine the maximum amount of groundwater that could be pumped from more than a dozen wells in the basin. The maximum rate of withdrawal was limited, however, by constraints placed on the minimum amount of streamflow required at four streamflow locations. Each of the proposed minimum streamflow standards was defined in terms of the minimum streamflow required at each streamflow site per square mile of drainage area to each site. For example, for a defined standard of 0.5 cubic foot per second per square mile \([\text{ft}^3/\text{s}/\text{mi}^2]\), the minimum flow required at a stream location having a 30 mi² drainage area would be 15 ft³/s.

The combined simulation-optimization model was run several times to determine a range of optimal withdrawal rates for alternative definitions of the minimum streamflow standard at the four stream sites. Not surprisingly, the results of the simulation-optimization model indicated that as the minimum
streamflow standard was increased, the total amount of pumping within the basin that would be possible decreased (fig. 50). Graphs such as the one shown in figure 50 are often referred to as trade-off curves, because they illustrate the trade-offs that decision makers must consider between minimum streamflow standards and maximum rates of groundwater development. For example, point A on the graph corresponds to a minimum streamflow standard of 0.5 \((\text{ft}^3/\text{s})/\text{mi}^2) at each of the four stream sites. For this proposed standard, an average annual pumping rate of 12 Mgal/d from the basin would be possible. Although the overall results of the study could be anticipated without a model—that is, that groundwater development would decrease as the streamflow standard was increased—the specific rates of pumping at each of the wells, and therefore from the basin as a whole, could not. The shape of the curve in figure 50 reflects the unique hydrogeologic and hydrologic conditions within the basin and the distribution of the pumping wells relative to the locations of the streamflow constraint sites.

Both of the examples described in this section and illustrated in figures 49 and 50 were related to managing groundwater withdrawals to meet specified rates of minimum streamflow. However, a number of studies have demonstrated the utility of artificial-recharge strategies at injection wells or artificial-recharge basins to increase streamflow or to offset the effects of withdrawals, such as was illustrated in figure 26. Additional examples of the use of artificial recharge to augment streamflow are provided in the studies by Burns (1984), Bredehoeft and Kendy (2008), and Barber and others (2009).
Conclusions

Understanding and managing streamflow depletion is a major challenge in regulation and management of groundwater use in coupled groundwater/surface-water systems. Scientific research in conjunction with practical applications of this research to real-world field settings over the past seven decades have made important contributions to the understanding of the processes and factors that affect the timing, locations, and rates of streamflow depletion, and for evaluating alternative approaches for managing depletion. The following primary conclusions can be drawn from this research and the many field applications:

Sources of water to a well: The sources of water to a well are reductions in aquifer storage, increases in the rates of recharge (inflow) to an aquifer, and decreases in the rates of discharge (outflow) from an aquifer. The latter two components are referred to as capture. In many groundwater systems, the primary components of capture are groundwater that would otherwise have discharged to a connected stream or river in the absence of pumping (referred to as captured groundwater discharge) and streamflow drawn into an aquifer because of the pumping (induced infiltration of streamflow).

Components of streamflow depletion: Both captured groundwater discharge and induced infiltration of streamflow result in reductions in the total rate of streamflow. Streamflow depletion, therefore, is the sum of captured groundwater discharge and induced infiltration. Captured groundwater discharge is often the primary component of streamflow depletion, but if pumping rates are relatively large or the locations of withdrawal relatively close to a stream, then induced infiltration may become an important component of streamflow depletion.

Time response of streamflow depletion: Reductions in aquifer storage are the primary source of water to a well during the early stages of pumping. The contribution of water from storage decreases and the contribution from streamflow depletion increases with time as the hydraulic stress caused by pumping expands outward away from the well and reaches one or more areas of the aquifer from which water can be captured. At some point in time, streamflow depletion will be the dominant source of water to the well (that is, more than 50 percent of the discharge from the well) and after an extended period of time may become the only source of water to the well. The time at which streamflow depletion is the only source of water to a well is referred to as the time to full capture.

Factors that affect streamflow depletion: Many factors affect the timing of the response of streamflow depletion to pumping at a particular well. These include the geologic structure, dimensions, and hydraulic properties of the groundwater system; the locations and hydrologic conditions along the boundaries of the groundwater system, including the streams and streambed hydraulic properties; the horizontal and vertical distances of wells from the streams; and pumping schedules at the wells. In a system with predominantly horizontal groundwater flow, well distance and the hydraulic diffusivity of the aquifer are two of the most important factors. Streamflow depletion will occur more rapidly for a well pumping relatively close to a stream from an aquifer having a relatively high value of hydraulic diffusivity and less rapidly for a well pumping far from a stream from an aquifer having a relatively low value of hydraulic diffusivity. In settings in which vertical groundwater-flow components are important, distributions of vertical and horizontal hydraulic conductivity, specific storage, specific yield, and aquifer thickness, in addition to well distance from the stream, are the key properties that control the timing of depletion. Aquifer extent is also an important variable. The time to full capture for wells pumping from narrow river-valley aquifers that are bounded at their margins by relatively impermeable materials can be short (days to years), whereas the time to full capture for wells pumping from regionally extensive aquifer systems can be quite long (years to centuries).

Effects of confining layers on depletion: Various geologic features that act as conduits or barriers to groundwater flow can affect the timing of streamflow depletion from groundwater pumping and also can affect which streams are affected by the pumping. Horizontal or nearly horizontal beds of clay, silt, or other geologic materials that are of substantially lower hydraulic conductivity than adjacent aquifer material may be laterally discontinuous or form laterally extensive confining units that separate adjacent aquifers. Even though confining layers can slow down the progression of depletion in comparison to equivalent aquifer systems without confining layers, it is not reasonable to expect that pumping beneath an extensive confining layer will entirely eliminate depletion. For some well locations, discontinuous confining beds of clay may actually increase the depletion process relative to a condition in which the beds are absent.

Aquifer recharge and streamflow depletion: The long-term average or transient rates of recharge to an aquifer (or the predevelopment rates and directions of flow within an aquifer) will not affect the total amount of depletion that results from pumping a well, because the sources of capture to a well result from changes in the predevelopment recharge and discharge rates to or from an aquifer and not the absolute rates of recharge or discharge themselves. Because the natural rate of recharge does not affect the quantity of streamflow that can be captured by a well, it cannot be assumed that the total amount of groundwater development from an aquifer system is “safe” or “sustainable” at rates up to the long-term average recharge rate. The amount of depletion that can be captured is dependent on the total amount of water in the stream and the amount of reduced streamflow that a community or regulatory authority is willing to accept. However, recharge rates do affect the relative contributions of captured groundwater discharge and induced infiltration to total streamflow depletion: relatively high rates of recharge (or predevelopment flow rates through the aquifer) will result in relatively high rates of captured groundwater discharge, whereas relatively low rates of recharge will result in relatively high rates of induced infiltration.
Distribution of streamflow depletion along stream reaches: Groundwater pumping causes streamflow depletion in streams and stream reaches that are both upgradient and downgradient from the location of withdrawal; the effect of pumping is not confined to those reaches that are immediately adjacent to the well. Some stream reaches will be affected more than others, depending on the distance of the pumped well from each reach and the three-dimensional distribution and hydraulic properties of the sediments that compose the groundwater system and adjoining streambeds. Cumulative streamflow depletion increases in the downstream direction of a basin, and the total amount of depletion in the direction of the outflow point (or points) from the basin will, over time, tend toward the total pumping rate of the well or wells that pump from the basin.

Disconnected and dry stream reaches: Two important assumptions that have been made throughout the report are that the stream and underlying aquifer remain hydraulically connected by a continuous saturated zone and that the stream does not become dry. In extreme cases of large-scale groundwater development and limited streamflow, groundwater levels can be drawn down below the bottom of the streambed and the stream may eventually lose all of its water to the aquifer. Under such conditions, there will not be enough water available from streamflow depletion to offset the pumping by a well or wells in the aquifer.

Streamflow depletion after pumping stops: Streamflow depletion continues after pumping stops because it takes time for groundwater levels to recover from the previous pumping stress and for the depleted aquifer defined by the cone of depression to be refilled with water. The time of maximum streamflow depletion often may occur after pumping has stopped. Eventually, the aquifer and stream may return to their pre-pumping conditions, but the time required for full recovery may be quite long and exceed the total time that the well was pumped. Over the time interval from when pumping starts until the system fully recovers to its pre-pumping levels, the volume of streamflow depletion will equal the volume of water pumped.

Variable- and cyclic-pumping effects: Pumping schedules at wells fluctuate in response to water-supply demands that change on daily, seasonal, and longer-term intervals. Intermittent- and cyclic-pumping schedules result in variable or cyclic patterns of streamflow depletion, but the overall effect of an aquifer is to damp the variability and amplitude (range) of pumping rates such that the resulting rates of streamflow depletion are less variable and smaller in amplitude than the pumping stress itself. The damping effect is enhanced as the distance of the pumped well increases from a stream or the diffusivity of the aquifer decreases, and at some distance the effects of an intermittent- or cyclic-pumping pattern become indistinguishable from a constant pumping pattern at a cycle (or long-term)-average pumping rate.

Basinwide analyses: Many groundwater basins have hundreds or thousands of pumped wells. Individually, these wells may have little effect on streamflow depletion, but small effects of many wells within a basin can combine to produce substantial effects on streamflow and aquatic habitats. Moreover, basinwide groundwater development typically occurs over a period of several decades, and the resulting cumulative effects on streamflow depletion may not be fully realized for years. As a result of the large number of wells and complex history of development, it is often necessary to take a basinwide perspective to assess the effects of groundwater withdrawals on streamflow depletion.

Streamflow depletion and water quality: Many of the problems associated with streamflow depletion do not require that the two components of depletion—captured groundwater discharge and induced infiltration—be differentiated, or individually quantified. This is the case, for example, for issues that are strictly related to questions of streamflow quantity, such as for water-rights administration or determination of minimum instream-flow requirements for aquatic habitats. For water-quality concerns, however, the relative contribution of captured groundwater discharge and induced infiltration has important implications to the resulting quality of the water in the stream, in the aquifer system, and pumped from wells. As a result, techniques of analysis that are needed to evaluate water-quality problems associated with streamflow depletion must be able to identify the specific components of depletion. For example, analytical solutions and superposition numerical models that can only identify changes in streamflow and not the absolute amount of streamflow will not be appropriate, whereas numerical models, particularly those that can track particles of water through a groundwater system or can simulate solute-transport processes may be.

Field methods for identifying and monitoring streamflow depletion: Two general approaches are used to monitor streamflow depletion: (1) short-term field tests lasting several hours to several months to determine local-scale effects of pumping from a specific well or well field on streams that are in relative close proximity to the location of withdrawal and (2) statistical analyses of hydrologic and climatic data collected over a period of many years to test correlations between long-term changes in streamflow conditions with basinwide development of groundwater resources. Direct measurement of streamflow depletion is made difficult by the limitations of streamflow-measurement techniques to accurately detect a pumping-induced change in streamflow, the ability to differentiate a pumping-induced change in streamflow from other stresses that cause streamflow fluctuations, and by the diffusive effects of a groundwater system that delay the arrival and reduce the peak effect of a particular pumping stress.

Analytical-modeling methods to estimate streamflow depletion: Several analytical solutions to the groundwaterflow equation have been developed to estimate streamflow depletion by wells. These solutions are based on highly simplified representations of field conditions that are necessary to develop mathematical solutions to the groundwater-flow equation but that limit their applicability to real-world field conditions. Some of the important limitations of analytical solutions are that they cannot adequately represent aquifer
heterogeneity, the presence of multiple streams or complex stream geometry, or aquifers having complex, three-dimensional geometries. Nevertheless, analytical solutions provide insight into several of the factors that affect streamflow depletion and are often used to make an initial estimate of the effect of a particular well on a nearby stream.

**Numerical-modeling methods to estimate streamflow depletion:** Numerical models are the most robust method for determining the rates, locations, and timing of streamflow depletion caused by pumping because they are capable of handling many of the common complexities of real groundwater systems. They are the only effective method for determining detailed, basinwide water budgets that account for the effects of complex pumping histories from large numbers of wells on all types of hydrologic features, including streams. Numerical models can be used to generate streamflow-depletion response functions and capture maps. Response functions characterize the unique functional relation between pumping at a particular location and the resulting depletion in a nearby stream or stream network, independently of other pumping or recharge stresses that may be occurring simultaneously within the aquifer. Capture maps, which are a type of response function, show the spatial distribution of response-function values for large regions of an aquifer, and provide a visual tool to illustrate the effects of pumping location on streamflow depletion within a large set of possible pumping locations within an aquifer.

**Management of streamflow depletion:** Managing streamflow depletion by wells is challenging because of the significant time delays that often occur between when pumping begins and when the effects of that pumping are realized in nearby streams. In many cases, it is not possible to reduce pumping rates during periods of low streamflow to substantially affect flow during the period of stress. Effective management of streamflow depletion requires both a long-term perspective and an understanding of how streamflow responds to pumping at each well individually and at all wells simultaneously. Numerical models are the most effective means to determine the effects of pumping on streamflow and to determine whether or not pumping schedules can be manipulated to meet minimum streamflow requirements. For conditions in which many wells pump from the same basin, the use of numerical models can be enhanced by their coupling with management models that identify the optimal pumping strategies to meet water-resource goals and constraints.

**Depletion of other hydrologic features:** Most aquifer systems are complex, with water moving from areas of recharge through geologic materials and discharging to streams, springs, rivers, and wetlands, and by plants that use groundwater. The introduction of groundwater pumping can affect all features connected to an aquifer. The emphasis of this report has been on the effects of pumping on connected streams, although most of the discussion that has been presented is equally applicable to other connected features.

### Acknowledgments

The authors thank William Alley, Marshall Gannett, Thomas Reilly, and Kay Hedrick-Naugle of the U.S. Geological Survey for their helpful technical and editorial comments on earlier drafts of this report and Christine Mendelsohn, U.S. Geological Survey, for preparation of the illustrations and layout of the final report. The authors also thank Michael Collier, Bob Herrmann, and U.S. Geological Survey personnel for allowing the use of their photographs in this report.

### References Cited


Bourg, A.C.M., and Bertin, Clotilde, 1993, Biogeochemical processes during the infiltration of river water into an alluvial aquifer: Environmental Science and Technology, v. 27, no. 4, p. 661–666.


Hunt, Bruce, 2003a, Unsteady stream depletion when pumping from semiconfined aquifer: Journal of Hydrologic Engineering, v. 8, no. 1, p. 12–19.


Jenkins, C.T., 1968c, Electric-analog and digital-computer model analysis of stream depletion by wells: Ground Water, v. 6, no. 6, p. 27–34.


Su, G.W., Jasperse, James, Seymour, Donald, Constantz, James, and Zhou, Quanlin, 2007, Analysis of pumping-induced unsaturated regions beneath a perennial river: Water Resources Research, v. 43, W08421, 14 p. (Also available at http://dx.doi.org/10.1029/2006WR005389.)


